Ephraim Suhir
Y.C. Lee
C.P. Wong
Editors

Micro- and Opto-Electronic Materials and Structures:

Physics, Mechanics, Design, Reliability, Packaging

Volume I

Springer

Micro- and Opto-Electronic Materials and Structures: Physics, Mechanics, Design, Reliability, Packaging

Volume I
Materials Physics—Materials Mechanics

Micro- and Opto-Electronic Materials and Structures: Physics, Mechanics, Design, Reliability, Packaging

Volume I
Materials Physics—Materials Mechanics

Edited by:

E. Suhir
University of California, Santa Cruz
Santa Cruz, California, USA

University of Maryland
College Park, Maryland, USA

Y.C. Lee
University of Colorado
Boulder, Colorado, USA

C.P. Wong
Georgia Tech
Atlanta, Georgia, USA

Springer

E. Suhir
University of California, Santa Cruz
Santa Cruz, California
and
University of Maryland
College Park, Maryland

Y.C. Lee
University of Colorado
Boulder, Colorado

C.P. Wong
Georgia Institute of Technology
Atlanta, Georgia

Micro- and Opto-Electronic Materials and Structures:
Physics, Mechanics, Design, Reliability, Packaging

Library of Congress Control Number: 2006922729

ISBN 0-387-27974-1 e-ISBN 0-387-32989-7
ISBN 978-0-387-27974-9

Printed on acid-free paper.

© 2007 Springer Science+Business Media, Inc.
All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Springer Science + Business Media, Inc., 233 Spring Street, New York, NY 10013, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now know or hereafter developed is forbidden.
The use in this publication of trade names, trademarks, service marks and similar terms, even if the are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

Printed in the United States of America.

9 8 7 6 5 4 3 2 1 SPIN 11055464

springer.com

Contents

Volume I

List of Contributors xxvii

Preface xxxi

Materials Physics

Chapter 1
Polymer Materials Characterization, Modeling and Application
L.J. Ernst, K.M.B. Jansen, D.G. Yang, C. van 't Hof, H.J.L. Bressers, J.H.J. Janssen and G.Q. Zhang 3

1.1. Introduction 3
1.2. Polymers in Microelectronics 4
1.3. Basics of Visco-Elastic Modeling 6
1.3.1. Preliminary: State Dependent Viscoelasticity 6
1.3.2. Incremental Relationship 10
1.3.3. Linear State Dependent Viscoelasticity 13
1.3.4. Isotropic Material Behavior 14
1.3.5. Interrelations between Property Functions 15
1.3.6. Elastic Approximations 17
1.4. Linear Visco-Elastic Modeling (Fully Cured Polymers) 18
1.4.1. Introduction 18
1.4.2. Static Testing of Relaxation Moduli 18
1.4.3. Time-Temperature Superposition Principle 23
1.4.4. Static Testing of Creep Compliances 24
1.4.5. Dynamic Testing 27
1.5. Modeling of Curing Polymers 34
1.5.1. "Partly State Dependent" Modeling (Curing Polymers) 35
1.5.2. "Fully State Dependent" Modeling (Curing Polymers) 49
1.6. Parameterized Polymer Modeling (PPM) 53
1.6.1. PPM Hypotheses 54
1.6.2. Experimental Characterizations 55
1.6.3. PPM Modeling in Virtual Prototyping 62
Acknowledgments 62
References 62

Chapter 2
Thermo-Optic Effects in Polymer Bragg Gratings
Avram Bar-Cohen, Bongtae Han and Kyoung Joon Kim 65

2.1. Introduction 65
2.2. Fundamentals of Bragg Gratings 67
2.2.1. Physical Descriptions 67
2.2.2. Basic Optical Principles 68
2.3. Thermo-Optical Modeling of Polymer Fiber Bragg Grating 70
2.3.1. Heat Generation by Intrinsic Absorption 70
2.3.2. Analytical Thermal Model of PFBG 78
2.3.3. FEA Thermal Model of PFBG 80
2.3.4. Thermo-Optical Model of PFBG 80
2.4. Thermo-Optical Behavior of PMMA-Based PFBG 84
2.4.1. Description of a PMMA-Based PFBG and Light Sources 85
2.4.2. Power Variation Along the PFBG 86
2.4.3. Thermo-Optical Behavior of the PFBG–LED Illumination 87
2.4.4. Thermo-Optical Behavior of the PFBG–SM LD Illumination 92
2.4.5. Thermo-Optical Behavior of the PFBG Associated with Other Light Sources 101
2.5. Concluding Remarks 102
References 102
Appendix 2.A: Solution Procedure to Obtain the Optical Power Along the PFBG 104
Appendix 2.B: Solution Procedure to Determine the Temperature Profile Along the PFBG 106
2.B.1. Solution Procedure of the Temperature Profile Along the PFBG with the LED 106
2.B.2. Solution Procedure of the Temperature Profile Along the PFBG with the SM LD 106

Chapter 3
Photorefractive Materials and Devices for Passive Components in WDM Systems
Claire Gu, Yisi Liu, Yuan Xu, J.J. Pan, Fengqing Zhou, Liang Dong and Henry He 111

3.1. Introduction 111
3.2. Tunable Flat-Topped Filter 114
3.2.1. Principle of Operation 114
3.2.2. Device Simulation 116
3.2.3. Design for Implementation 117
3.3. Wavelength Selective 2×2 Switch 117
3.3.1. Principle of Operation 118
3.3.2. Experimental Demonstration 119
3.3.3. Theoretical Analysis 121
3.3.4. Optimized Switch Design 123
3.3.5. Discussion 125
3.4. High Performance Dispersion Compensators 126
3.4.1. Multi-Channel Dispersion-Slope Compensator 126
3.4.2. High Precision FBG Fabrication Method and Dispersion Management Filters 129
3.5. Conclusions 133
References 133

Chapter 4
Thin Films for Microelectronics and Photonics: Physics, Mechanics, Characterization, and Reliability
David T. Read and Alex A. Volinsky 135

4.1. Terminology and Scope 135
4.1.1. Thin Films 135
4.1.2. Motivation 136
4.1.3. Chapter Outline 136
4.2. Thin Film Structures and Materials 137
4.2.1. Substrates 137
4.2.2. Epitaxial Films 137
4.2.3. Dielectric Films 140
4.2.4. Metal Films 141
4.2.5. Organic and Polymer Films 142
4.2.6. MEMS Structures 142
4.2.7. Intermediate Layers: Adhesion, Barrier, Buffer, and Seed Layers 142
4.3. Manufacturability/Reliability Challenges 143
4.3.1. Film Deposition and Stress 144
4.3.2. Grain Structure and Texture 147
4.3.3. Impurities 151
4.3.4. Dislocations 152
4.3.5. Electromigration and Voiding 153
4.3.6. Structural Considerations 155
4.3.7. Need for Mechanical Characterization 155
4.3.8. Properties of Interest 156
4.4. Methods for mechanical characterization of thin films 157
4.4.1. Microtensile Testing 157
4.4.2. Instrumented Indentation 159
4.4.3. Other Techniques 164
4.4.4. Adhesion Tests 165
4.5. Materials and Properties 172
4.5.1. Grain Size and Structure Size Effects 172
4.6. Properties of Specific Materials 173
4.7. Future Research 175
4.7.1. Techniques 175
4.7.2. Properties 175
4.7.3. Length Scale 175
References 176

Chapter 5
Carbon Nanotube Based Interconnect Technology: Opportunities and Challenges
Alan M. Cassell and Jun Li 181

5.1. Introduction: Physical Characteristics of Carbon Nanotubes 181
5.1.1. Structural 181
5.1.2. Electrical 182
5.1.3. Mechanical 185
5.1.4. Thermal 186
5.2. CNT Fabrication Technologies 186

5.2.1. Chemical Vapor Deposition of Carbon Nanotubes 187
5.2.2. Process Integration and Development 189
5.3. Carbon Nanotubes as Interconnects 191
5.3.1. Limitations of the Current Technology 191
5.3.2. Architecture, Geometry and Performance Potential Using Carbon Nanotubes 191
5.4. Design, Manufacture and Reliability 194
5.4.1. Microstructural Attributes and Effects on Electrical Characteristics 194
5.4.2. Interfacial Contact Materials 196
5.4.3. End-contacted Metal–CNT Junction 198
5.4.4. Thermal Stress Characteristics 198
5.4.5. Reliability Test 199
5.5. Summary 200
References 200

Chapter 6
Virtual Thermo-Mechanical Prototyping of Microelectronics and Microsystems
A. Wymysłowski, G.Q. Zhang, W.D. van Driel and L.J. Ernst 205
6.1. Introduction 205
6.2. Physical Aspects for Numerical Simulations 206
6.2.1. Numerical Modeling 208
6.2.2. Material Properties and Models 211
6.2.3. Thermo-Mechanical Related Failures 215
6.2.4. Designing for Reliability 219
6.3. Mathematical Aspects of Optimization 225
6.3.1. Design of Experiments 226
6.3.2. Response Surface Modeling 236
6.3.3. Advanced Approach to Virtual Prototyping 242
6.3.4. Designing for Quality 249
6.4. Application Case 252
6.4.1. Problem Description 252
6.4.2. Numerical Approach to QFN Package Design 253
6.5. Conclusion and Challenges 259
6.6. List of Acronyms 264
Acknowledgments 264
References 264

Materials Mechanics

Chapter 7
Fiber Optics Structural Mechanics and Nano-Technology Based New Generation of Fiber Coatings: Review and Extension
E. Suhir 269
7.1. Introduction 269
7.2. Fiber Optics Structural Mechanics 270
7.2.1. Review 270
7.3. New Nano-Particle Material (NPM) for Micro- and Opto-Electronic Applications 273
7.3.1. New Nano-Particle Material (NPM) 273
7.3.2. NPM-Based Optical Silica Fibers 274

7.4. Conclusions 277
Acknowledgment 277
References 277

Chapter 8
Area Array Technology for High Reliability Applications
Reza Ghaffarian 283

8.1. Introduction 283
8.2. Area Array Packages (AAPs) 284
8.2.1. Advantages of Area Array Packages 285
8.2.2. Disadvantages of Area Arrays 285
8.2.3. Area Array Types 286
8.3. Chip Scale Packages (CSPs) 286
8.4. Plastic Packages 288
8.4.1. Background 288
8.4.2. Plastic Area Array Packages 288
8.4.3. Plastic Package Assembly Reliability 289
8.4.4. Reliability Data for BGA, Flip Chip BGA, and CSP 291
8.5. Ceramic Packages 293
8.5.1. Background 293
8.5.2. Ceramic Package Assembly Reliability 294
8.5.3. Literature Survey on CBGA/CCGA Assembly Reliability 295
8.5.4. CBGA Thermal Cycle Test 297
8.5.5. Comparison of 560 I/O PBGA and CCGA assembly reliability 302
8.5.6. Designed Experiment for Assembly 305
8.6. Summary 309
8.7. List of Acronyms and Symbols 310
Acknowledgments 311
References 311

Chapter 9
Metallurgical Factors Behind the Reliability of High-Density Lead-Free Interconnections
Toni T. Mattila, Tomi T. Laurila and Jorma K. Kivilahti 313

9.1. Introduction 313
9.2. Approaches and Methods 315
9.2.1. The Four Steps of The Iterative Approach 315
9.2.2. The Role of Different Simulation Tools in Reliability Engineering 321
9.3. Interconnection Microstructures and Their Evolution 324
9.3.1. Solidification 324
9.3.2. Solidification Structure and the Effect of Contact Metalization Dissolution 325
9.3.3. Interfacial Reactions Products 330
9.3.4. Deformation Structures (Due to Slip and Twinning) 333
9.3.5. Recovery, Recrystallization and Grain Growth 335
9.4. Two Case Studies on Reliability Testing 335
9.4.1. Case 1: Reliability of Lead-Free CSPs in Thermal cycling 337
9.4.2. Case 2: Reliability of Lead-Free CSPs in Drop Testing 341
9.5. Summary 347

Acknowledgments 348
References 348

Chapter 10
Metallurgy, Processing and Reliability of Lead-Free Solder Joint Interconnections
Jin Liang, Nader Dariavach and Dongkai Shangguan 351

10.1. Introduction 351
10.2. Physical Metallurgy of Lead-Free Solder Alloys 352
10.2.1. Tin-Lead Solders 352
10.2.2. Lead-Free Solder Alloys 353
10.2.3. Interfacial Reaction: Wetting and Spreading 357
10.2.4. Interfacial Intermetallic Formation and Growth at Liquid–Solid Interfaces 363
10.3. Lead-Free Soldering Processes and Compatibility 377
10.3.1. Lead-Free Soldering Materials 378
10.3.2. PCB Substrates and Metalization Finishes 380
10.3.3. Lead-Free Soldering Processes 381
10.3.4. Components for Lead-Free Soldering 384
10.3.5. Design, Equipment and Cost Considerations 387
10.4. Reliability of Pb-Free Solder Interconnects 388
10.4.1. Reliability and Failure Distribution of Pb-Free Solder Joints 388
10.4.2. Effects of Loading and Thermal Conditions on Reliability of Solder Interconnection 389
10.4.3. Reliability of Pb-Free Solder Joints in Comparison to Sn-Pb Eutectic Solder Joints 395
10.5. Guidelines for Pb-free Soldering and Improvement in Reliability 406
References 406

Chapter 11
Fatigue Life Assessment for Lead-Free Solder Joints
Masaki Shiratori and Qiang Yu 411

11.1. Introduction 411
11.2. The Intermetallic Compound Formed at the Interface of the Solder Joints and the Cu-pad 412
11.3. Mechanical Fatigue Testing Equipment and Load Condition in the Lead Free Solder 413
11.4. Results of Mechanical Fatigue Test 414
11.5. Critical Fatigue Stress Limit for the Intermetallic Compound Layer 417
11.6. Influence of the Plating Material on the Fatigue Life of Sn-Zn (Sn-9Zn and Sn-8Zn-3Bi) Solder Joints 424
11.7. Conclusion 426
References 426

Chapter 12
Lead-Free Solder Materials: Design For Reliability
John H.L. Pang 429

12.1. Introduction 429
12.2. Mechanics of Solder Materials 430
12.2.1. Fatigue Behavior of Solder Materials 431
12.3. Design For Reliability (DFR) 433

12.4. Constitutive Models For Lead Free Solders 435
12.4.1. Tensile Test Results 435
12.4.2. Creep Test Results 440
12.5. Low Cycle Fatigue Models 443
12.6. FEA Modeling and Simulation 448
12.7. Reliability Test and Analysis 454
12.8. Conclusions 456
Acknowledgments 456
References 456

Chapter 13
Application of Moire Interferometry to Strain Analysis of PCB Deformations at Low Temperatures
Arkady Voloshin 459
13.1. Introduction 459
13.2. Optical Method and Recording of Fringe Patterns 460
13.2.1. Fractional Fringe Approach 461
13.2.2. Grating Frequency Increase 461
13.2.3. Creation of a High-Frequency Master Grating 462
13.2.4. Combination of the High Grating Frequency and Fractional Fringe Approach 463
13.3. Data Processing 463
13.4. Test Boards and Specimen Grating 463
13.5. Elevated Temperature Test 465
13.6. Low Temperature Test 468
13.7. Conclusions 470
Acknowledgment 472
References 473

Chapter 14
Characterization of Stresses and Strains in Microelectronics and Photonics Devices Using Photomechanics Methods
Bongtae Han 475
14.1. Introduction 475
14.2. Stress/Strain analysis 476
14.2.1. Moiré Interferometry 476
14.2.2. Extension: Microscopic Moiré Interferometry 477
14.2.3. Specimen Gratings 479
14.2.4. Strain Analysis 480
14.2.5. Thermal Deformation Measured at Room Temperature 481
14.2.6. Deformation as a Function of Temperature 485
14.2.7. Hygroscopic Deformation 494
14.2.8. Micromechanics 501
14.3. Warpage Analysis 505
14.3.1. Twyman/Green Interferometry 505
14.3.2. Shadow Moiré 509
14.3.3. Far Infrared Fizeau Interferometry 514
Acknowledgment 520
References 520

Chapter 15
Analysis of Reliability of IC Packages Using the Fracture Mechanics Approach
Andrew A.O. Tay 523

15.1. Introduction 523
15.2. Heat Transfer and Moisture Diffusion in IC Packages 525
15.3. Fundamentals of Interfacial Fracture Mechanics 527
15.4. Criterion for Crack Propagation 529
15.5. Interface Fracture Toughness 529
15.6. Total Stress Intensity Factor 530
15.7. Calculation of SERR and Mode Mixity 531
15.7.1. Crack Surface Displacement Extrapolation Method 531
15.7.2. Modified J-integral Method 532
15.7.3. Modified Virtual Crack Closure Method 533
15.7.4. Variable Order Boundary Element Method 536
15.7.5. Interaction Integral Method 536
15.8. Experimental Verification 538
15.9. Case Studies 542
15.9.1. Delamination Along Pad-Encapsulant Interface 542
15.9.2. Delamination Along Die-Attach/Pad Interface 544
15.9.3. Analysis Using Variable Order Boundary Element Method 546
15.10. Discussion of the Various Numerical Methods for Calculating G and ψ 549
15.11. Conclusion 551
References 551

Chapter 16
Dynamic Response of Micro- and Opto-Electronic Systems to Shocks and Vibrations: Review and Extension
E. Suhir 555

16.1. Introduction 555
16.2. Review 556
16.3. Extension: Quality of Shock Protection with a Flexible Wire Elements 557
16.4. Analysis 558
16.4.1. Pre-Buckling Mode: Small Displacements 558
16.4.2. Post-Buckling Mode: Large Displacements 564
16.5. Conclusions 567
References 568

Chapter 17
Dynamic Physical Reliability in Application to Photonic Materials
Dov Ingman, Tatiana Mirer and Ephraim Suhir 571

17.1. Introduction: Dynamic Reliability Approach to the Evolution of Silica Fiber Performance 571
17.1.1. Dynamic Physical Model of Damage Accumulation 572
17.1.2. Impact of the Three-Dimensional Mechanical-Temperature-Humidity Load on the Optical Fiber Reliability 575
17.1.3. Effect of Bimodality and Its Explanation Based on the Suggested Model 576
17.2. Reliability Improvement through NPM-Based Fiber Structures 585

17.2.1. Environmental Protection by NPM-Based Coating and Overall Self-Curing Effect of NPM Layers 585
17.2.2. Improvement in the Reliability Characteristics by Employing NPM Structures in Optical Fibers 587
17.3. Conclusions 593
References 593

Chapter 18
High-Speed Tensile Testing of Optical Fibers—New Understanding for Reliability Prediction
Sergey Semjonov and G. Scott Glaesemann 595
18.1. INTRODUCTION 595
18.2. Theory 596
18.2.1. Single-Region Power-Law Model 596
18.2.2. Two-Region Power-Law Model 598
18.2.3. Universal Static and Dynamic Fatigue Curves 599
18.3. Experimental 602
18.3.1. Sample Preparation 602
18.3.2. Dynamic Fatigue Tests 604
18.3.3. Static Fatigue Tests 605
18.4. Results and Discussion 606
18.4.1. High-Speed Testing 606
18.4.2. Static Fatigue 610
18.4.3. Influence of Multiregion Model on Lifetime Prediction 613
18.5. Conclusion 613
References 614
Appendix 18.A: High Speed Axial Strength Testing: Measurement Limits 616
Appendix 18.B: Incorporating Static Fatigue Results into Dynamic Fatigue Curves 620
18.B.1. Static Fatigue Test 620
18.B.2. Dynamic Fatigue Test 621
18.B.3. Discussion 622

Chapter 19
The Effect of Temperature on the Microstructure Nonlinear Dynamics Behavior
Xiaoling He 627
19.1. Introduction 627
19.2. Theoretical Development 630
19.2.1. Background on Nonlinear Dynamics and Nonlinear Thermo-Elasticity Theories 630
19.2.2. Nonlinear Thermo-Elasticity Development for an Isotropic Laminate Subject to Thermal and Mechanical and Load 631
19.3. Thin Laminate Deflection Response Subject to Thermal Effect and Mechanical Load 633
19.3.1. Steady State Temperature Effect 633
19.3.2. Transient Thermal Field Effect 638
19.4. Stress Field in Nonlinear Dynamics Response 653
19.4.1. Stress Field Formulation 653
19.4.2. Stress Distribution 654
19.4.3. Failure Analysis 654

19.5. Discussions 660
19.6. Summary 661
Nomenclature 662
Acknowledgment 663
References 663

Chapter 20
Effect of Material's Nonlinearity on the Mechanical Response of some Piezoelectric and Photonic Systems
Victor Birman and Ephraim Suhir 667

20.1. Introduction 667
20.2. Effect of Physical Nonlinearity on Vibrations of Piezoelectric Rods Driven by Alternating Electric Field 668
20.2.1. Physically Nonlinear Constitutive Relationships for an Orthotropic Cylindrical Piezoelectric Rod Subject to an Electric Field in the Axial Direction 670
20.2.2. Analysis of Uncoupled Axial Vibrations 673
20.2.3. Solution for Coupled Axial-Radial Axisymmetric Vibrations by the Generalized Galerkin Procedure 677
20.2.4. Numerical Results and Discussion 678
20.3. The Effect of the Nonlinear Stress–Strain Relationship on the Response of Optical Fibers 683
20.3.1. Stability of Optical Fibers 684
20.3.2. Stresses and Strains in a Lightwave Coupler Subjected to Tension 686
20.3.3. Free Vibrations 690
20.3.4. Bending of an Optical Fiber 692
20.4. Conclusions 695
Acknowledgment 696
References 697

Index 701

Volume II

List of Contributors xxvii

Preface xxxi

Physical Design

Chapter 1
Analytical Thermal Stress Modeling in Physical Design for Reliability of Micro- and Opto-Electronic Systems: Role, Attributes, Challenges, Results
E. Suhir 3

1.1. Thermal Loading and Thermal Stress Failures 3
1.2. Thermal Stress Modeling 4
1.3. Bi-Metal Thermostats and other Bi-Material Assemblies 5
1.4. Finite-Element Analysis 5

1.5. Die-Substrate and other Bi-Material Assemblies 6
1.6. Solder Joints 8
1.7. Design Recommendations 9
1.8. “Global” and “Local” Mismatch and Assemblies Bonded at the Ends 10
1.9. Assemblies with Low Modulus Adhesive Layer at the Ends 11
1.10. thermally Matched Assemblies 11
1.11. Thin Films 12
1.12. Polymeric Materials And Plastic Packages 13
1.13. Thermal Stress Induced Bowing and Bow-Free Assemblies 14
1.14. Probabilistic Approach 15
1.15. Optical Fibers and other Photonic Structures 15
1.16. Conclusion 16
References 17

Chapter 2
Probabilistic Physical Design of Fiber-Optic Structures
Satish Radhakrishnan, Ganesh Subbarayan and Luu Nguyen 23
2.1. Introduction 23
2.1.1. Demonstration Vehicle 24
2.2. Optical Model 25
2.2.1. Mode Field Diameter 26
2.2.2. Refraction and Reflection Losses 27
2.2.3. Calculations for Coupling Losses 27
2.2.4. Coupling Efficiency 28
2.3. Interactions in System and Identification of Critical Variables 30
2.3.1. Function Variable Incidence Matrix 30
2.3.2. Function Variable Incidence Matrix to Graph Conversion 31
2.3.3. Graph Partitioning Techniques 34
2.3.4. System Decomposition using Simulated Annealing 34
2.4. Deterministic Design Procedures 37
2.4.1. Optimal and Robust Design 40
2.4.2. A Brief Review of Multi-Objective Optimization 42
2.4.3. Implementation 43
2.4.4. Results 43
2.5. Stochastic Analysis 44
2.5.1. The First and Second Order Second Moment Methods 44
2.6. Probabilistic Design for Maximum Reliability 46
2.6.1. Results 49
2.7. Stochastic Characterization of Epoxy Behavior 51
2.7.1. Viscoelastic Models 52
2.7.2. Modeling the Creep Test 53
2.7.3. Dynamic Mechanical Analysis 54
2.7.4. Experimental Results 55
2.8. Analytical Model to Determine VCSEL Displacement 57
2.8.1. Results 63
2.9. Summary 67
References 67

Chapter 3
The Wirebonded Interconnect: A Mainstay for Electronics
Harry K. Charles, Jr. 71

3.1. Introduction 71
3.1.1. Integrated Circuit Revolution 71
3.1.2. Interconnection Types 72
3.1.3. Wirebond Importance 80
3.2. Wirebonding Basics 81
3.2.1. Thermocompression Bonding 81
3.2.2. Ultrasonic Bonding 83
3.2.3. Thermosonic Bonding 85
3.2.4. Wirebond Reliability 87
3.2.5. Wirebond Testing 89
3.2.6. Bonding Automation and Optimization 93
3.3. Materials 95
3.3.1. Bonding Wire 95
3.3.2. Bond Pad Metallurgy 100
3.3.3. Gold Plating 102
3.3.4. Pad Cleaning 104
3.4. Advanced Bonding Methods 105
3.4.1. Fine Pitch Bonding 105
3.4.2. Soft Substrates 108
3.4.3. Machine Improvements 110
3.4.4. Higher Frequency Wirebonding 110
3.4.5. Stud Bumping 115
3.5. Summary 116
Acknowledgments 116
References 116

Chapter 4
Metallurgical Interconnections for Extreme High and Low Temperature Environments
George G. Harman 121

4.1. Introduction 121
4.2. High Temperature Interconnections Requirements 122
4.2.1. Wire Bonding 122
4.2.2. The Use of Flip Chips in HTE 127
4.2.3. General Overview of Metallurgical Interfaces for Both HTE and LTE 129
4.3. Low Temperature Environment Interconnection Requirements 129
4.4. Corrosion and Other Problems in Both *HTE*, and *LTE* 130
4.5. The Potential Use of High Temperature Polymers in HTE 131
4.6. Conclusions 132
Acknowledgments 132
References 132

Chapter 5
Design, Process, and Reliability of Wafer Level Packaging
Zhuqing Zhang and C.P. Wong 135

5.1. Introduction 135

5.2. WLCSP 137
5.2.1. Thin Film Redistribution 137
5.2.2. Encapsulated Package 139
5.2.3. Compliant Interconnect 139
5.3. Wafer Level Underfill 141
5.3.1. Challenges of Wafer Level Underfill 142
5.3.2. Examples of Wafer Level Underfill Process 143
5.4. Comparison of Flip-Chip and WLCSP 145
5.5. Wafer Level Test and Burn-In 145
5.6. Summary 149
References 149

Chapter 6
Passive Alignment of Optical Fibers in V-grooves with Low Viscosity Epoxy Flow
S.W. Ricky Lee and C.C. Lo 151
6.1. Introduction 151
6.2. Design and Fabrication of Silicon Optical Bench with V-grooves 152
6.3. Issues of Conventional Passive Alignment Methods 158
6.3.1. V-grooves with Cover Plate 158
6.3.2. Edge Dispensing of Epoxy 161
6.4. Modified Passive Alignment Method 162
6.4.1. Working Principle 162
6.4.2. Alignment Mechanism 163
6.4.3. Design of Experiment 164
6.4.4. Experimental Procedures 164
6.4.5. Experimental Results 165
6.5. Effects of Epoxy Viscosity and Dispensing Volume 168
6.6. Application to Fiber Array Passive Alignment 170
6.7. Conclusions and Discussion 172
References 172

Reliability and Packaging

Chapter 7
Fundamentals of Reliability and Stress Testing
H. Anthony Chan 177
7.1. More Performance at Lower Cost in Shorter Time-to-market 178
7.1.1. Rapid Technological Developments 178
7.1.2. Integration of More Products into Human Life 178
7.1.3. Diverse Environmental Stresses 178
7.1.4. Competitive Market 179
7.1.5. Short Product Cycles 179
7.1.6. The Bottom Line 179
7.2. Measure of Reliability 180
7.2.1. Failure Rate 180
7.2.2. Systems with Multiple Independent Failure Modes 181
7.2.3. Failure Rate Distribution 182
7.3. Failure Mechanisms in Electronics and Packaging 184

7.3.1. Failure Mechanisms at Chip Level Include 184
7.3.2. Failure Mechanisms at Bonding Include 184
7.3.3. Failure Mechanisms in Device Packages Include 185
7.3.4. Failure Mechanisms in Epoxy Compounds Include 185
7.3.5. Failure Mechanisms at Shelf Level Include 185
7.3.6. Failure Mechanisms in Material Handling Include 185
7.3.7. Failure Mechanisms in Fiber Optics Include 185
7.3.8. Failure Mechanisms in Flat Panel Displays Include 186
7.4. Reliability Programs and Strategies 186
7.5. Product Weaknesses and Stress Testing 187
7.5.1. Why do Products Fail? 187
7.5.2. Stress Testing Principle 189
7.6. Stress Testing Formulation 191
7.6.1. Threshold and Cumulative Stress Failures 191
7.6.2. Stress Stimuli and Flaws 192
7.6.3. Modes of Stress Testing 193
7.6.4. Lifetime Failure Fraction 194
7.6.5. Robustness Against Maximum Service Life Stress 195
7.6.6. Stress–Strength Contour 197
7.6.7. Common Issues 198
7.7. Further Reading 201

Chapter 8
How to Make a Device into a Product: *Accelerated Life Testing (ALT), Its Role, Attributes, Challenges, Pitfalls, and Interaction with Qualification Tests*
E. Suhir 203

8.1. Introduction 203
8.2. Some Major Definitions 204
8.3. Engineering Reliability 204
8.4. Field Failures 205
8.5. Reliability is a Complex Property 206
8.6. Three Major Classes of Engineering Products and Market Demands 206
8.7. Reliability, Cost and Time-to-Market 208
8.8. Reliability Costs Money 208
8.9. Reliability Should Be Taken Care of on a Permanent Basis 209
8.10. Ways to Prevent and Accommodate Failures 210
8.11. Redundancy 211
8.12. Maintenance and Warranty 211
8.13. Test Types 212
8.14. Accelerated Tests 212
8.15. Accelerated Test Levels 213
8.16. Qualification Standards 213
8.17. Accelerated Life Tests (ALTs) 214
8.18. Accelerated Test Conditions 215
8.19. Acceleration Factor 216
8.20. Accelerated Stress Categories 217
8.21. Accelerated Life Tests (ALTs) and Highly Accelerated Life Tests (HALTs) 218
8.22. Failure Mechanisms and Accelerated Stresses 219

8.23. ALTs: Pitfalls and Challenges 219
8.24. Burn-ins 220
8.25. Wear-Out Failures 221
8.26. Non-Destructive Evaluations (NDE's) 222
8.27. Predictive Modeling 222
8.28. Some Accelerated Life Test (ALT) Models 223
8.28.1. Power Law 224
8.28.2. Boltzmann-Arrhenius Equation 224
8.28.3. Coffin-Manson Equation (Inverse Power Law) 225
8.28.4. Paris-Erdogan Equation 226
8.28.5. Bueche-Zhurkov Equation 227
8.28.6. Eyring Equation 227
8.28.7. Peck and Black Equations 227
8.28.8. Fatigue Damage Model (Miner's Rule) 228
8.28.9. Creep Rate Equations 228
8.28.10. Weakest Link Models 228
8.28.11. Stress–Strength Models 229
8.29. Probability of Failure 229
8.30. Conclusions 230
References 230

Chapter 9
Micro-Deformation Analysis and Reliability Estimation of Micro-Components by Means of NanoDAC Technique
Bernd Michel and Jürgen Keller 233

9.1. Introduction 233
9.2. Basics of Digital Image Correlation 234
9.2.1. Cross Correlation Algorithms on Gray Scale Images 234
9.2.2. Subpixel Analysis for Enhanced Resolution 236
9.2.3. Results of Digital Image Correlation 238
9.3. Displacement and Strain Measurements on SFM Images 239
9.3.1. Digital Image Correlation under SPM Conditions 239
9.3.2. Technical Requirements for the Application of the Correlation Technique 241
9.4. Deformation Analysis on Thermally and Mechanically Loaded Objects under the SFM 241
9.4.1. Reliability Aspects of Sensors and Micro Electro-Mechanical Systems (MEMS) 241
9.4.2. Thermally Loaded Gas Sensor under SFM 242
9.4.3. Crack Detection and Evaluation by SFM 243
9.5. Conclusion and Outlook 250
References 250

Chapter 10
Interconnect Reliability Considerations in Portable Consumer Electronic Products
Sridhar Canumalla and Puligandla Viswanadham 253

10.1. Introduction 253
10.2. Reliability—Thermal, Mechanical and Electrochemical 255
10.2.1. Accelerated Life Testing 255
10.2.2. Thermal Environment 257

10.2.3. Mechanical Environment 257
10.2.4. Electrochemical Environment 264
10.2.5. Tin Whiskers 267
10.3. Reliability Comparisons in Literature 267
10.3.1. Thermomechanical Reliability 268
10.3.2. Mechanical Reliability 270
10.4. Influence of Material Properties on Reliability 271
10.4.1. Printed Wiring Board 271
10.4.2. Package 272
10.4.3. Surface Finish 272
10.5. Failure Mechanisms 273
10.5.1. Thermal Environment 273
10.5.2. Mechanical Environment 276
10.5.3. Electrochemical Environment 286
10.6. reliability test Practices 291
10.7. Summary 294
Acknowledgments 295
References 295

Chapter 11
MEMS Packaging and Reliability
Y.C. Lee 299
11.1. Introduction 299
11.2. Flip-Chip Assembly for Hybrid Integration 304
11.3. Soldered Assembly for Three-Dimensional MEMS 309
11.4. Flexible Circuit Boards for MEMS 313
11.5. Atomic Layer Deposition for Reliable MEMS 316
11.6. Conclusions 320
Acknowledgments 320
References 320

Chapter 12
Advances in Optoelectronic Methodology for MOEMS Testing
Ryszard J. Pryputniewicz 323
12.1. Introduction 323
12.2. MOEMS Samples 324
12.3. Analysis 328
12.4. Optoelectronic Methodology 330
12.5. Representative Applications 334
12.6. Conclusions and Recommendations 338
Acknowledgments 339
References 339

Chapter 13
Durability of Optical Nanostructures: Laser Diode Structures and Packages, A Case Study
Ajay P. Malshe and Jay Narayan 341

13.1. High Efficiency Quantum Confined (Nanostructured) III-Nitride Based Light Emitting Diodes And Lasers 342
13.1.1. Introduction 342
13.2. Investigation of Reliability Issues in High Power Laser Diode Bar Packages 348
13.2.1. Introduction 348
13.2.2. Preparation of Packaged Samples for Reliability Testing 349
13.2.3. Finding and Model of Reliability Results 350
13.3. Conclusions 357
Acknowledgments 358
References 358

Chapter 14
Review of the Technology and Reliability Issues Arising as Optical Interconnects Migrate onto the Circuit Board
P. Misselbrook, D. Gwyer, C. Bailey, D. Gwyer, C. Bailey, P.P. Conway and K. Williams 361
14.1. Background to Optical Interconnects 362
14.2. Transmission Equipment for Optical Interconnects 362
14.3. Very Short Reach Optical Interconnects 365
14.4. Free Space USR Optical Interconnects 366
14.5. Guided Wave USR Interconnects 367
14.6. Component Assembly of OECB's 370
14.7. Computational Modeling of Optical Interconnects 373
14.8. Conclusions 380
Acknowledgments 380
References 381

Chapter 15
Adhesives for Micro- and Opto-Electronics Application: Chemistry, Reliability and Mechanics
D.W. Dahringer 383
15.1. Introduction 383
15.1.1. Use of Adhesives in Micro and Opto-Electronic Assemblies 383
15.1.2. Specific Applications 384
15.2. Adhesive Characteristics 385
15.2.1. General Properties of Adhesives 385
15.2.2. Adhesive Chemistry 390
15.3. Design Objective 393
15.3.1. Adhesive Joint Design 393
15.3.2. Manufacturing Issues 397
15.4. Failure Mechanism 401
15.4.1. General 401
15.4.2. Adhesive Changes 401
15.4.3. Interfacial Changes 401
15.4.4. Interfacial Stress 401
15.4.5. External Stress 402
References 402

Chapter 16
Multi-Stage Peel Tests and Evaluation of Interfacial Adhesion Strength for Micro- and Opto-Electronic Materials
Masaki Omiya, Kikuo Kishimoto and Wei Yang 403

16.1. Introduction 403
16.2. Multi-Stage Peel Test (MPT) 407
16.2.1. Testing Setup 407
16.2.2. Multi-Stage Peel Test 408
16.2.3. Energy Variation in Steady State Peeling 409
16.3. Interfacial Adhesion Strength of Copper Thin Film 413
16.3.1. Preparation of Specimen 413
16.3.2. Measurement of Adhesion Strength by the MPT 414
16.3.3. Discussions 415
16.4. UV-Irradiation Effect on Ceramic/Polymer Interfacial Strength 419
16.4.1. Preparation of PET/ITO Specimen 419
16.4.2. Measurement of Interfacial Strength by MPT 422
16.4.3. Surface Crack Formation on ITO Layer under Tensile Loading 424
16.5. Concluding Remarks 426
Acknowledgment 427
References 427

Chapter 17
The Effect of Moisture on the Adhesion and Fracture of Interfaces in Microelectronic Packaging
Timothy P. Ferguson and Jianmin Qu 431

17.1. Introduction 432
17.2. Moisture Transport Behavior 433
17.2.1. Background 433
17.2.2. Diffusion Theory 434
17.2.3. Underfill Moisture Absorption Characteristics 435
17.2.4. Moisture Absorption Modeling 438
17.3. Elastic Modulus Variation Due to Moisture Absorption 442
17.3.1. Background 442
17.3.2. Effect of Moisture Preconditioning 444
17.3.3. Elastic Modulus Recovery from Moisture Uptake 447
17.4. Effect of Moisture on Interfacial Adhesion 449
17.4.1. Background 449
17.4.2. Interfacial Fracture Testing 451
17.4.3. Effect of Moisture Preconditioning on Adhesion 452
17.4.4. Interfacial Fracture Toughness Recovery from Moisture Uptake 461
17.4.5. Interfacial Fracture Toughness Moisture Degradation Model 462
References 469

Chapter 18
Highly Compliant Bonding Material for Micro- and Opto-Electronic Applications
E. Suhir and D. Ingman 473

18.1. Introduction 473

18.2. Effect of the Interfacial Compliance on the interfacial Shearing Stress 474
18.3. Internal Compressive Forces 476
18.4. Advanced Nano-Particle Material (NPM) 476
18.5. Highly-Compliant Nano-Systems 478
18.6. Conclusions 479
References 480
Appendix 18.A: Bimaterial Assembly Subjected to an External Shearing Load and Change in Temperature: Expected Stress Relief due to the Elevated Interfacial Compliance 480
Appendix 18.B: Cantilever Wire ("Beam") Subjected at its Free End to a Lateral (Bending) and an Axial (Compressive) Force 483
Appendix 18.C: Compressive Forces in the NPM-Based Compound Structure 485

Chapter 19
Adhesive Bonding of Passive Optical Components
Anne-Claire Pliska and Christian Bosshard 487

19.1. Introduction 487
19.2. Optical Devices and Assemblies 489
19.2.1. Optical Components 489
19.2.2. Opto-electronics Assemblies: Specific Requirements 489
19.3. Adhesive Bonding in Optical Assemblies 503
19.3.1. Origin of Adhesion 503
19.3.2. Adhesive Selection and Dispensing 508
19.3.3. Dispensing Technologies 515
19.4. Some Applications 518
19.4.1. Laser to Fiber Assembly 518
19.4.2. Planar Lightwave Circuit (PLC) Pigtailing 520
19.5. Summary and Recommendations 522
Acknowledgments 523
References 523

Chapter 20
Electrically Conductive Adhesives: A Research Status Review
James E. Morris and Johan Liu 527

20.1. Introduction 527
20.1.1. Technology Drivers 527
20.1.2. Isotropic Conductive Adhesives (ICAs) 529
20.1.3. Anisotropic Conductive Adhesives (ACAs) 529
20.1.4. Non-Conductive Adhesive (NCA) 529
20.2. Structure 529
20.2.1. ICA 529
20.2.2. ACA 532
20.2.3. Modeling 534
20.3. Materials and Processing 534
20.3.1. Polymers 534
20.3.2. ICA Filler 536
20.3.3. ACA Processing 536
20.4. Electrical Properties 538

20.4.1. ICA 538
20.4.2. Electrical Measurements 544
20.4.3. ACA 544
20.5. Mechanical Properties 546
20.5.1. ICA 546
20.5.2. ACA 547
20.6. Thermal Properties 553
20.6.1. Thermal Characteristics 553
20.6.2. Maximum Current Carrying Capacity 553
20.7. Reliability 554
20.7.1. ICA 554
20.7.2. ACA 557
20.7.3. General Comments 565
20.8. Environmental Impact 565
20.9. Further Study 565
References 565

Chapter 21
Electrically Conductive Adhesives
Johann Nicolics and Martin Mündlein 571
21.1. Introduction and Historical Background 571
21.2. Contact Formation 574
21.2.1. Percolation and Critical Filler Content 574
21.2.2. ICA Contact Model 575
21.2.3. Results 578
21.3. Aging Behavior and Quality Assessment 595
21.3.1. Introduction 595
21.3.2. Material Selection and Experimental Parameters 595
21.3.3. Curing Parameters and Definition of Curing Time 597
21.3.4. Testing Conditions, Typical Results, and Conclusions 598
21.4. About Typical Applications 602
21.4.1. ICA for Attachment of Power Devices 602
21.4.2. ICA for Interconnecting Parts with Dissimilar Thermal Expansion Coefficient 604
21.4.3. ICA for Cost-Effective Assembling of Multichip Modules 606
21.5. Summary 607
Notations and Definitions 607
References 608

Chapter 22
Recent Advances of Conductive Adhesives: A Lead-Free Alternative in Electronic Packaging
Grace Y. Li and C.P. Wong 611
22.1. Introduction 611
22.2. Isotropic Conductive Adhesives (ICAs) 613
22.2.1. Improvement of Electrical Conductivity of ICAs 614
22.2.2. Stabilization of Contact Resistance on Non-Noble Metal Finishes 615
22.2.3. Silver Migration Control of ICA 618

22.2.4. Improvement of Reliability in Thermal Shock Environment 618
22.2.5. Improvement of Impact Performance of ICA 619
22.3. Anisotropic Conductive Adhesives (ACAs)/Anisotropic Conductive Film (ACF) 619
22.3.1. Materials 620
22.3.2. Application of ACA/ACF in Flip Chip 621
22.3.3. Improvement of Electrical Properties of ACAs 621
22.3.4. Thermal Conductivity of ACA 623
22.4. Future Advances of ECAs 623
22.4.1. Electrical Characteristics 623
22.4.2. High Frequency Compatibility 623
22.4.3. Reliability 623
22.4.4. ECAs with Nano-filler for Wafer Level Application 625
References 625

Chapter 23
Die Attach Quality Testing by Structure Function Evaluation
Márta Rencz, Vladimir Székely and Bernard Courtois 629

Nomenclature 629
Greek symbols 629
Subscripts 630
23.1. Introduction 630
23.2. Theoretical Background 630
23.3. Detecting Voids in the Die Attach of Single Die Packages 634
23.4. Simulation Experiments for Locating the Die Attach Failure on Stacked Die Packages 636
23.4.1. Simulation Tests Considering Stacked Dies of the Same Size 637
23.4.2. Simulation Experiments on a Pyramidal Structure 639
23.5. Verification of the Methodology by Measurements 642
23.5.1. Comparison of the Transient Behavior of Stacked Die Packages Containing Test Dies, Prior Subjected to Accelerated Moisture and Temperature Testing 642
23.5.2. Comparison of the Transient Behavior of Stacked Die Packages Containing Real Functional Dies, Subjected Prior to Accelerated Moisture and Temperature Testing 644
23.6. Conclusions 649
Acknowledgments 649
References 650

Chapter 24
Mechanical Behavior of Flip Chip Packages under Thermal Loading
Enboa Wu, Shoulung Chen, C.Z. Tsai and Nicholas Kao 651

24.1. Introduction 651
24.2. Flip Chip Packages 652
24.3. Measurement Methods 654
24.3.1. Phase Shifted Shadow Moiré Method 654
24.3.2. Electronic Speckle Pattern Interferometry (ESPI) Method 655
24.4. Substrate CTE Measurement 656
24.5. Behavior of Flip Chip Packages under Thermal Loading 661
24.5.1. Warpage at Room Temperature 661

24.5.2. Warpage at Elevated Temperatures 662
24.5.3. Effect of Underfill on Warpage 666
24.6. Finite Element Analysis of Flip Chip Packages under Thermal Loading 668
24.7. Parametric Study of Warpage for Flip Chip Packages 669
24.7.1. Change of the Chip Thickness 670
24.7.2. Change of the Substrate Thickness 670
24.7.3. Change of the Young's Modulus of the Underfill 671
24.7.4. Change of the CTE of the Underfill 672
24.7.5. Effect of the Geometry of the Underfill Fillet 672
24.8. Summary 674
References 674

Chapter 25
Stress Analysis for Processed Silicon Wafers and Packaged Micro-devices
Li Li, Yifan Guo and Dawei Zheng 677
25.1. Intrinsic Stress Due to Semiconductor Wafer Processing 677
25.1.1. Testing Device Structure 678
25.1.2. Membrane Deformations 679
25.1.3. Intrinsic Stress 681
25.1.4. Intrinsic Stress in Processed Wafer: Summary 683
25.2. Die Stress Result from Flip-chip Assembly 685
25.2.1. Consistent Composite Plate Model 685
25.2.2. Free Thermal Deformation 687
25.2.3. Bimaterial Plate (BMP) Case 688
25.2.4. Validation of the Bimaterial Model 691
25.2.5. Flip-Chip Package Design 695
25.2.6. Die Stress in Flip Chip Assembly: Summary 697
25.3. Thermal Stress Due to Temperature Cycling 698
25.3.1. Finite Element Analysis 698
25.3.2. Constitutive Equation for Solder 699
25.3.3. Time-Dependent Thermal Stresses of Solder Joint 700
25.3.4. Solder Joint Reliability Estimation 701
25.3.5. Thermal Stress Due to Temperature Cycling: Summary 703
25.4. Residual Stress in Polymer-based Low Dielectric Constant (low-k) Materials 703
References 708

Index 711

List of Contributors

VOLUME I

Avram Bar-Cohen
University of Maryland
College Park, Maryland, USA

Victor Birman
University of Missouri-Rolla
St. Louis, Missouri, USA

H.J.L. Bressers
Philips Semiconductors
Nijmegen, The Netherlands

Alan M. Cassell
NASA Ames Research Center
Moffett Field, California, USA

N. Dariavach
EMC Corp
Hopkinton, Massachusetts, USA

Liang Dong
Lightwaves 2020 Inc.
Milpitas, California, USA

L.J. Ernst
Delft University of Technology
Delft, The Netherlands

Reza Ghaffarian
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California, USA

G. Scott Glaesemann
Corning Incorporated
Corning, New York, USA

Claire Gu
University of California, Santa Cruz
Santa Cruz, California, USA

Bongtae Han
University of Maryland
College Park, Maryland, USA

Henry He
Lightwaves 2020 Inc.
Milpitas, California, USA

Xiaoling He
University of Wisconsin
Milwaukee, Wisconsin, USA

Dov Ingman
Technion, Israel Institute of Technology
Haifa, Israel

K.M.B. Jansen
Delft University of Technology
Delft, The Netherlands

J.H.J. Janssen
Philips Semiconductors
Nijmegen, The Netherlands

Kyoung Joon Kim
University of Maryland
College Park, Maryland, USA

Jorma K. Kivilahti
Helsinki University of Technology
Helsinki, Finland

Tomi T. Laurila
Helsinki University of Technology
Helsinki, Finland

J. Liang
EMC Corp
Hopkinton, Massachusetts, USA

Yisi Liu
University of California, Santa Cruz
Santa Cruz, California, USA

Jun Li
NASA Ames Research Center
Moffett Field, California, USA

Toni T. Mattila
Helsinki University of Technology
Helsinki, Finland

Tatiana Mirer
Technion, Israel Institute of Technology
Haifa, Israel

J.J. Pan
Lightwaves 2020 Inc.
Milpitas, California, USA

John H.L. Pang
Nanyang Technological University
Nanyang, Singapore

David T. Read
National Institute of Standards and Technology
Boulder, Colorado, USA

Sergey Semjonov
Fiber Optics Research Center
Moscow, Russia

D. Shangguan
FLEXTRONICS
San Jose, California, USA

Masaki Shiratori
Yokohama National University
Yokohama, Japan

Andrew A.O. Tay
National University of Singapore
Republic of Singapore

W.D. van Driel
Delft University of Technology
Delft, The Netherlands

C. van't Hof
Delft University of Technology
Delft, The Netherlands

Alex A. Volinsky
University of South Florida
Tampa, Florida, USA

Arkady Voloshin
Lehigh University
Bethlehem, Pennsylvania, USA

A. Wymyslowski
Wroclaw University of Technology
Wroclaw, Poland

Yuan Xu
University of California, Santa Cruz
Santa Cruz, California, USA

D.G. Yang
Delft University of Technology
Delft, The Netherlands

Qiang Yu
Yokohama National University
Yokohama, Japan

G.Q. Zhang
Delft University of Technology
Delft, The Netherlands and
Philips Semiconductors
Eindhoven, The Netherlands

Fengqing Zhou
Lightwaves 2020 Inc.
Milpitas, California, USA

VOLUME II

C. Bailey
University of Greenwich
London, United Kingdom

Christian Bosshard
CSEM SA, Untere Grundlistrasse 1
Alpnach Dorf, Switzerland

Sridhar Canumalla
Nokia
Irving, Texas, USA

H. Anthony Chan
University of Cape Town
Rondebosch, South Africa

Harry K. Charles, Jr.
The Johns Hopkins University
Laurel, Maryland, USA

Shoulung Chen
National Taiwan University
Taiwan

Bernard Courtois
TIMA-CMP
Grenoble Cedex, France

D.W. Dahringer
D.W. Dahringer Consultants
Glen Ridge, New Jersey, USA

Timothy P. Ferguson
Southern Research Institute
Birmingham, Alabama, USA

Yifan Guo
Skyworks Solutions, Inc.
Irvine, California, USA

D. Gwyer
University of Greenwich
London, United Kingdom

George G. Harman
National Institute of Standards and Technology
Gaithersburg, Maryland, USA

D. Ingman
Technion, Israel Institute of Technology
Haifa, Israel

Nicholas Kao
National Taiwan University
Taiwan

J. Keller
Fraunhofer Institute for Reliability and Micro Integration (IZM)
Berlin, Germany

Kikuo Kishimoto
Tokyo Institute of Technology
Tokyo, Japan

S.W.R. Lee
Hong Kong University of Science and Technology
Clear Water Bay, Kowloon, Hong Kong

Grace Y. Li
Georgia Institute of Technology
Atlanta, Georgia, USA

Li Li
Cisco Systems, Inc.
San Jose, California, USA

Johan Liu
Chalmers University of Technology
Goteborg, Sweden

C.C. Lo
Hong Kong University of Science and Technology
Clear Water Bay, Kowloon, Hong Kong

Ajay P. Malshe
University of Arkansas
Fayetteville, Arkansas, USA

Bernd Michel
Fraunhofer MicroMaterials Center
Berlin, Germany

P. Misselbrook
Celestica
Kidsgrove, Stoke-on-Trent, United Kingdom
and
University of Greenwich
London, United Kingdom

James E. Morris
Portland State University
Portland, Oregon, USA

Martin Mündlein
Vienna Institute of Technology
Vienna, Austria

Jay Narayan
North Carolina State University
Raleigh, North Carolina, USA

Johann Nicolics
Vienna Institute of Technology
Vienna, Austria

Luu Nguyen
National Semiconductor Corporation
Santa Clara, California, USA

Masaki Omiya
Tokyo Institute of Technology
Tokyo, Japan

Anne-Claire Pliska
CSEM SA, Untere Grundlistrasse 1
Alpnach Dorf, Switzerland

Jianmin Qu
Georgia Institute of Technology
Atlanta, Georgia, USA

Satish Radhakrishanan
Purdue University
West Lafayette, Indiana, USA

Marta Rencz
MicReD Ltd.
Budapest, Hungary

Ganesh Subbarayan
Purdue University
West Lafayette, Indiana, USA

Vladimir Szekely
Budapest University of Technology and Economics
Budapest, Hungary

C.Z. Tsai
National Taiwan University
Taiwan

Puligandla Viswanadham
Nokia Research Center
Irving, Texas, USA

K. Williams
Loughborough University
Loughborough, United Kingdom

Enboa Wu
National Taiwan University
Taipei, Taiwan

Wei Yang
Tsinghua University
Beijing, P.R. China

Z. Zhang
Georgia Institute of Technology
Atlanta, Georgia, USA

Dawei Zheng
Kotura, Inc.
Monterey Park, California, USA

Preface

This book encompasses a broad area of micro- and opto-electronic engineering materials: their physics, mechanics, reliability, and packaging, with an emphasis on physical design issues and problems. The editors tried to bring in the most eminent engineers and scientists as chapter authors and put together the most comprehensive book ever written on the subjects of materials, mechanics, physics, packaging, functional performance, mechanical reliability, environmental durability and other aspects of reliability of micro- and opto-electronic assemblies, components, devices, and systems. University professors and leading industrial engineers contributed to the book. The contents of the book reflect the state-of-the-art in the above listed fields of applied science and engineering.

The intended audience are all those who work in micro- and opto-electronics, and photonics; electronic and optical materials; applied and industrial physics; mechanical and reliability engineering; electron and optical devices and systems. The expected and targeted readers are practitioners and professionals, scientists and researchers, lecturers and continuing education course directors, graduate and undergraduate students, technical supervisors and entrepreneurs. The book can serve, to a great extent, as an encyclopedia in the field of physics and mechanics of micro- and opto-electronic materials and structures. In the editors' opinion, it can serve also as a textbook, as a reference book, and as a guidance for self- and continuing education, i.e., as a source of comprehensive and in-depth information in its areas. The book's chapters contain both the description of the state-of-the-art in a particular field, as well as new results obtained by the chapter authors and their colleagues.

We would like to point out that many methods and approaches addressed in this book extend far beyond microelectronics and photonics. Although these methods and approaches were developed, advanced and reported primarily in application to micro- and opto-electronic systems, they are applicable also in many related areas of engineering and physics.

The editors are proud of the broad scope of the book, and of the quality of the contributed chapters, and would like to take this opportunity to deeply acknowledge, with thanks, the conscientious effort of the numerous contributors.

February 2006
E. Suhir
C.-P. Wong
Y.-C. Lee

Micro- and Opto-Electronic Materials and Structures: Physics, Mechanics, Design, Reliability, Packaging

Volume II
Physics Design—Reliability and Packaging

Micro- and Opto-Electronic Materials and Structures: Physics, Mechanics, Design, Reliability, Packaging

Volume II
Physics Design—Reliability and Packaging

Edited by:

E. Suhir
University of California, Santa Cruz
Santa Cruz, California, USA

University of Maryland
College Park, Maryland, USA

Y.C. Lee
University of Colorado
Boulder, Colorado, USA

C.P. Wong
Georgia Tech
Atlanta, Georgia, USA

Springer

E. Suhir
University of California, Santa Cruz
Santa Cruz, California
and
University of Maryland
College Park, Maryland

Y.C. Lee
University of Colorado
Boulder, Colorado

C.P. Wong
Georgia Institute of Technology
Atlanta, Georgia

Micro- and Opto-Electronic Materials and Structures:
Physics, Mechanics, Design, Reliability, Packaging

Library of Congress Control Number: 2006922729

ISBN 0-387-27974-1 e-ISBN 0-387-32989-7
ISBN 978-0-387-27974-9

Printed on acid-free paper.

© 2007 Springer Science+Business Media, Inc.
All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Springer Science + Business Media, Inc., 233 Spring Street, New York, NY 10013, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now know or hereafter developed is forbidden.
The use in this publication of trade names, trademarks, service marks and similar terms, even if the are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

Printed in the United States of America.

9 8 7 6 5 4 3 2 1 SPIN 11055464

springer.com

Contents

Volume I

List of Contributors xxvii

Preface xxxi

Materials Physics

Chapter 1
Polymer Materials Characterization, Modeling and Application
L.J. Ernst, K.M.B. Jansen, D.G. Yang, C. van 't Hof, H.J.L. Bressers, J.H.J. Janssen and G.Q. Zhang 3

1.1. Introduction 3
1.2. Polymers in Microelectronics 4
1.3. Basics of Visco-Elastic Modeling 6
1.3.1. Preliminary: State Dependent Viscoelasticity 6
1.3.2. Incremental Relationship 10
1.3.3. Linear State Dependent Viscoelasticity 13
1.3.4. Isotropic Material Behavior 14
1.3.5. Interrelations between Property Functions 15
1.3.6. Elastic Approximations 17
1.4. Linear Visco-Elastic Modeling (Fully Cured Polymers) 18
1.4.1. Introduction 18
1.4.2. Static Testing of Relaxation Moduli 18
1.4.3. Time-Temperature Superposition Principle 23
1.4.4. Static Testing of Creep Compliances 24
1.4.5. Dynamic Testing 27
1.5. Modeling of Curing Polymers 34
1.5.1. "Partly State Dependent" Modeling (Curing Polymers) 35
1.5.2. "Fully State Dependent" Modeling (Curing Polymers) 49
1.6. Parameterized Polymer Modeling (PPM) 53
1.6.1. PPM Hypotheses 54
1.6.2. Experimental Characterizations 55
1.6.3. PPM Modeling in Virtual Prototyping 62
Acknowledgments 62
References 62

Chapter 2
Thermo-Optic Effects in Polymer Bragg Gratings
Avram Bar-Cohen, Bongtae Han and Kyoung Joon Kim 65

2.1. Introduction 65
2.2. Fundamentals of Bragg Gratings 67
2.2.1. Physical Descriptions 67
2.2.2. Basic Optical Principles 68
2.3. Thermo-Optical Modeling of Polymer Fiber Bragg Grating 70
2.3.1. Heat Generation by Intrinsic Absorption 70
2.3.2. Analytical Thermal Model of PFBG 78
2.3.3. FEA Thermal Model of PFBG 80
2.3.4. Thermo-Optical Model of PFBG 80
2.4. Thermo-Optical Behavior of PMMA-Based PFBG 84
2.4.1. Description of a PMMA-Based PFBG and Light Sources 85
2.4.2. Power Variation Along the PFBG 86
2.4.3. Thermo-Optical Behavior of the PFBG–LED Illumination 87
2.4.4. Thermo-Optical Behavior of the PFBG–SM LD Illumination 92
2.4.5. Thermo-Optical Behavior of the PFBG Associated with Other Light Sources 101
2.5. Concluding Remarks 102
References 102
Appendix 2.A: Solution Procedure to Obtain the Optical Power Along the PFBG 104
Appendix 2.B: Solution Procedure to Determine the Temperature Profile Along the PFBG 106
2.B.1. Solution Procedure of the Temperature Profile Along the PFBG with the LED 106
2.B.2. Solution Procedure of the Temperature Profile Along the PFBG with the SM LD 106

Chapter 3
Photorefractive Materials and Devices for Passive Components in WDM Systems
Claire Gu, Yisi Liu, Yuan Xu, J.J. Pan, Fengqing Zhou, Liang Dong and Henry He 111

3.1. Introduction 111
3.2. Tunable Flat-Topped Filter 114
3.2.1. Principle of Operation 114
3.2.2. Device Simulation 116
3.2.3. Design for Implementation 117
3.3. Wavelength Selective 2×2 Switch 117
3.3.1. Principle of Operation 118
3.3.2. Experimental Demonstration 119
3.3.3. Theoretical Analysis 121
3.3.4. Optimized Switch Design 123
3.3.5. Discussion 125
3.4. High Performance Dispersion Compensators 126
3.4.1. Multi-Channel Dispersion-Slope Compensator 126
3.4.2. High Precision FBG Fabrication Method and Dispersion Management Filters 129
3.5. Conclusions 133
References 133

Chapter 4
Thin Films for Microelectronics and Photonics: Physics, Mechanics, Characterization, and Reliability
David T. Read and Alex A. Volinsky 135

4.1. Terminology and Scope 135
4.1.1. Thin Films 135
4.1.2. Motivation 136
4.1.3. Chapter Outline 136
4.2. Thin Film Structures and Materials 137
4.2.1. Substrates 137
4.2.2. Epitaxial Films 137
4.2.3. Dielectric Films 140
4.2.4. Metal Films 141
4.2.5. Organic and Polymer Films 142
4.2.6. MEMS Structures 142
4.2.7. Intermediate Layers: Adhesion, Barrier, Buffer, and Seed Layers 142
4.3. Manufacturability/Reliability Challenges 143
4.3.1. Film Deposition and Stress 144
4.3.2. Grain Structure and Texture 147
4.3.3. Impurities 151
4.3.4. Dislocations 152
4.3.5. Electromigration and Voiding 153
4.3.6. Structural Considerations 155
4.3.7. Need for Mechanical Characterization 155
4.3.8. Properties of Interest 156
4.4. Methods for mechanical characterization of thin films 157
4.4.1. Microtensile Testing 157
4.4.2. Instrumented Indentation 159
4.4.3. Other Techniques 164
4.4.4. Adhesion Tests 165
4.5. Materials and Properties 172
4.5.1. Grain Size and Structure Size Effects 172
4.6. Properties of Specific Materials 173
4.7. Future Research 175
4.7.1. Techniques 175
4.7.2. Properties 175
4.7.3. Length Scale 175
References 176

Chapter 5
Carbon Nanotube Based Interconnect Technology: Opportunities and Challenges
Alan M. Cassell and Jun Li 181

5.1. Introduction: Physical Characteristics of Carbon Nanotubes 181
5.1.1. Structural 181
5.1.2. Electrical 182
5.1.3. Mechanical 185
5.1.4. Thermal 186
5.2. CNT Fabrication Technologies 186

5.2.1. Chemical Vapor Deposition of Carbon Nanotubes 187
5.2.2. Process Integration and Development 189
5.3. Carbon Nanotubes as Interconnects 191
5.3.1. Limitations of the Current Technology 191
5.3.2. Architecture, Geometry and Performance Potential Using Carbon Nanotubes 191
5.4. Design, Manufacture and Reliability 194
5.4.1. Microstructural Attributes and Effects on Electrical Characteristics 194
5.4.2. Interfacial Contact Materials 196
5.4.3. End-contacted Metal–CNT Junction 198
5.4.4. Thermal Stress Characteristics 198
5.4.5. Reliability Test 199
5.5. Summary 200
References 200

Chapter 6
Virtual Thermo-Mechanical Prototyping of Microelectronics and Microsystems
A. Wymysłowski, G.Q. Zhang, W.D. van Driel and L.J. Ernst 205
6.1. Introduction 205
6.2. Physical Aspects for Numerical Simulations 206
6.2.1. Numerical Modeling 208
6.2.2. Material Properties and Models 211
6.2.3. Thermo-Mechanical Related Failures 215
6.2.4. Designing for Reliability 219
6.3. Mathematical Aspects of Optimization 225
6.3.1. Design of Experiments 226
6.3.2. Response Surface Modeling 236
6.3.3. Advanced Approach to Virtual Prototyping 242
6.3.4. Designing for Quality 249
6.4. Application Case 252
6.4.1. Problem Description 252
6.4.2. Numerical Approach to QFN Package Design 253
6.5. Conclusion and Challenges 259
6.6. List of Acronyms 264
Acknowledgments 264
References 264

Materials Mechanics

Chapter 7
Fiber Optics Structural Mechanics and Nano-Technology Based New Generation of Fiber Coatings: Review and Extension
E. Suhir 269
7.1. Introduction 269
7.2. Fiber Optics Structural Mechanics 270
7.2.1. Review 270
7.3. New Nano-Particle Material (NPM) for Micro- and Opto-Electronic Applications 273
7.3.1. New Nano-Particle Material (NPM) 273
7.3.2. NPM-Based Optical Silica Fibers 274

7.4. Conclusions 277
Acknowledgment 277
References 277

Chapter 8
Area Array Technology for High Reliability Applications
Reza Ghaffarian 283

8.1. Introduction 283
8.2. Area Array Packages (AAPs) 284
8.2.1. Advantages of Area Array Packages 285
8.2.2. Disadvantages of Area Arrays 285
8.2.3. Area Array Types 286
8.3. Chip Scale Packages (CSPs) 286
8.4. Plastic Packages 288
8.4.1. Background 288
8.4.2. Plastic Area Array Packages 288
8.4.3. Plastic Package Assembly Reliability 289
8.4.4. Reliability Data for BGA, Flip Chip BGA, and CSP 291
8.5. Ceramic Packages 293
8.5.1. Background 293
8.5.2. Ceramic Package Assembly Reliability 294
8.5.3. Literature Survey on CBGA/CCGA Assembly Reliability 295
8.5.4. CBGA Thermal Cycle Test 297
8.5.5. Comparison of 560 I/O PBGA and CCGA assembly reliability 302
8.5.6. Designed Experiment for Assembly 305
8.6. Summary 309
8.7. List of Acronyms and Symbols 310
Acknowledgments 311
References 311

Chapter 9
Metallurgical Factors Behind the Reliability of High-Density Lead-Free Interconnections
Toni T. Mattila, Tomi T. Laurila and Jorma K. Kivilahti 313

9.1. Introduction 313
9.2. Approaches and Methods 315
9.2.1. The Four Steps of The Iterative Approach 315
9.2.2. The Role of Different Simulation Tools in Reliability Engineering 321
9.3. Interconnection Microstructures and Their Evolution 324
9.3.1. Solidification 324
9.3.2. Solidification Structure and the Effect of Contact Metalization Dissolution 325
9.3.3. Interfacial Reactions Products 330
9.3.4. Deformation Structures (Due to Slip and Twinning) 333
9.3.5. Recovery, Recrystallization and Grain Growth 335
9.4. Two Case Studies on Reliability Testing 335
9.4.1. Case 1: Reliability of Lead-Free CSPs in Thermal cycling 337
9.4.2. Case 2: Reliability of Lead-Free CSPs in Drop Testing 341
9.5. Summary 347

Acknowledgments 348
References 348

Chapter 10
Metallurgy, Processing and Reliability of Lead-Free Solder Joint Interconnections
Jin Liang, Nader Dariavach and Dongkai Shangguan 351

10.1. Introduction 351
10.2. Physical Metallurgy of Lead-Free Solder Alloys 352
10.2.1. Tin-Lead Solders 352
10.2.2. Lead-Free Solder Alloys 353
10.2.3. Interfacial Reaction: Wetting and Spreading 357
10.2.4. Interfacial Intermetallic Formation and Growth at Liquid–Solid Interfaces 363
10.3. Lead-Free Soldering Processes and Compatibility 377
10.3.1. Lead-Free Soldering Materials 378
10.3.2. PCB Substrates and Metalization Finishes 380
10.3.3. Lead-Free Soldering Processes 381
10.3.4. Components for Lead-Free Soldering 384
10.3.5. Design, Equipment and Cost Considerations 387
10.4. Reliability of Pb-Free Solder Interconnects 388
10.4.1. Reliability and Failure Distribution of Pb-Free Solder Joints 388
10.4.2. Effects of Loading and Thermal Conditions on Reliability of Solder Interconnection 389
10.4.3. Reliability of Pb-Free Solder Joints in Comparison to Sn-Pb Eutectic Solder Joints 395
10.5. Guidelines for Pb-free Soldering and Improvement in Reliability 406
References 406

Chapter 11
Fatigue Life Assessment for Lead-Free Solder Joints
Masaki Shiratori and Qiang Yu 411

11.1. Introduction 411
11.2. The Intermetallic Compound Formed at the Interface of the Solder Joints and the Cu-pad 412
11.3. Mechanical Fatigue Testing Equipment and Load Condition in the Lead Free Solder 413
11.4. Results of Mechanical Fatigue Test 414
11.5. Critical Fatigue Stress Limit for the Intermetallic Compound Layer 417
11.6. Influence of the Plating Material on the Fatigue Life of Sn-Zn (Sn-9Zn and Sn-8Zn-3Bi) Solder Joints 424
11.7. Conclusion 426
References 426

Chapter 12
Lead-Free Solder Materials: Design For Reliability
John H.L. Pang 429

12.1. Introduction 429
12.2. Mechanics of Solder Materials 430
12.2.1. Fatigue Behavior of Solder Materials 431
12.3. Design For Reliability (DFR) 433

12.4. Constitutive Models For Lead Free Solders 435
12.4.1. Tensile Test Results 435
12.4.2. Creep Test Results 440
12.5. Low Cycle Fatigue Models 443
12.6. FEA Modeling and Simulation 448
12.7. Reliability Test and Analysis 454
12.8. Conclusions 456
Acknowledgments 456
References 456

Chapter 13
Application of Moire Interferometry to Strain Analysis of PCB Deformations at Low Temperatures
Arkady Voloshin 459
13.1. Introduction 459
13.2. Optical Method and Recording of Fringe Patterns 460
13.2.1. Fractional Fringe Approach 461
13.2.2. Grating Frequency Increase 461
13.2.3. Creation of a High-Frequency Master Grating 462
13.2.4. Combination of the High Grating Frequency and Fractional Fringe Approach 463
13.3. Data Processing 463
13.4. Test Boards and Specimen Grating 463
13.5. Elevated Temperature Test 465
13.6. Low Temperature Test 468
13.7. Conclusions 470
Acknowledgment 472
References 473

Chapter 14
Characterization of Stresses and Strains in Microelectronics and Photonics Devices Using Photomechanics Methods
Bongtae Han 475
14.1. Introduction 475
14.2. Stress/Strain analysis 476
14.2.1. Moiré Interferometry 476
14.2.2. Extension: Microscopic Moiré Interferometry 477
14.2.3. Specimen Gratings 479
14.2.4. Strain Analysis 480
14.2.5. Thermal Deformation Measured at Room Temperature 481
14.2.6. Deformation as a Function of Temperature 485
14.2.7. Hygroscopic Deformation 494
14.2.8. Micromechanics 501
14.3. Warpage Analysis 505
14.3.1. Twyman/Green Interferometry 505
14.3.2. Shadow Moiré 509
14.3.3. Far Infrared Fizeau Interferometry 514
Acknowledgment 520
References 520

Chapter 15
Analysis of Reliability of IC Packages Using the Fracture Mechanics Approach
Andrew A.O. Tay 523

15.1. Introduction 523
15.2. Heat Transfer and Moisture Diffusion in IC Packages 525
15.3. Fundamentals of Interfacial Fracture Mechanics 527
15.4. Criterion for Crack Propagation 529
15.5. Interface Fracture Toughness 529
15.6. Total Stress Intensity Factor 530
15.7. Calculation of SERR and Mode Mixity 531
15.7.1. Crack Surface Displacement Extrapolation Method 531
15.7.2. Modified J-integral Method 532
15.7.3. Modified Virtual Crack Closure Method 533
15.7.4. Variable Order Boundary Element Method 536
15.7.5. Interaction Integral Method 536
15.8. Experimental Verification 538
15.9. Case Studies 542
15.9.1. Delamination Along Pad-Encapsulant Interface 542
15.9.2. Delamination Along Die-Attach/Pad Interface 544
15.9.3. Analysis Using Variable Order Boundary Element Method 546
15.10. Discussion of the Various Numerical Methods for Calculating G and ψ 549
15.11. Conclusion 551
References 551

Chapter 16
Dynamic Response of Micro- and Opto-Electronic Systems to Shocks and Vibrations: Review and Extension
E. Suhir 555

16.1. Introduction 555
16.2. Review 556
16.3. Extension: Quality of Shock Protection with a Flexible Wire Elements 557
16.4. Analysis 558
16.4.1. Pre-Buckling Mode: Small Displacements 558
16.4.2. Post-Buckling Mode: Large Displacements 564
16.5. Conclusions 567
References 568

Chapter 17
Dynamic Physical Reliability in Application to Photonic Materials
Dov Ingman, Tatiana Mirer and Ephraim Suhir 571

17.1. Introduction: Dynamic Reliability Approach to the Evolution of Silica Fiber Performance 571
17.1.1. Dynamic Physical Model of Damage Accumulation 572
17.1.2. Impact of the Three-Dimensional Mechanical-Temperature-Humidity Load on the Optical Fiber Reliability 575
17.1.3. Effect of Bimodality and Its Explanation Based on the Suggested Model 576
17.2. Reliability Improvement through NPM-Based Fiber Structures 585

17.2.1. Environmental Protection by NPM-Based Coating and Overall Self-Curing Effect of NPM Layers 585
17.2.2. Improvement in the Reliability Characteristics by Employing NPM Structures in Optical Fibers 587
17.3. Conclusions 593
References 593

Chapter 18
High-Speed Tensile Testing of Optical Fibers—New Understanding for Reliability Prediction
Sergey Semjonov and G. Scott Glaesemann 595
18.1. INTRODUCTION 595
18.2. Theory 596
18.2.1. Single-Region Power-Law Model 596
18.2.2. Two-Region Power-Law Model 598
18.2.3. Universal Static and Dynamic Fatigue Curves 599
18.3. Experimental 602
18.3.1. Sample Preparation 602
18.3.2. Dynamic Fatigue Tests 604
18.3.3. Static Fatigue Tests 605
18.4. Results and Discussion 606
18.4.1. High-Speed Testing 606
18.4.2. Static Fatigue 610
18.4.3. Influence of Multiregion Model on Lifetime Prediction 613
18.5. Conclusion 613
References 614
Appendix 18.A: High Speed Axial Strength Testing: Measurement Limits 616
Appendix 18.B: Incorporating Static Fatigue Results into Dynamic Fatigue Curves 620
18.B.1. Static Fatigue Test 620
18.B.2. Dynamic Fatigue Test 621
18.B.3. Discussion 622

Chapter 19
The Effect of Temperature on the Microstructure Nonlinear Dynamics Behavior
Xiaoling He 627
19.1. Introduction 627
19.2. Theoretical Development 630
19.2.1. Background on Nonlinear Dynamics and Nonlinear Thermo-Elasticity Theories 630
19.2.2. Nonlinear Thermo-Elasticity Development for an Isotropic Laminate Subject to Thermal and Mechanical and Load 631
19.3. Thin Laminate Deflection Response Subject to Thermal Effect and Mechanical Load 633
19.3.1. Steady State Temperature Effect 633
19.3.2. Transient Thermal Field Effect 638
19.4. Stress Field in Nonlinear Dynamics Response 653
19.4.1. Stress Field Formulation 653
19.4.2. Stress Distribution 654
19.4.3. Failure Analysis 654

19.5. Discussions 660
19.6. Summary 661
Nomenclature 662
Acknowledgment 663
References 663

Chapter 20
Effect of Material's Nonlinearity on the Mechanical Response of some Piezoelectric and Photonic Systems
Victor Birman and Ephraim Suhir 667

20.1. Introduction 667
20.2. Effect of Physical Nonlinearity on Vibrations of Piezoelectric Rods Driven by Alternating Electric Field 668
20.2.1. Physically Nonlinear Constitutive Relationships for an Orthotropic Cylindrical Piezoelectric Rod Subject to an Electric Field in the Axial Direction 670
20.2.2. Analysis of Uncoupled Axial Vibrations 673
20.2.3. Solution for Coupled Axial-Radial Axisymmetric Vibrations by the Generalized Galerkin Procedure 677
20.2.4. Numerical Results and Discussion 678
20.3. The Effect of the Nonlinear Stress–Strain Relationship on the Response of Optical Fibers 683
20.3.1. Stability of Optical Fibers 684
20.3.2. Stresses and Strains in a Lightwave Coupler Subjected to Tension 686
20.3.3. Free Vibrations 690
20.3.4. Bending of an Optical Fiber 692
20.4. Conclusions 695
Acknowledgment 696
References 697

Index 701

Volume II

List of Contributors xxvii

Preface xxxi

Physical Design

Chapter 1
Analytical Thermal Stress Modeling in Physical Design for Reliability of Micro- and Opto-Electronic Systems: Role, Attributes, Challenges, Results
E. Suhir 3

1.1. Thermal Loading and Thermal Stress Failures 3
1.2. Thermal Stress Modeling 4
1.3. Bi-Metal Thermostats and other Bi-Material Assemblies 5
1.4. Finite-Element Analysis 5

1.5. Die-Substrate and other Bi-Material Assemblies 6
1.6. Solder Joints 8
1.7. Design Recommendations 9
1.8. “Global” and “Local” Mismatch and Assemblies Bonded at the Ends 10
1.9. Assemblies with Low Modulus Adhesive Layer at the Ends 11
1.10. thermally Matched Assemblies 11
1.11. Thin Films 12
1.12. Polymeric Materials And Plastic Packages 13
1.13. Thermal Stress Induced Bowing and Bow-Free Assemblies 14
1.14. Probabilistic Approach 15
1.15. Optical Fibers and other Photonic Structures 15
1.16. Conclusion 16
References 17

Chapter 2
Probabilistic Physical Design of Fiber-Optic Structures
Satish Radhakrishnan, Ganesh Subbarayan and Luu Nguyen 23
2.1. Introduction 23
2.1.1. Demonstration Vehicle 24
2.2. Optical Model 25
2.2.1. Mode Field Diameter 26
2.2.2. Refraction and Reflection Losses 27
2.2.3. Calculations for Coupling Losses 27
2.2.4. Coupling Efficiency 28
2.3. Interactions in System and Identification of Critical Variables 30
2.3.1. Function Variable Incidence Matrix 30
2.3.2. Function Variable Incidence Matrix to Graph Conversion 31
2.3.3. Graph Partitioning Techniques 34
2.3.4. System Decomposition using Simulated Annealing 34
2.4. Deterministic Design Procedures 37
2.4.1. Optimal and Robust Design 40
2.4.2. A Brief Review of Multi-Objective Optimization 42
2.4.3. Implementation 43
2.4.4. Results 43
2.5. Stochastic Analysis 44
2.5.1. The First and Second Order Second Moment Methods 44
2.6. Probabilistic Design for Maximum Reliability 46
2.6.1. Results 49
2.7. Stochastic Characterization of Epoxy Behavior 51
2.7.1. Viscoelastic Models 52
2.7.2. Modeling the Creep Test 53
2.7.3. Dynamic Mechanical Analysis 54
2.7.4. Experimental Results 55
2.8. Analytical Model to Determine VCSEL Displacement 57
2.8.1. Results 63
2.9. Summary 67
References 67

Chapter 3
The Wirebonded Interconnect: A Mainstay for Electronics
Harry K. Charles, Jr. 71

3.1. Introduction 71
3.1.1. Integrated Circuit Revolution 71
3.1.2. Interconnection Types 72
3.1.3. Wirebond Importance 80
3.2. Wirebonding Basics 81
3.2.1. Thermocompression Bonding 81
3.2.2. Ultrasonic Bonding 83
3.2.3. Thermosonic Bonding 85
3.2.4. Wirebond Reliability 87
3.2.5. Wirebond Testing 89
3.2.6. Bonding Automation and Optimization 93
3.3. Materials 95
3.3.1. Bonding Wire 95
3.3.2. Bond Pad Metallurgy 100
3.3.3. Gold Plating 102
3.3.4. Pad Cleaning 104
3.4. Advanced Bonding Methods 105
3.4.1. Fine Pitch Bonding 105
3.4.2. Soft Substrates 108
3.4.3. Machine Improvements 110
3.4.4. Higher Frequency Wirebonding 110
3.4.5. Stud Bumping 115
3.5. Summary 116
Acknowledgments 116
References 116

Chapter 4
Metallurgical Interconnections for Extreme High and Low Temperature Environments
George G. Harman 121

4.1. Introduction 121
4.2. High Temperature Interconnections Requirements 122
4.2.1. Wire Bonding 122
4.2.2. The Use of Flip Chips in HTE 127
4.2.3. General Overview of Metallurgical Interfaces for Both HTE and LTE 129
4.3. Low Temperature Environment Interconnection Requirements 129
4.4. Corrosion and Other Problems in Both *HTE*, and *LTE* 130
4.5. The Potential Use of High Temperature Polymers in HTE 131
4.6. Conclusions 132
Acknowledgments 132
References 132

Chapter 5
Design, Process, and Reliability of Wafer Level Packaging
Zhuqing Zhang and C.P. Wong 135

5.1. Introduction 135

5.2. WLCSP 137
5.2.1. Thin Film Redistribution 137
5.2.2. Encapsulated Package 139
5.2.3. Compliant Interconnect 139
5.3. Wafer Level Underfill 141
5.3.1. Challenges of Wafer Level Underfill 142
5.3.2. Examples of Wafer Level Underfill Process 143
5.4. Comparison of Flip-Chip and WLCSP 145
5.5. Wafer Level Test and Burn-In 145
5.6. Summary 149
References 149

Chapter 6
Passive Alignment of Optical Fibers in V-grooves with Low Viscosity Epoxy Flow
S.W. Ricky Lee and C.C. Lo 151
6.1. Introduction 151
6.2. Design and Fabrication of Silicon Optical Bench with V-grooves 152
6.3. Issues of Conventional Passive Alignment Methods 158
6.3.1. V-grooves with Cover Plate 158
6.3.2. Edge Dispensing of Epoxy 161
6.4. Modified Passive Alignment Method 162
6.4.1. Working Principle 162
6.4.2. Alignment Mechanism 163
6.4.3. Design of Experiment 164
6.4.4. Experimental Procedures 164
6.4.5. Experimental Results 165
6.5. Effects of Epoxy Viscosity and Dispensing Volume 168
6.6. Application to Fiber Array Passive Alignment 170
6.7. Conclusions and Discussion 172
References 172

Reliability and Packaging

Chapter 7
Fundamentals of Reliability and Stress Testing
H. Anthony Chan 177
7.1. More Performance at Lower Cost in Shorter Time-to-market 178
7.1.1. Rapid Technological Developments 178
7.1.2. Integration of More Products into Human Life 178
7.1.3. Diverse Environmental Stresses 178
7.1.4. Competitive Market 179
7.1.5. Short Product Cycles 179
7.1.6. The Bottom Line 179
7.2. Measure of Reliability 180
7.2.1. Failure Rate 180
7.2.2. Systems with Multiple Independent Failure Modes 181
7.2.3. Failure Rate Distribution 182
7.3. Failure Mechanisms in Electronics and Packaging 184

7.3.1. Failure Mechanisms at Chip Level Include 184
7.3.2. Failure Mechanisms at Bonding Include 184
7.3.3. Failure Mechanisms in Device Packages Include 185
7.3.4. Failure Mechanisms in Epoxy Compounds Include 185
7.3.5. Failure Mechanisms at Shelf Level Include 185
7.3.6. Failure Mechanisms in Material Handling Include 185
7.3.7. Failure Mechanisms in Fiber Optics Include 185
7.3.8. Failure Mechanisms in Flat Panel Displays Include 186
7.4. Reliability Programs and Strategies 186
7.5. Product Weaknesses and Stress Testing 187
7.5.1. Why do Products Fail? 187
7.5.2. Stress Testing Principle 189
7.6. Stress Testing Formulation 191
7.6.1. Threshold and Cumulative Stress Failures 191
7.6.2. Stress Stimuli and Flaws 192
7.6.3. Modes of Stress Testing 193
7.6.4. Lifetime Failure Fraction 194
7.6.5. Robustness Against Maximum Service Life Stress 195
7.6.6. Stress–Strength Contour 197
7.6.7. Common Issues 198
7.7. Further Reading 201

Chapter 8
How to Make a Device into a Product: *Accelerated Life Testing (ALT), Its Role, Attributes, Challenges, Pitfalls, and Interaction with Qualification Tests*
E. Suhir 203

8.1. Introduction 203
8.2. Some Major Definitions 204
8.3. Engineering Reliability 204
8.4. Field Failures 205
8.5. Reliability is a Complex Property 206
8.6. Three Major Classes of Engineering Products and Market Demands 206
8.7. Reliability, Cost and Time-to-Market 208
8.8. Reliability Costs Money 208
8.9. Reliability Should Be Taken Care of on a Permanent Basis 209
8.10. Ways to Prevent and Accommodate Failures 210
8.11. Redundancy 211
8.12. Maintenance and Warranty 211
8.13. Test Types 212
8.14. Accelerated Tests 212
8.15. Accelerated Test Levels 213
8.16. Qualification Standards 213
8.17. Accelerated Life Tests (ALTs) 214
8.18. Accelerated Test Conditions 215
8.19. Acceleration Factor 216
8.20. Accelerated Stress Categories 217
8.21. Accelerated Life Tests (ALTs) and Highly Accelerated Life Tests (HALTs) 218
8.22. Failure Mechanisms and Accelerated Stresses 219

8.23. ALTs: Pitfalls and Challenges 219
8.24. Burn-ins 220
8.25. Wear-Out Failures 221
8.26. Non-Destructive Evaluations (NDE's) 222
8.27. Predictive Modeling 222
8.28. Some Accelerated Life Test (ALT) Models 223
8.28.1. Power Law 224
8.28.2. Boltzmann-Arrhenius Equation 224
8.28.3. Coffin-Manson Equation (Inverse Power Law) 225
8.28.4. Paris-Erdogan Equation 226
8.28.5. Bueche-Zhurkov Equation 227
8.28.6. Eyring Equation 227
8.28.7. Peck and Black Equations 227
8.28.8. Fatigue Damage Model (Miner's Rule) 228
8.28.9. Creep Rate Equations 228
8.28.10. Weakest Link Models 228
8.28.11. Stress–Strength Models 229
8.29. Probability of Failure 229
8.30. Conclusions 230
References 230

Chapter 9
Micro-Deformation Analysis and Reliability Estimation of Micro-Components by Means of NanoDAC Technique
Bernd Michel and Jürgen Keller 233

9.1. Introduction 233
9.2. Basics of Digital Image Correlation 234
9.2.1. Cross Correlation Algorithms on Gray Scale Images 234
9.2.2. Subpixel Analysis for Enhanced Resolution 236
9.2.3. Results of Digital Image Correlation 238
9.3. Displacement and Strain Measurements on SFM Images 239
9.3.1. Digital Image Correlation under SPM Conditions 239
9.3.2. Technical Requirements for the Application of the Correlation Technique 241
9.4. Deformation Analysis on Thermally and Mechanically Loaded Objects under the SFM 241
9.4.1. Reliability Aspects of Sensors and Micro Electro-Mechanical Systems (MEMS) 241
9.4.2. Thermally Loaded Gas Sensor under SFM 242
9.4.3. Crack Detection and Evaluation by SFM 243
9.5. Conclusion and Outlook 250
References 250

Chapter 10
Interconnect Reliability Considerations in Portable Consumer Electronic Products
Sridhar Canumalla and Puligandla Viswanadham 253

10.1. Introduction 253
10.2. Reliability—Thermal, Mechanical and Electrochemical 255
10.2.1. Accelerated Life Testing 255
10.2.2. Thermal Environment 257

10.2.3. Mechanical Environment 257
10.2.4. Electrochemical Environment 264
10.2.5. Tin Whiskers 267
10.3. Reliability Comparisons in Literature 267
10.3.1. Thermomechanical Reliability 268
10.3.2. Mechanical Reliability 270
10.4. Influence of Material Properties on Reliability 271
10.4.1. Printed Wiring Board 271
10.4.2. Package 272
10.4.3. Surface Finish 272
10.5. Failure Mechanisms 273
10.5.1. Thermal Environment 273
10.5.2. Mechanical Environment 276
10.5.3. Electrochemical Environment 286
10.6. reliability test Practices 291
10.7. Summary 294
Acknowledgments 295
References 295

Chapter 11
MEMS Packaging and Reliability
Y.C. Lee 299
11.1. Introduction 299
11.2. Flip-Chip Assembly for Hybrid Integration 304
11.3. Soldered Assembly for Three-Dimensional MEMS 309
11.4. Flexible Circuit Boards for MEMS 313
11.5. Atomic Layer Deposition for Reliable MEMS 316
11.6. Conclusions 320
Acknowledgments 320
References 320

Chapter 12
Advances in Optoelectronic Methodology for MOEMS Testing
Ryszard J. Pryputniewicz 323
12.1. Introduction 323
12.2. MOEMS Samples 324
12.3. Analysis 328
12.4. Optoelectronic Methodology 330
12.5. Representative Applications 334
12.6. Conclusions and Recommendations 338
Acknowledgments 339
References 339

Chapter 13
Durability of Optical Nanostructures: Laser Diode Structures and Packages, A Case Study
Ajay P. Malshe and Jay Narayan 341

13.1. High Efficiency Quantum Confined (Nanostructured) III-Nitride Based Light Emitting Diodes And Lasers 342
13.1.1. Introduction 342
13.2. Investigation of Reliability Issues in High Power Laser Diode Bar Packages 348
13.2.1. Introduction 348
13.2.2. Preparation of Packaged Samples for Reliability Testing 349
13.2.3. Finding and Model of Reliability Results 350
13.3. Conclusions 357
Acknowledgments 358
References 358

Chapter 14
Review of the Technology and Reliability Issues Arising as Optical Interconnects Migrate onto the Circuit Board
P. Misselbrook, D. Gwyer, C. Bailey, D. Gwyer, C. Bailey, P.P. Conway and K. Williams 361
14.1. Background to Optical Interconnects 362
14.2. Transmission Equipment for Optical Interconnects 362
14.3. Very Short Reach Optical Interconnects 365
14.4. Free Space USR Optical Interconnects 366
14.5. Guided Wave USR Interconnects 367
14.6. Component Assembly of OECB's 370
14.7. Computational Modeling of Optical Interconnects 373
14.8. Conclusions 380
Acknowledgments 380
References 381

Chapter 15
Adhesives for Micro- and Opto-Electronics Application: Chemistry, Reliability and Mechanics
D.W. Dahringer 383
15.1. Introduction 383
15.1.1. Use of Adhesives in Micro and Opto-Electronic Assemblies 383
15.1.2. Specific Applications 384
15.2. Adhesive Characteristics 385
15.2.1. General Properties of Adhesives 385
15.2.2. Adhesive Chemistry 390
15.3. Design Objective 393
15.3.1. Adhesive Joint Design 393
15.3.2. Manufacturing Issues 397
15.4. Failure Mechanism 401
15.4.1. General 401
15.4.2. Adhesive Changes 401
15.4.3. Interfacial Changes 401
15.4.4. Interfacial Stress 401
15.4.5. External Stress 402
References 402

Chapter 16
Multi-Stage Peel Tests and Evaluation of Interfacial Adhesion Strength for Micro- and Opto-Electronic Materials
Masaki Omiya, Kikuo Kishimoto and Wei Yang 403

16.1. Introduction 403
16.2. Multi-Stage Peel Test (MPT) 407
16.2.1. Testing Setup 407
16.2.2. Multi-Stage Peel Test 408
16.2.3. Energy Variation in Steady State Peeling 409
16.3. Interfacial Adhesion Strength of Copper Thin Film 413
16.3.1. Preparation of Specimen 413
16.3.2. Measurement of Adhesion Strength by the MPT 414
16.3.3. Discussions 415
16.4. UV-Irradiation Effect on Ceramic/Polymer Interfacial Strength 419
16.4.1. Preparation of PET/ITO Specimen 419
16.4.2. Measurement of Interfacial Strength by MPT 422
16.4.3. Surface Crack Formation on ITO Layer under Tensile Loading 424
16.5. Concluding Remarks 426
Acknowledgment 427
References 427

Chapter 17
The Effect of Moisture on the Adhesion and Fracture of Interfaces in Microelectronic Packaging
Timothy P. Ferguson and Jianmin Qu 431

17.1. Introduction 432
17.2. Moisture Transport Behavior 433
17.2.1. Background 433
17.2.2. Diffusion Theory 434
17.2.3. Underfill Moisture Absorption Characteristics 435
17.2.4. Moisture Absorption Modeling 438
17.3. Elastic Modulus Variation Due to Moisture Absorption 442
17.3.1. Background 442
17.3.2. Effect of Moisture Preconditioning 444
17.3.3. Elastic Modulus Recovery from Moisture Uptake 447
17.4. Effect of Moisture on Interfacial Adhesion 449
17.4.1. Background 449
17.4.2. Interfacial Fracture Testing 451
17.4.3. Effect of Moisture Preconditioning on Adhesion 452
17.4.4. Interfacial Fracture Toughness Recovery from Moisture Uptake 461
17.4.5. Interfacial Fracture Toughness Moisture Degradation Model 462
References 469

Chapter 18
Highly Compliant Bonding Material for Micro- and Opto-Electronic Applications
E. Suhir and D. Ingman 473

18.1. Introduction 473

18.2. Effect of the Interfacial Compliance on the interfacial Shearing Stress 474
18.3. Internal Compressive Forces 476
18.4. Advanced Nano-Particle Material (NPM) 476
18.5. Highly-Compliant Nano-Systems 478
18.6. Conclusions 479
References 480
Appendix 18.A: Bimaterial Assembly Subjected to an External Shearing Load and Change in Temperature: Expected Stress Relief due to the Elevated Interfacial Compliance 480
Appendix 18.B: Cantilever Wire ("Beam") Subjected at its Free End to a Lateral (Bending) and an Axial (Compressive) Force 483
Appendix 18.C: Compressive Forces in the NPM-Based Compound Structure 485

Chapter 19
Adhesive Bonding of Passive Optical Components
Anne-Claire Pliska and Christian Bosshard 487

19.1. Introduction 487
19.2. Optical Devices and Assemblies 489
19.2.1. Optical Components 489
19.2.2. Opto-electronics Assemblies: Specific Requirements 489
19.3. Adhesive Bonding in Optical Assemblies 503
19.3.1. Origin of Adhesion 503
19.3.2. Adhesive Selection and Dispensing 508
19.3.3. Dispensing Technologies 515
19.4. Some Applications 518
19.4.1. Laser to Fiber Assembly 518
19.4.2. Planar Lightwave Circuit (PLC) Pigtailing 520
19.5. Summary and Recommendations 522
Acknowledgments 523
References 523

Chapter 20
Electrically Conductive Adhesives: A Research Status Review
James E. Morris and Johan Liu 527

20.1. Introduction 527
20.1.1. Technology Drivers 527
20.1.2. Isotropic Conductive Adhesives (ICAs) 529
20.1.3. Anisotropic Conductive Adhesives (ACAs) 529
20.1.4. Non-Conductive Adhesive (NCA) 529
20.2. Structure 529
20.2.1. ICA 529
20.2.2. ACA 532
20.2.3. Modeling 534
20.3. Materials and Processing 534
20.3.1. Polymers 534
20.3.2. ICA Filler 536
20.3.3. ACA Processing 536
20.4. Electrical Properties 538

20.4.1. ICA 538
20.4.2. Electrical Measurements 544
20.4.3. ACA 544
20.5. Mechanical Properties 546
20.5.1. ICA 546
20.5.2. ACA 547
20.6. Thermal Properties 553
20.6.1. Thermal Characteristics 553
20.6.2. Maximum Current Carrying Capacity 553
20.7. Reliability 554
20.7.1. ICA 554
20.7.2. ACA 557
20.7.3. General Comments 565
20.8. Environmental Impact 565
20.9. Further Study 565
References 565

Chapter 21
Electrically Conductive Adhesives
Johann Nicolics and Martin Mündlein 571
21.1. Introduction and Historical Background 571
21.2. Contact Formation 574
21.2.1. Percolation and Critical Filler Content 574
21.2.2. ICA Contact Model 575
21.2.3. Results 578
21.3. Aging Behavior and Quality Assessment 595
21.3.1. Introduction 595
21.3.2. Material Selection and Experimental Parameters 595
21.3.3. Curing Parameters and Definition of Curing Time 597
21.3.4. Testing Conditions, Typical Results, and Conclusions 598
21.4. About Typical Applications 602
21.4.1. ICA for Attachment of Power Devices 602
21.4.2. ICA for Interconnecting Parts with Dissimilar Thermal Expansion Coefficient 604
21.4.3. ICA for Cost-Effective Assembling of Multichip Modules 606
21.5. Summary 607
Notations and Definitions 607
References 608

Chapter 22
Recent Advances of Conductive Adhesives: A Lead-Free Alternative in Electronic Packaging
Grace Y. Li and C.P. Wong 611
22.1. Introduction 611
22.2. Isotropic Conductive Adhesives (ICAs) 613
22.2.1. Improvement of Electrical Conductivity of ICAs 614
22.2.2. Stabilization of Contact Resistance on Non-Noble Metal Finishes 615
22.2.3. Silver Migration Control of ICA 618

22.2.4. Improvement of Reliability in Thermal Shock Environment 618
22.2.5. Improvement of Impact Performance of ICA 619
22.3. Anisotropic Conductive Adhesives (ACAs)/Anisotropic Conductive Film (ACF) 619
22.3.1. Materials 620
22.3.2. Application of ACA/ACF in Flip Chip 621
22.3.3. Improvement of Electrical Properties of ACAs 621
22.3.4. Thermal Conductivity of ACA 623
22.4. Future Advances of ECAs 623
22.4.1. Electrical Characteristics 623
22.4.2. High Frequency Compatibility 623
22.4.3. Reliability 623
22.4.4. ECAs with Nano-filler for Wafer Level Application 625
References 625

Chapter 23
Die Attach Quality Testing by Structure Function Evaluation
Márta Rencz, Vladimir Székely and Bernard Courtois 629

Nomenclature 629
Greek symbols 629
Subscripts 630
23.1. Introduction 630
23.2. Theoretical Background 630
23.3. Detecting Voids in the Die Attach of Single Die Packages 634
23.4. Simulation Experiments for Locating the Die Attach Failure on Stacked Die Packages 636
23.4.1. Simulation Tests Considering Stacked Dies of the Same Size 637
23.4.2. Simulation Experiments on a Pyramidal Structure 639
23.5. Verification of the Methodology by Measurements 642
23.5.1. Comparison of the Transient Behavior of Stacked Die Packages Containing Test Dies, Prior Subjected to Accelerated Moisture and Temperature Testing 642
23.5.2. Comparison of the Transient Behavior of Stacked Die Packages Containing Real Functional Dies, Subjected Prior to Accelerated Moisture and Temperature Testing 644
23.6. Conclusions 649
Acknowledgments 649
References 650

Chapter 24
Mechanical Behavior of Flip Chip Packages under Thermal Loading
Enboa Wu, Shoulung Chen, C.Z. Tsai and Nicholas Kao 651

24.1. Introduction 651
24.2. Flip Chip Packages 652
24.3. Measurement Methods 654
24.3.1. Phase Shifted Shadow Moiré Method 654
24.3.2. Electronic Speckle Pattern Interferometry (ESPI) Method 655
24.4. Substrate CTE Measurement 656
24.5. Behavior of Flip Chip Packages under Thermal Loading 661
24.5.1. Warpage at Room Temperature 661

24.5.2. Warpage at Elevated Temperatures 662
24.5.3. Effect of Underfill on Warpage 666
24.6. Finite Element Analysis of Flip Chip Packages under Thermal Loading 668
24.7. Parametric Study of Warpage for Flip Chip Packages 669
24.7.1. Change of the Chip Thickness 670
24.7.2. Change of the Substrate Thickness 670
24.7.3. Change of the Young's Modulus of the Underfill 671
24.7.4. Change of the CTE of the Underfill 672
24.7.5. Effect of the Geometry of the Underfill Fillet 672
24.8. Summary 674
References 674

Chapter 25
Stress Analysis for Processed Silicon Wafers and Packaged Micro-devices
Li Li, Yifan Guo and Dawei Zheng 677
25.1. Intrinsic Stress Due to Semiconductor Wafer Processing 677
25.1.1. Testing Device Structure 678
25.1.2. Membrane Deformations 679
25.1.3. Intrinsic Stress 681
25.1.4. Intrinsic Stress in Processed Wafer: Summary 683
25.2. Die Stress Result from Flip-chip Assembly 685
25.2.1. Consistent Composite Plate Model 685
25.2.2. Free Thermal Deformation 687
25.2.3. Bimaterial Plate (BMP) Case 688
25.2.4. Validation of the Bimaterial Model 691
25.2.5. Flip-Chip Package Design 695
25.2.6. Die Stress in Flip Chip Assembly: Summary 697
25.3. Thermal Stress Due to Temperature Cycling 698
25.3.1. Finite Element Analysis 698
25.3.2. Constitutive Equation for Solder 699
25.3.3. Time-Dependent Thermal Stresses of Solder Joint 700
25.3.4. Solder Joint Reliability Estimation 701
25.3.5. Thermal Stress Due to Temperature Cycling: Summary 703
25.4. Residual Stress in Polymer-based Low Dielectric Constant (low-k) Materials 703
References 708

Index 711

List of Contributors

VOLUME I

Avram Bar-Cohen
University of Maryland
College Park, Maryland, USA

Victor Birman
University of Missouri-Rolla
St. Louis, Missouri, USA

H.J.L. Bressers
Philips Semiconductors
Nijmegen, The Netherlands

Alan M. Cassell
NASA Ames Research Center
Moffett Field, California, USA

N. Dariavach
EMC Corp
Hopkinton, Massachusetts, USA

Liang Dong
Lightwaves 2020 Inc.
Milpitas, California, USA

L.J. Ernst
Delft University of Technology
Delft, The Netherlands

Reza Ghaffarian
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California, USA

G. Scott Glaesemann
Corning Incorporated
Corning, New York, USA

Claire Gu
University of California, Santa Cruz
Santa Cruz, California, USA

Bongtae Han
University of Maryland
College Park, Maryland, USA

Henry He
Lightwaves 2020 Inc.
Milpitas, California, USA

Xiaoling He
University of Wisconsin
Milwaukee, Wisconsin, USA

Dov Ingman
Technion, Israel Institute of Technology
Haifa, Israel

K.M.B. Jansen
Delft University of Technology
Delft, The Netherlands

J.H.J. Janssen
Philips Semiconductors
Nijmegen, The Netherlands

Kyoung Joon Kim
University of Maryland
College Park, Maryland, USA

Jorma K. Kivilahti
Helsinki University of Technology
Helsinki, Finland

Tomi T. Laurila
Helsinki University of Technology
Helsinki, Finland

J. Liang
EMC Corp
Hopkinton, Massachusetts, USA

Yisi Liu
University of California, Santa Cruz
Santa Cruz, California, USA

Jun Li
NASA Ames Research Center
Moffett Field, California, USA

Toni T. Mattila
Helsinki University of Technology
Helsinki, Finland

Tatiana Mirer
Technion, Israel Institute of Technology
Haifa, Israel

J.J. Pan
Lightwaves 2020 Inc.
Milpitas, California, USA

John H.L. Pang
Nanyang Technological University
Nanyang, Singapore

David T. Read
National Institute of Standards and Technology
Boulder, Colorado, USA

Sergey Semjonov
Fiber Optics Research Center
Moscow, Russia

D. Shangguan
FLEXTRONICS
San Jose, California, USA

Masaki Shiratori
Yokohama National University
Yokohama, Japan

Andrew A.O. Tay
National University of Singapore
Republic of Singapore

W.D. van Driel
Delft University of Technology
Delft, The Netherlands

C. van't Hof
Delft University of Technology
Delft, The Netherlands

Alex A. Volinsky
University of South Florida
Tampa, Florida, USA

Arkady Voloshin
Lehigh University
Bethlehem, Pennsylvania, USA

A. Wymyslowski
Wroclaw University of Technology
Wroclaw, Poland

Yuan Xu
University of California, Santa Cruz
Santa Cruz, California, USA

D.G. Yang
Delft University of Technology
Delft, The Netherlands

Qiang Yu
Yokohama National University
Yokohama, Japan

G.Q. Zhang
Delft University of Technology
Delft, The Netherlands and
Philips Semiconductors
Eindhoven, The Netherlands

Fengqing Zhou
Lightwaves 2020 Inc.
Milpitas, California, USA

VOLUME II

C. Bailey
University of Greenwich
London, United Kingdom

Christian Bosshard
CSEM SA, Untere Grundlistrasse 1
Alpnach Dorf, Switzerland

Sridhar Canumalla
Nokia
Irving, Texas, USA

H. Anthony Chan
University of Cape Town
Rondebosch, South Africa

Harry K. Charles, Jr.
The Johns Hopkins University
Laurel, Maryland, USA

Shoulung Chen
National Taiwan University
Taiwan

Bernard Courtois
TIMA-CMP
Grenoble Cedex, France

D.W. Dahringer
D.W. Dahringer Consultants
Glen Ridge, New Jersey, USA

Timothy P. Ferguson
Southern Research Institute
Birmingham, Alabama, USA

Yifan Guo
Skyworks Solutions, Inc.
Irvine, California, USA

D. Gwyer
University of Greenwich
London, United Kingdom

George G. Harman
National Institute of Standards and Technology
Gaithersburg, Maryland, USA

D. Ingman
Technion, Israel Institute of Technology
Haifa, Israel

Nicholas Kao
National Taiwan University
Taiwan

J. Keller
Fraunhofer Institute for Reliability and Micro Integration (IZM)
Berlin, Germany

Kikuo Kishimoto
Tokyo Institute of Technology
Tokyo, Japan

S.W.R. Lee
Hong Kong University of Science and Technology
Clear Water Bay, Kowloon, Hong Kong

Grace Y. Li
Georgia Institute of Technology
Atlanta, Georgia, USA

Li Li
Cisco Systems, Inc.
San Jose, California, USA

Johan Liu
Chalmers University of Technology
Goteborg, Sweden

C.C. Lo
Hong Kong University of Science and Technology
Clear Water Bay, Kowloon, Hong Kong

Ajay P. Malshe
University of Arkansas
Fayetteville, Arkansas, USA

Bernd Michel
Fraunhofer MicroMaterials Center
Berlin, Germany

P. Misselbrook
Celestica
Kidsgrove, Stoke-on-Trent, United Kingdom
and
University of Greenwich
London, United Kingdom

James E. Morris
Portland State University
Portland, Oregon, USA

Martin Mündlein
Vienna Institute of Technology
Vienna, Austria

Jay Narayan
North Carolina State University
Raleigh, North Carolina, USA

Johann Nicolics
Vienna Institute of Technology
Vienna, Austria

Luu Nguyen
National Semiconductor Corporation
Santa Clara, California, USA

Masaki Omiya
Tokyo Institute of Technology
Tokyo, Japan

Anne-Claire Pliska
CSEM SA, Untere Grundlistrasse 1
Alpnach Dorf, Switzerland

Jianmin Qu
Georgia Institute of Technology
Atlanta, Georgia, USA

Satish Radhakrishanan
Purdue University
West Lafayette, Indiana, USA

Marta Rencz
MicReD Ltd.
Budapest, Hungary

Ganesh Subbarayan
Purdue University
West Lafayette, Indiana, USA

Vladimir Szekely
Budapest University of Technology and Economics
Budapest, Hungary

C.Z. Tsai
National Taiwan University
Taiwan

Puligandla Viswanadham
Nokia Research Center
Irving, Texas, USA

K. Williams
Loughborough University
Loughborough, United Kingdom

Enboa Wu
National Taiwan University
Taipei, Taiwan

Wei Yang
Tsinghua University
Beijing, P.R. China

Z. Zhang
Georgia Institute of Technology
Atlanta, Georgia, USA

Dawei Zheng
Kotura, Inc.
Monterey Park, California, USA

Preface

This book encompasses a broad area of micro- and opto-electronic engineering materials: their physics, mechanics, reliability, and packaging, with an emphasis on physical design issues and problems. The editors tried to bring in the most eminent engineers and scientists as chapter authors and put together the most comprehensive book ever written on the subjects of materials, mechanics, physics, packaging, functional performance, mechanical reliability, environmental durability and other aspects of reliability of micro- and opto-electronic assemblies, components, devices, and systems. University professors and leading industrial engineers contributed to the book. The contents of the book reflect the state-of-the-art in the above listed fields of applied science and engineering.

The intended audience are all those who work in micro- and opto-electronics, and photonics; electronic and optical materials; applied and industrial physics; mechanical and reliability engineering; electron and optical devices and systems. The expected and targeted readers are practitioners and professionals, scientists and researchers, lecturers and continuing education course directors, graduate and undergraduate students, technical supervisors and entrepreneurs. The book can serve, to a great extent, as an encyclopedia in the field of physics and mechanics of micro- and opto-electronic materials and structures. In the editors' opinion, it can serve also as a textbook, as a reference book, and as a guidance for self- and continuing education, i.e., as a source of comprehensive and in-depth information in its areas. The book's chapters contain both the description of the state-of-the-art in a particular field, as well as new results obtained by the chapter authors and their colleagues.

We would like to point out that many methods and approaches addressed in this book extend far beyond microelectronics and photonics. Although these methods and approaches were developed, advanced and reported primarily in application to micro- and opto-electronic systems, they are applicable also in many related areas of engineering and physics.

The editors are proud of the broad scope of the book, and of the quality of the contributed chapters, and would like to take this opportunity to deeply acknowledge, with thanks, the conscientious effort of the numerous contributors.

February 2006
E. Suhir
C.-P. Wong
Y.-C. Lee

MATERIALS PHYSICS

1

Polymer Materials Characterization, Modeling and Application

L.J. Ernst[a], K.M.B. Jansen[a], D.G. Yang[a], C. van 't Hof[a], H.J.L. Bressers[b], J.H.J. Janssen[b], and G.Q. Zhang[a,c]

[a]*Delft University of Technology, Department Precision and Microsystems Engineering, Mekelweg 2, 2628 CD Delft, The Netherlands*
[b]*Philips Semiconductors, IMO BE Innovation, Nijmegen, The Netherlands*
[c]*Philips Semiconductors/CTO/Technology Partnership Office, Eindhoven, The Netherlands*

Abstract In computational prototyping of electronic packages an appropriate description of the mechanical behavior of polymers being included is required. An overview of presently available material models is presented. In particular a cure dependent linear-viscoelastic model is discussed more in detail. With this model the investigation of processing induced stress fields during and after fabrication is possible.

1.1. INTRODUCTION

Among various materials, polymers are widely used in microelectronics as different product constituents, such as encapsulants, conductive or non-conductive adhesives, underfills, molding compounds, insulators, dielectrics, and coatings. The behavior of these polymer constituents determines the performance, such as functionality and reliability, of the final products. Therefore, the successful development of microelectronics depends to some extend, on the optimal design and processing of polymer materials. Due to the development trends of microelectronics, characterized mainly by ongoing miniaturization down to nano scale, technology and functionality integration, eco-designing, shorter-time-to-market, the development and application of polymers becomes one of the bottlenecks for the microelectronic industry. With the development and introduction of new packaging materials there are many new requirements to packaged device reliability. Most of these new materials require extensive characterization, due to the lack of historical reliability data. As such, the new chemistries of these materials are linked to the eventual reliability performance of the package. Even more, in-depth knowledge of the relevant failure mechanisms, coupled with knowledge of its chemistry will be required to successfully introduce the materials and bring new package technologies to the marketplace.

Aiming at optimizing the product/process development, much effort is directed to understanding and designing polymer behavior in microelectronics, such as in material preselection, processing, characterization and modeling. Although these efforts are necessary, the ultimate benefits can only be realized if the relationship between chemistry and the behavior can be understood, predicted and modeled. When appropriate thermal-mechanical models are available, simulations of product fabrication steps and subsequent testing can be performed. On the basis of simulation results products can subsequently be optimized prior to actual product fabrication and prototyping (Zhang et al. [26–28]).

Polymers are characterized by a strongly temperature and time dependent mechanical behavior, combined with a relatively high thermal expansion. Generally the behavior is visco-elastic, such that combined phenomena of creep and relaxation occur in packages at various levels.

In the sequence, an overview of presently available constitutive models for the description of thermal-mechanical behavior of polymers is presented, together with appropriate parameter characterization methods. When giving this overview of polymer material modeling and characterization, it is realized that many models, from quite simple to very complicated, are presently available. For electronics packaging simulations the moderately simple linear visco-elastic models appear to be appropriate in most cases. Therefore, the discussion will be mainly focused to linear visco-elastic behavior.

It is realized that visco-elasticity-based models appear to be appropriate for thermal cycling simulations, provided that the processing induced stress fields during and after fabrication are adequately established. The residual stress field is merely due to chemical shrinkage and simultaneous stiffness built-up in packaging polymers during the curing process, and afterwards the cooling down phase. The levels of processing induced stress can seriously influence the stress fields under operating conditions and thus affect the critical states of stress and deformation. Therefore, a major part of this work is focussed on cure-temperature- and time-dependent modeling and characterization of packaging polymers. The article ends with some results of most recent effort to establish the links between chemical details of the polymers and microelectronics reliability.

1.2. POLYMERS IN MICROELECTRONICS

According to general thermal mechanical characteristics in relation to their usual application temperature, polymers often are classified as: thermoplastic polymers, thermosetting polymers and rubbers.

Thermoplastic polymers consist of long, linear chains and are processed by heating them to the melt stage, bringing them in the desired shape followed by a cooling stage in which the shape is frozen-in. This process does not involve chemical reactions and is thus reversible: the polymer product can be melted and reprocessed to a different shape. Thermoplastic polymers with a randomly arranged molecular structure do not crystallize upon cooling but form an *amorphous*, glassy solid. Typical examples are polystyrene and polycarbonate (safety glass). The temperature below which these polymers show solid-like, glassy behavior is called the glass transition temperature, T_g. Note that although the glass transition is often characterized by a single temperature, the process of liquid-to-glass transition takes place in a temperature range of typically 20 to 30 degrees. Another group of thermoplastic polymers crystallize upon cooling: the macromolecular chains form groups of highly oriented clusters. The size and amount of clusters depends on the molecular structure as well as on the presence of nuclei and on the cooling conditions. The crystallization is never perfect and crystalline clusters are alternated with amorphous regions.

This group of thermoplastic polymers is therefore called *semi-crystalline*. Semi-crystalline polymers have both a melting temperature and a glass transition temperature. Below the melting temperature of the crystalline clusters the amorphous regions are still in the rubbery state and can easily deform. The macroscopic deformation is restricted by the crystalline regions which act as physical crosslinks. Only after further cooling the molecular motion in the amorphous regions freezes and the material shows a glass-like behavior. Semi-crystalline polymers with application temperatures between the crystalline melting point and the glass transition temperature show a soft solid-like behavior. Examples are polyethylene and polypropylene. Other semi-crystalline polymers are in the glassy state at room temperature. PET (polyethylene terephthalate), mainly used for soft drink bottles, is an example of this class.

In *thermosetting polymers* all macromolecular chains are connected by chemical bonds and form a large, three dimensional network. A thermoset product therefore in fact consists of one, big molecule. The crosslinking may occur either during or after the growth of the macromolecular chain segments. The in-situ reaction of epoxy and polyester resins are examples of the fist type and the chemical crosslinking after shaping of rubber parts or the radiation crosslinking of polyethylene are examples of the second type. Note that thermoset polymers may also partly crystallize upon cooling.

Rubbers are thermosetting polymers with an application temperature above the glass transition temperature and the crystalline melting temperature. Examples are polybutadiene and polyisoprene rubbers (both semi-crystalline). The modulus of crosslinked rubbers depends on the crosslink density. Typical rubber moduli are about 1 MPa which is three orders of magnitude lower than the moduli in the glassy state.

For practical purposes often additives are used within these polymers: for example, for high temperature protection (anti-oxidants), for flame-retardancy, for influencing the thermal expansion, stiffness, strength or fracture toughness and for reducing the price. Additives, however, can make the mechanical modeling much more complicated.

The application of polymer materials in microelectronics is generally focused to thermosetting polymers. We will therefore limit the following discussion to the mechanical behavior of these thermosetting materials.

1.2.1.1. Molding Compound The chemistry of widely used molding compounds can be described by a combination of different building blocks with epoxy and hydroxyl reactive groups. Phenol novolac- and cresol novolac-based resins and hardeners are common, but also newcomers such as biphenyl-, multi-aromatic- and DCPD-based precursors and mixtures thereof are being used regarding "environmentally green" and/or very good MSL-performing materials when it comes to 260°C reflow soldering conditions. Moulding compounds are complex mixtures of epoxy resin(s), hardener(s), accelerator(s), high filler loadings, adhesion promotors, release agents, flow additives, carbon black, ion trapping agents, stress absorbers, flame retarders, etc.

With advanced packages, e.g., warpage control is becoming more and more important. Properties to control this warpage are many, such as the E-modulus, coefficients of thermal expansion, position of T_g, but also coefficients of moisture expansion and curing shrinkage. Recent examples of computer simulations including curing shrinkage are given in (Yang et al. [22,23]) and show that with appropriate modeling of curing shrinkage we can get better predictability in subsequent reliability tests such as thermal cycling. It is predicted that warpage caused by the curing process accounts for about 30 to 50% of the total warpage for different map mould configurations.

Another important subject is the moisture ingress in packages and the consequences for warpage and failures during accelerated moisture tests (Gils et al. [8]). Moisture induced damages form one of the most important failure mechanisms within microelectronic products (Fan et al. [6], Wong [30]). The material chemistry is not only linked to its wet properties but also to the strength of interfaces with adjacent materials. The effect of moisture and delamination on the warpage and reliability for moisture sensitive packages, e.g., Ball Grid Arrays (BGA) needs more insight in the relevant mechanisms related to material chemistry.

1.2.1.2. Underfill Underfill polymers are often used to compensate mechanical stresses caused by thermal mismatch between the silicon chip and, e.g., an organic substrate. Furthermore, underfill protects the device against corrosion of, e.g., aluminum structures and against fast moisture penetration.

In [1] various different underfills have been evaluated, as well as different curing schedules with their effect on the reliability. Although being a relatively old example, it still is valid and illustrates the importance of the knowledge of the gelation point, and the influence of the chemical curing shrinkage on the reliability in terms of temperature cycling and accelerated moisture tests. Another example (Yang et al. [3,20,21]) shows a parameter sensitivity study of cure-dependent underfill properties on Flip Chip failures.

1.2.1.3. Liquid Encapsulant As an example a low-stress encapsulant for Hyperred Light Emitted Diodes's (LED) is discussed in [29]. The correlation between chemistry and final optical and mechanical properties of the encapsulant, plus their influence on the lifetime of the LED crystal, is elucidated. The basic chemistry is an epoxy resin/anhydride hardener system. Nevertheless, there are a lot of differences in these chemical building blocks, as well as in the variations of the combinations of ingredients. Basically 5 different reactions can occur between an epoxy and an anhydride. Factors that influence the occurrence of these reactions and their mutual dependence, and thus the formation of the polymer network and hence their mechanical properties, are as follows:

- Type of epoxy/accelerator and their mixing ratio.
- Curing temperature and curing profile.
- Mixing quality and presence of hydroxyl groups and moisture.

With a thorough knowledge of the chemistry, combined with accurate processing, it is possible not only to understand the different chemical reactions, but also to control them. Within limits, this means that the physical properties can be tailored to the application.

1.3. BASICS OF VISCO-ELASTIC MODELING

1.3.1. Preliminary: State Dependent Viscoelasticity

Before discussing the typical modeling of thermal-mechanical behavior of polymers, first some simple "illustrative" characteristic mechanical models are discussed. These are mechanical models, made up of "springs" and "dashpots," sometimes combined with "friction elements" which often are used to illustrate simple 1D material responses: well known are the Kelvin element for the illustration of "creep," the Maxwell element for illustration of stress relaxation (Figure 1.1) and the Burgers model for a combination of both phenomena. For the understanding of subsequent modeling, some characteristics of the Maxwell element (a spring and dashpot in line) are shown.

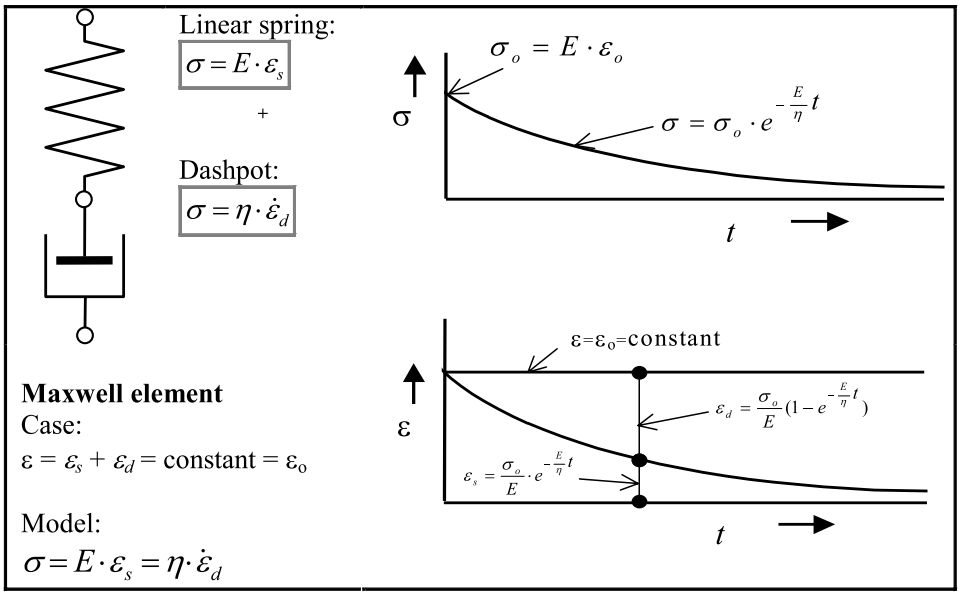


FIGURE 1.1. Illustration of the Maxwell-element.

A constant strain, being applied at time 0, induces an initial stress σ_0, which subsequently is relaxed with time. At time 0 the deformation is completely located within the spring, while subsequently the dashpot is taking over part of the deformation, ending up with zero deformation in the spring at time infinity. The stress relaxation is described by a decaying exponential function:

$$\sigma = \sigma_0 \cdot e^{-\frac{E}{\eta}t} = \sigma_0 \cdot e^{-t/\tau}. \tag{1.1}$$

Here τ in the exponent is the so-called "relaxation time." This is the time to obtain a 37% stress reduction. For the presentation of generalized 3D-models first vectors of stress, strain and initial strain components are introduced.

$$\text{State of stress:} \qquad [\mathbf{S}]^T = [\sigma_{11}\sigma_{22}\sigma_{33}\sigma_{12}\sigma_{23}\sigma_{31}] = [S_1 S_2 S_3 S_4 S_5 S_6]. \tag{1.2}$$

$$\text{State of deformation:} \qquad [\mathbf{E}]^T = [\varepsilon_{11}\varepsilon_{22}\varepsilon_{33}\varepsilon_{12}\varepsilon_{23}\varepsilon_{31}] = [E_1 E_2 E_3 E_4 E_5 E_6]. \tag{1.3}$$

Initial state of deformation (= "stress free" state of an infinitesimal volume):

$$[\mathbf{E}^*]^T = \left[\varepsilon_{11}^*\varepsilon_{22}^*\varepsilon_{33}^*\varepsilon_{12}^*\varepsilon_{23}^*\varepsilon_{31}^*\right] = \left[E_1^* E_2^* E_3^* E_4^* E_5^* E_6^*\right]. \tag{1.4}$$

Note that $\mathbf{E}^*$ is in fact the sum of all volumetric effects (thermal and moisture expansion, reaction shrinkage, physical aging, etc.). The term $\mathbf{E} - \mathbf{E}^*$ can thus be regarded as the effective strain. Now small increments in the state of deformation ΔE_j and in the initial state of deformation ΔE_j^* are assumed at a so-called "load application time" ξ_0 (Figure 1.2). (The actual application of this load increment is assumed to take place in an infinitesimal time increment $\Delta\xi$.)

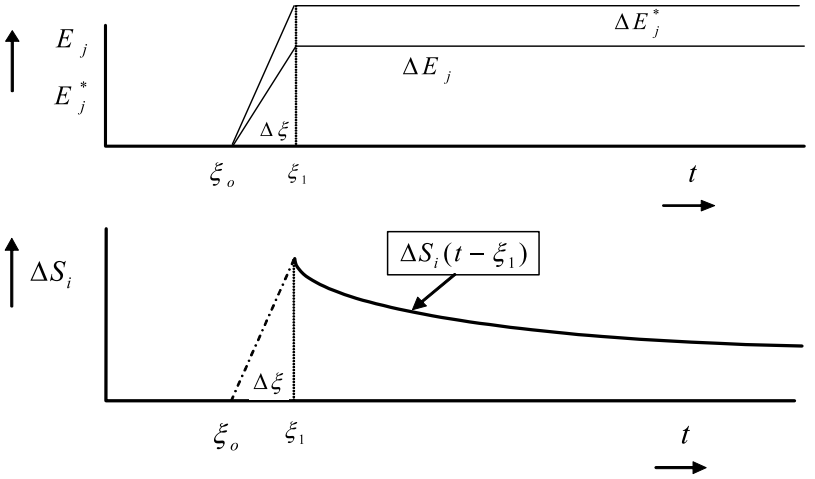


FIGURE 1.2. Step relaxation response.

These strain increments will result into stress increments, with initial values ΔS_i, which subsequently relax with time $(t - \xi_1)$. (Note that ξ_0 and ξ_1 will be the same for an infinitesimal time increment.) The stress relaxation can formally be described as follows:

$$\Delta S_i(t-\xi_1) = C_{ij}(t-\xi_1)\cdot\left(\Delta E_j - \Delta E_j^*\right). \tag{1.5}$$

Here C_{ij} represent so-called *relaxation modulus functions*. It should be noted that we could straightforwardly establish these functions by an experiment, only in case that the strains and initial strains are not continuously changing.

Further it might be so, that the initial stress increment as well as the subsequent relaxation depends on "the stress level" present before application of the strain increment. In that case the relaxation modulus functions should be considered as stress-level dependent, formally noted by:

$$C_{ij} = C_{ij}[\sigma(\xi_1), (t-\xi_1)]. \tag{1.6}$$

Let us now consider continuously changing states of deformation and initial deformation. Here the strains and initial strains are functions of the "load application time" ξ. They can be approximated by considering "load increments (= strain level increments)" occurring over discrete load-application-time increments $\Delta\xi$, so that subsequent "load application times" are:

$$\xi_1 = \xi_0 + \Delta\xi, \xi_2 = \xi_0 + 2\Delta\xi, \ldots,\ \xi_k = \xi_0 + k\Delta\xi, \ldots,\ \xi_n = \xi_0 + n\Delta\xi. \tag{1.7}$$

Corresponding deformation and initial deformation increments than are:

$$\begin{aligned}\Delta E_j^1 &= \left(E_{j,\xi}\right)_{\xi_0}\Delta\xi, \Delta E_j^2 = \left(E_{j,\xi}\right)_{\xi_1}\Delta\xi, \ldots, \Delta E_j^k \\ &= \left(E_{j,\xi}\right)_{\xi_{k-1}}\Delta\xi, \ldots, \Delta E_j^n = \left(E_{j,\xi}\right)_{\xi_{n-1}}\Delta\xi,\end{aligned} \tag{1.8}$$

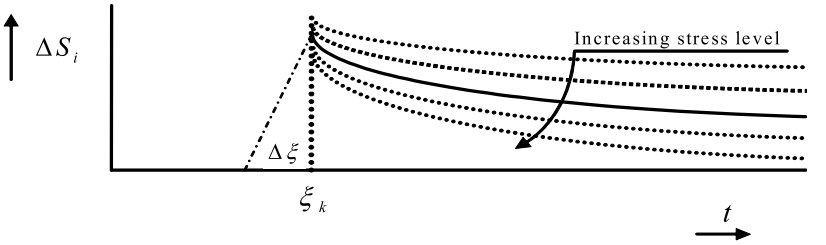


FIGURE 1.3. Stress responses: effect of increasing stress level.

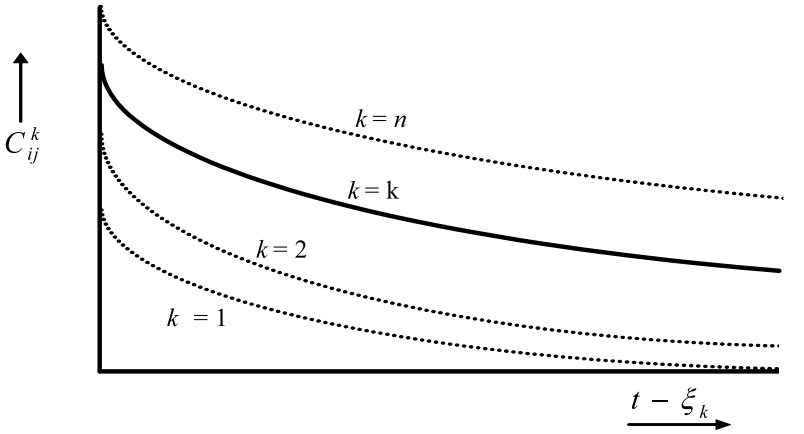


FIGURE 1.4. Effect of changing material properties on the relaxation moduli.

$$\Delta E_j^{*1} = \left(E_{j,\xi}^*\right)_{\xi_0}\Delta\xi, \Delta E_j^{*2} = \left(E_{j,\xi}^*\right)_{\xi_1}\Delta\xi, \ldots, \Delta E_j^{*k}$$
$$= \left(E_{j,\xi}^*\right)_{\xi_{k-1}}\Delta\xi, \ldots, \Delta E_j^{*n} = \left(E_{j,\xi}^*\right)_{\xi_{n-1}}\Delta\xi. \tag{1.9}$$

For load-application time, number k (Figure 1.3), the stress increment-function is:

$$\Delta S_i^k(t-\xi_k) = C_{ij}^k(t-\xi_k)\cdot\left[\left(E_{j,\xi}\right)_{\xi_{k-1}} - \left(E_{j,\xi}^*\right)_{\xi_{k-1}}\right]\cdot\Delta\xi. \tag{1.10}$$

The curing reaction does not only change the initial deformation term $\mathbf{E}^*$ (cure shrinkage effect) but will also change the shape of the relaxation curve itself (Figure 1.4). Therefore the relaxation modulus function also depends on state functions $x^l(\xi_k)$, such as temperature, humidity or degree of cure:

$$C_{ij}^k = C_{ij}\left[x^l(\xi_k), \sigma(\xi_k), (t-\xi_k)\right]. \tag{1.11}$$

The stress state at time t is now found by summing the incremental contributions of all infinitesimal strain increments, yielding following description for stress- and state-dependent viscoelasticity:

$$S_i(t) = \int_{\xi=-\infty}^{t} C_{ij}\big[x^l(\xi), \sigma(\xi), (t-\xi)\big] \cdot \Big[\big(E_{j,\xi}\big)_\xi - \big(E^*_{j,\xi}\big)_\xi\Big] d\xi. \tag{1.12a}$$

Alternatively we can use the following (tensor-) component expression:

$$\sigma_{ij}(t) = \int_{\xi=-\infty}^{t} C_{ijkl}\big[x^l(\xi), \sigma(\xi), (t-\xi)\big] \cdot \Big[\big(\varepsilon_{kl,\xi}\big)_\xi - \big(\varepsilon^*_{kl,\xi}\big)_\xi\Big] d\xi. \tag{1.12b}$$

(In the above expressions and in coming expressions the subscript with comma is used to denote differentiation.)

It should be noted, that so far the relaxation curves are assumed to depend on the "load set-up time" only. Later, in Section 1.5.2 it is discussed that the actual curves could also depend on state variable changes in the time interval from load application to the current time t. The present description is referred to as "***partly state dependent***," whereas the model discussed in Section 1.5.2.2 will be referred to as "***fully state dependent***."

With these formulations the state of stress at time t as the resulting response on a prescribed history of strain and initial strain and state variables $x^l(\xi)$ is defined. Such a description will be referred to as a "relaxation description" of viscoelastic behavior. An alternative formulation that can be constructed analogously, gives the state of deformation at time t as the result of a prescribed history of stress and initial stress (= stress at deformation 0) and state variables $x^l(\xi)$:

$$E_i(t) = \int_{\xi=-\infty}^{t} J_{ij}\big[x^l(\xi), \sigma(\xi), (t-\xi)\big] \cdot \Big[\big(S_{j,\xi}\big)_\xi - \big(S^*_{j,\xi}\big)_\xi\Big] d\xi. \tag{1.13a}$$

Or, alternatively:

$$\varepsilon_{ij}(t) = \int_{\xi=-\infty}^{t} J_{ijkl}\big[x^l(\xi), \sigma(\xi), (t-\xi)\big] \cdot \Big[\big(\sigma_{kl,\xi}\big)_\xi - \big(\sigma^*_{kl,\xi}\big)_\xi\Big] d\xi. \tag{1.13b}$$

Here J_{ij} (or J_{ijkl}) represent the so called creep compliance functions. The description according to (1.13a, b) is often referred to as a "creep description." In principle, the creep compliance functions and the relaxation modulus functions are related (see also Section 1.3.5). In the sequence we will use a "relaxation description" rather than a "creep description" with a single exception in Section 1.4.4.2.

1.3.2. Incremental Relationship

In order to be able to actually investigate processing induced stress fields in electronic packages the state dependent constitutive Equations (1.12a, b) should be implemented into a standard FEM package. Most standard FEM packages facilitate a simple implementation of incremental stress–strain relations (or rate equations) in user subroutines. According to the constitutive Equation (1.12a) a stress-update ΔSi for a time step Δt between time t and time $\hat{t} = t + \Delta t$ is defined by:

$$\Delta Si = Si(\hat{t}) - Si(t), \tag{1.14}$$

with

$$S_i(t) = \int_{\xi=-\infty}^{t} C_{ij}\big[x^l(\xi), \sigma(\xi), (t-\xi)\big] \cdot \Big[\big(E_{j,\xi}\big)_\xi - \big(E^*_{j,\xi}\big)_\xi\Big] d\xi, \tag{1.15}$$

and

$$S_i(\hat{t}) = \int_{\xi=-\infty}^{\hat{t}} C_{ij}\big[x^l(\xi), \sigma(\xi), (\hat{t}-\xi)\big] \cdot \Big[\big(E_{j,\xi}\big)_\xi - \big(E^*_{j,\xi}\big)_\xi\Big] d\xi. \tag{1.16}$$

Note that $\hat{t}$ is not only present in the upper bound of integral (1.16) but also in the kernel function. This reflects the change in relaxation behavior due to cure. Consequently ΔS_i can not simply be written as an integral with bounds t and $\hat{t}$:

$$\Delta S_i = S_i(\hat{t}) - S_i(t) \neq \int_{\xi=t}^{\hat{t}} C_{ij}\big[x^l(\xi), \sigma(\xi), (t-\xi)\big] \cdot \Big[\big(E_{j,\xi}\big)_\xi - \big(E^*_{j,\xi}\big)_\xi\Big] d\xi. \tag{1.17}$$

In an incremental iterative FEM solution the stress $S_i(t)$ represents the equilibrium solution for time t. Then for the calculation of ΔS_i according to (1.17), the convolution integral (1.16) should be evaluated. Since time $\hat{t}$ $(= t + \Delta t)$ is present in the upper bound of the integral, as well as in the kernel, the integral must be evaluated for each time step (and actually also for each iteration). For this evaluation the whole history of stress and strain is involved and should be kept available for each time step. The amount of data is thus progressively increasing with each successive time step, which in turn would object the practical application of the model within a finite element environment.

Hence, an approximate method is sought to circumvent these huge data storage and data handling problems. A more accessible manner for the integration problem was achieved in [16] for an analogous integration problem in the description of time-dependent behavior of rubber-like materials. He employed a (very) limited number of strain history data, by adopting a so-called multi-points approach. Here the kernel of the integral was approximated by a polynomial through a limited number of time points. In this method the selection of time points for which strain data have to be stored and, for the current time step, the actual selection of time points for the polynomial representation (of the kernel), requires a good engineering judgment. The choices made may affect the results of integration and thus can influence the (converged) solutions. Another approximate method was employed in [24,25] in a study on non-linear viscoelasticity of polyester resins and glass fiber reinforced composites. Here the data storage and data handling problem due to the evaluation of the convolution integral being involved was circumvented or reduced by approximating the kernel function with a Prony series. For the present state dependent model such an approximation can also successfully be employed. Here the following Prony series approximation of the state-dependent relaxation moduli functions can be adopted:

$$C_{ij}\big[x^l(\xi), \sigma(\xi), (t-\xi)\big] \approx \sum_{n=1}^{N} C^n_{ij}\big[x^l(\xi), \sigma(\xi)\big] \cdot e^{-(t-\xi)/\tau_n}. \tag{1.18}$$

In principle, the participation functions $C^n_{ij} = C^n_{ij}[x^l(\xi), \sigma(\xi)]$ and the so-called relaxation times (τ_n) can directly be fitted on appropriate experimental data. Substitution of this Prony

series approximation into the integrals (1.15) and (1.16) and applying some rearrangements results into:

$$S_i(t) = \sum_{n=1}^{N} \left\{ e^{-t/\tau_n} \int_{\xi=-\infty}^{t} C_{ij}^n \left[x^l(\xi), \sigma(\xi) \right] \cdot e^{\xi/\tau_n} \cdot \left[\left(E_{j,\xi} \right)_\xi - \left(E_{j,\xi}^* \right)_\xi \right] d\xi \right\}, \tag{1.19}$$

and

$$S_i(\hat{t}) = \sum_{n=1}^{N} \left\{ e^{-\hat{t}/\tau_n} \int_{\xi=-\infty}^{\hat{t}} C_{ij}^n \left[x^l(\xi), \sigma(\xi) \right] \cdot e^{\xi/\tau_n} \cdot \left[\left(E_{j,\xi} \right)_\xi - \left(E_{j,\xi}^* \right)_\xi \right] d\xi \right\}. \tag{1.20}$$

Substituting $\hat{t} = t + \Delta t$ into Equation (1.20) yields:

$$S_i(\hat{t}) = \sum_{n=1}^{N} \left\{ e^{-(t+\Delta t)/\tau_n} \int_{\xi=-\infty}^{(t+\Delta t)} C_{ij}^n \left[x^l(\xi), \sigma(\xi) \right] \cdot e^{\xi/\tau_n} \cdot \left[\left(E_{j,\xi} \right)_\xi - \left(E_{j,\xi}^* \right)_\xi \right] d\xi \right\}. \tag{1.21}$$

Or equivalently:

$$S_i(\hat{t}) = \sum_{n=1}^{N} \left\{ e^{-\Delta t/\tau_n} \cdot e^{-t/\tau_n} \int_{\xi=-\infty}^{(t+\Delta t)} C_{ij}^n \left[x^l(\xi), \sigma(\xi) \right] \cdot e^{\xi/\tau_n} \cdot \left[\left(E_{j,\xi} \right)_\xi - \left(E_{j,\xi}^* \right)_\xi \right] d\xi \right\}. \tag{1.22}$$

Splitting the integral into two parts yields:

$$\begin{aligned} S_i(\hat{t}) = & \sum_{n=1}^{N} \left\{ e^{-\Delta t/\tau_n} \cdot e^{-t/\tau_n} \int_{\xi=-\infty}^{t} C_{ij}^n \left[x^l(\xi), \sigma(\xi) \right] \cdot e^{\xi/\tau_n} \cdot \left[\left(E_{j,\xi} \right)_\xi - \left(E_{j,\xi}^* \right)_\xi \right] d\xi \right\} \\ & + \sum_{n=1}^{N} \left\{ e^{-\Delta t/\tau_n} \cdot e^{-t/\tau_n} \int_{\xi=t}^{(t+\Delta t)} C_{ij}^n \left[x^l(\xi), \sigma(\xi) \right] \cdot e^{\xi/\tau_n} \cdot \left[\left(E_{j,\xi} \right)_\xi - \left(E_{j,\xi}^* \right)_\xi \right] d\xi \right\}. \end{aligned} \tag{1.23}$$

Equations (1.19) is now alternatively formulated by introducing the time function $\Re^n = \Re^n(t)$:

$$S_i(t) = \sum_{n=1}^{N} \Re^n(t), \tag{1.24}$$

with

$$\Re^n(t) = e^{-t/\tau_n} \int_{\xi=-\infty}^{t} C_{ij}^n \left[x^l(\xi), \sigma(\xi) \right] \cdot e^{\xi/\tau_n} \cdot \left[\left(E_{j,\xi} \right)_\xi - \left(E_{j,\xi}^* \right)_\xi \right] d\xi. \tag{1.25}$$

The stress increment is now found by subtracting Equation (1.24) from Equation (1.25), yielding:

$$\begin{aligned}\Delta S_i &= S_i(\hat{t}) - S_i(t) \\ &= \sum_{n=1}^{N}\left[\left(e^{-\Delta t/\tau_n} - 1\right)\cdot \Re^n(t)\right] \\ &\quad + \sum_{n=1}^{N}\left\{e^{-\Delta t/\tau_n}\cdot e^{-t/\tau_n}\int_{\xi=t}^{(t+\Delta t)} C_{ij}^n(\xi)\cdot e^{\xi/\tau_n}\cdot\left[\left(E_{j,\xi}\right)_\xi - \left(E_{j,\xi}^*\right)_\xi\right]d\xi\right\}.\end{aligned} \tag{1.26}$$

In this stress update expression, the integral does not contain the time $\hat{t}$ $(= t + \Delta t)$ in the kernel function and thus its evaluation does not require any stored data from previous time steps, such as it was the case with the original exact description according to Equations (1.14–1.17). The evaluation can straightforwardly be performed adapting an appropriate time integration scheme. The stress update now only requires stored data from the previous time step such as the time functions $\Re^n(t)$. With a stress update procedure, based on the above expressions being implemented into a FEM program, an adequate simulation possibility for establishing curing induced stress and strain fields is obtained. This stress update only requires stored data for the functions $\Re^n(t)$ of the previous time step and can easily be implemented into a FEM program.

1.3.3. Linear State Dependent Viscoelasticity

In order to reduce the complexity in application and in experimental investigation of the modulus functions, first of all it is assumed that the stress level dependency can be neglected. This often can be justified as the polymer-stress levels in electronic packages remain relatively low (for $\sigma < 0.4 \times$ [ultimate strength], the material is generally observed to behave linear viscoelastic). Without stress-level dependency we obtain so-called linear-state dependent viscoelasticity. The convolution integral Equations (1.12a, b) and (1.13a, b) thus are simplified:

"Stress relaxation" description

$$S_i(t) = \int_{\xi=-\infty}^{t} C_{ij}\left[x^l(\xi), (t-\xi)\right]\cdot\left[\left(E_{j,\xi}\right)_\xi - \left(E_{j,\xi}^*\right)_\xi\right]d\xi, \tag{1.27a}$$

$$\sigma_{ij}(t) = \int_{\xi=-\infty}^{t} C_{ijkl}\left[x^l(\xi), (t-\xi)\right]\cdot\left[\left(\varepsilon_{kl,\xi}\right)_\xi - \left(\varepsilon_{kl,\xi}^*\right)_\xi\right]d\xi, \tag{1.27b}$$

$$E_i(t) = \int_{\xi=-\infty}^{t} J_{ij}\left[x^l(\xi), (t-\xi)\right]\cdot\left[\left(S_{j,\xi}\right)_\xi - \left(S_{j,\xi}^*\right)_\xi\right]d\xi, \tag{1.28a}$$

$$\varepsilon_{ij}(t) = \int_{\xi=-\infty}^{t} J_{ijkl}\left[x^l(\xi), (t-\xi)\right]\cdot\left[\left(\sigma_{kl,\xi}\right)_\xi - \left(\sigma_{kl,\xi}^*\right)_\xi\right]d\xi. \tag{1.28b}$$

Assuming linear viscoelasticity, in Section 1.4 the modeling and characterization of fully cured polymers is discussed. Here the temperature is the only state variable taken into

account. The generally applied principle of time–temperature superposition is discussed in Section 1.4.3. Various experimental methods for the characterization of the modulus functions are discussed in Sections 1.4.2–1.4.5.

In Section 1.5 the "partly state dependent" modeling for curing polymer materials is discussed, together with appropriate experimental methods for the characterization of the modulus functions, the curing shrinkage and the reaction kinetics.

Section 1.6 finally describes the theory of so-called "fully state dependent" modeling for curing polymers.

1.3.4. Isotropic Material Behavior

For isotropic materials the relaxation modulus functions $C_{ij}[x^l(\xi), (t-\xi)]$ have only two independent components: the bulk- and the shear-relaxation moduli $K[x^l(\xi), (t-\xi)]$ and$G[x^l(\xi), (t-\xi)]$. The relaxation modulus function C_{ij} can be expressed as

$$C_{ij}\big[x^l(\xi), (t-\xi)\big] = G\big[x^l(\xi), (t-\xi)\big] \cdot D_{ij} + K\big[x^l(\xi), (t-\xi)\big] \cdot V_{ij}, \tag{1.29}$$

where: $V_{ij} = 1$ for ($i \in 1, 2, 3$ and $j \in 1, 2, 3$), $V_{ij} = 0$ for ($i \in 1, 2, 3$ and $j \in 4, 5, 6$) or ($i \in 1, 2, 3$ and $j \in 1, 2, 3$) or ($i \in 4, 5, 6$ and $j \in 4, 5, 6$), $D_{ij} = 3/4$ for ($i \equiv j \in 1, 2, 3$), $D_{ij} = -2/3$ for ($i \neq j \in 1, 2, 3$), $D_{ij} = 2$ for ($i \equiv j \in 4, 5, 6$), $D_{ij} = 0$ for ($i \neq j \in 4, 5, 6$) or ($i \in 4, 5, 6$ and $j \in 1, 2, 3$) or ($i \in 4, 5, 6$ and $j \in 4, 5, 6$).

For isotropic materials Equations (1.27a, b) are alternatively written as:

"Stress relaxation" description

$$S_i(t) = \int_{\xi=-\infty}^{t} \Big\{ G\big[x^l(\xi), (t-\xi)\big] \cdot D_{ij} + K\big[x^l(\xi), (t-\xi)\big] \cdot V_{ij} \Big\} \times \Big[\big(E_{j,\xi}\big)_\xi - \big(E^*_{j,\xi}\big)_\xi \Big] d\xi, \tag{1.30a}$$

$$\sigma_{ij}(t) = \int_{\xi=-\infty}^{t} \Big\{ 2G\big[x^l(\xi), (t-\xi)\big] \cdot \big(\varepsilon^d_{ij,\xi}\big)_\xi + K\big[x^l(\xi), (t-\xi)\big] \cdot \big(\varepsilon^{\mathit{eff}}_{v,\xi}\big)_\xi \Big\} d\xi. \tag{1.30b}$$

The bulk relaxation modulus function fully defines the relation between volumetric stress and the volumetric strain:

$$\sigma_v(t) = \int_{\xi=-\infty}^{t} \Big\{ K\big[x^l(\xi), (t-\xi)\big] \cdot \big(\varepsilon^{\mathit{eff}}_{v,\xi}\big)_\xi \Big\} d\xi. \tag{1.31}$$

Here $\varepsilon_v^{\mathit{eff}}$ and σ_v represent the effective volumetric strain and the volumetric stress, respectively:

$$\varepsilon_v^{\mathit{eff}} = \sum_{i=1}^{3} [\varepsilon_{ii} - \varepsilon^*_{ii}]; \quad \sigma_v = \frac{1}{3} \sum_{i=1}^{3} \sigma_{ii} = -p. \tag{1.32a, b}$$

The shear relaxation modulus function fully defines the relation between deviatoric stress and deviatoric strain:

$$\sigma^d_{ij}(t) = \int_{\xi=-\infty}^{t} \Big\{ 2G\big[x^l(\xi), (t-\xi)\big] \cdot \big(\varepsilon^d_{ij,\xi}\big)_\xi \Big\} d\xi. \tag{1.33}$$

Here σ_{ij}^{d} and ε_{ij}^{d} denote the deviatoric stresses and deviatoric strains, respectively:

$$\sigma_{ij}^{d} = \sigma_{ij} - \sigma_{v} \cdot \delta_{ij}; \quad \varepsilon_{ij}^{d} = \varepsilon_{ij} - \frac{1}{3}\varepsilon_{v}^{\mathit{eff}}. \tag{1.34a, b}$$

Analogously, for isotropic materials the creep description is formulized through:

"Creep" description

$$E_i(t) = \int_{\xi=-\infty}^{t} \left\{ J\left[x^{l}(\xi), (t-\xi)\right] \cdot C_{ij}^{D} + B\left[x^{l}(\xi), (t-\xi)\right] \cdot C_{ij}^{V} \right\} \cdot \left[\left(S_{j,\xi}\right)_{\xi} - \left(S_{j,\xi}^{*}\right)_{\xi} \right] d\xi, \tag{1.35a}$$

$$\varepsilon_{ij}(t) = \int_{\xi=-\infty}^{t} \left\{ \frac{1}{2} J\left[x^{l}(\xi), (t-\xi)\right] \cdot \left(\sigma_{ij,\xi}^{d}\right)_{\xi} + B\left[x^{l}(\xi), (t-\xi)\right] \cdot \left(\sigma_{v,\xi}^{\mathit{eff}}\right)_{\xi} \right\} d\xi, \tag{1.35b}$$

where:

$$\sigma_{v}^{\mathit{eff}} = \frac{1}{3} \sum_{i=1}^{3} [\sigma_{ii} - \sigma_{ii}^{*}]. \tag{1.36}$$

Here $B[x^{l}(\xi), (t-\xi)]$ and $J[x^{l}(\xi), (t-\xi)]$ are the so-called volumetric creep compliance and the deviatoric creep compliance, respectively.

Alternatively, the counterparts of (1.33) and (1.35) for a "creep formulation" are:

$$\varepsilon_{v}(t) = \int_{\xi=-\infty}^{t} \left\{ B\left[x^{l}(\xi), (t-\xi)\right] \cdot \left(\sigma_{v,\xi}^{\mathit{eff}}\right)_{\xi} \right\} d\xi, \tag{1.37}$$

$$\varepsilon_{ij}^{d}(t) = \int_{\xi=-\infty}^{t} \left\{ \frac{1}{2} J\left[x^{l}(\xi), (t-\xi)\right] \cdot \left(\sigma_{ij,\xi}^{d}\right)_{\xi} \right\} d\xi. \tag{1.38}$$

1.3.5. Interrelations between Property Functions

The interrelation between the shear relaxation modulus $G(t)$ and the deviatoric creep compliance $J(t)$ follows by combining the Laplace transforms of the constitutive equations for shear relaxation and shear creep [Equations (1.33) and (1.38)]:

$$\bar{\sigma}_{ij}^{d}(s) = \overline{G}(s) * s\bar{\varepsilon}_{ij}^{d}(s), \tag{1.39}$$

$$\bar{\varepsilon}_{ij}^{d}(s) = \bar{J}(s) * s\bar{\sigma}_{ij}^{d}(s), \tag{1.40}$$

where the overbar denotes the Laplace transform and s is the Laplace variable. Inserting Equation (1.40) into (1.39) and elimination of σ_{ij}^{d} results in

$$\overline{G}(s) \cdot \bar{J}(s) = \frac{1}{s^{2}}. \tag{1.41}$$

The Laplace inverse is

$$\int_0^t G(\xi) \cdot J(t-\xi) d\xi = t. \tag{1.42}$$

Similarly, interrelations for the bulk relaxation and creep moduli $K(t)$ and $B(t)$ as well as for the tensile relaxation and creep moduli $E(t)$ and $D(t)$ can be derived:

$$\int_0^t K(\xi) B(t-\xi) d\xi = t, \tag{1.43}$$

$$\int_0^t E(\xi) D(t-\xi) d\xi = t. \tag{1.44}$$

A simple iterative scheme can be used to evaluate the above integral relations (see [18]). Suppose the creep compliance $D(t)$ is known from experiments and the corresponding relaxation modulus $E(t)$ is the target function. We start by dividing the integration interval into m subintervals and obtain:

$$\sum_{i=1}^{m} \int_{t_{i-1}}^{t_i} E(\xi) D(t_m - \xi) d\xi = t_m. \tag{1.45}$$

Next the function $E(\xi)$ is approximated as its average value over the subinterval: $(1/2)[E(t_{i-1}) + E(t_i)]$. Application of the trapezium rule for the integral over the subinterval and collecting all $E(t_m)$ terms then results into:

$$E_m = -E_{m-1} + \frac{4t_m - \sum_{i=1}^{m-1} [E(t_i) + E(t_{i-1})] * [D(t_m - t_i) + D(t_m - t_{i-1})](t_i - t_{i-1})}{[D_g - D(t_m - t_{m-1})](t_m - t_{m-1})}. \tag{1.46}$$

As a starting value $E(t_1) = 1/D(t_1)$ should be used, where $D(t_1)$ must be close to the glassy modulus D_g. For the interconversion from a known $E(t)$ to $D(t)$ as well as for the interconversion of Equations (1.42–1.43) the same numerical scheme can be used.

Instead of using the (state dependent) volumetric- and deviatoric-creep compliances as the primary variable functions for creep of isotropic materials, often the so-called tensile creep compliance $D(t-\xi)$ $(= D[x^l(\xi), (t-\xi)])$ and the Poisson's ratio $\nu(t-\xi)$ $(= \nu[x^l(\xi), (t-\xi)])$ are used as primary variable functions. These characteristic mechanical properties of the material can be obtained from the time-dependent longitudinal and lateral strains in a tensile creep test (see also Section 1.4.4) through the following Laplace transform relations:

$$s \cdot \overline{D}(s) = \frac{\mathrm{s} \cdot \bar{\varepsilon}_{11}(s)}{\sigma_0} \quad \text{and} \quad s \cdot \bar{\nu}(s) = \frac{-\mathrm{s} \cdot \bar{\varepsilon}_{22}(s)}{\mathrm{s} \cdot \bar{\varepsilon}_{11}(s)}. \tag{3.47a, b}$$

Here σ_0 represents the applied creep stress (applied at time 0). For the tensile creep test (see Section 1.4.4) the Laplace inverses of these relations are given by:

$$D(t) = \frac{\varepsilon_{11}(t-\xi)}{\sigma_0}, \tag{1.48}$$

$$\varepsilon_{22}(t) = -\varepsilon_{11}(0) \cdot \nu(t) + \int_0^t \nu(\eta) \frac{d\varepsilon_{11}(t-\eta)}{d\eta} d\eta. \tag{1.49}$$

The interrelations between bulk- and shear-relaxation modulus functions and the Young's modulus and Poisson's ratio functions can also be formulated through Laplace transform relations of these property functions:

$$s \cdot \overline{E}(s) = \frac{9s \cdot \overline{G}(s) \cdot \overline{K}(s)}{\overline{G}(s) + 3\overline{K}(s)} \quad \text{and} \quad s \cdot \bar{\upsilon}(s) = \frac{1.5\overline{K}(s) - \overline{G}(s)}{\overline{G}(s) + 3\overline{K}(s)}. \tag{3.50a, b}$$

Iterative schemes for the Laplace inverses of these equations can be obtained in a similar way as described above.

1.3.6. Elastic Approximations

Figure 1.5 shows an illustration of a typical relaxation curve on a log–log scale. Generally we than observe a more or less horizontal plateau for small time values. This plateau is referred to as the "glassy plateau." The corresponding modulus values are the so-called "glassy modulus" values (i.e., G^G). Than we observe a transient part of the modulus curve that actually describes the visco-elastic relaxations. For larger time values, we observe a second more or less horizontal plateau that is referred to as the "rubbery plateau." The corresponding modulus values are the so-called "rubbery moduli" or "equilibrium moduli" (i.e., G^R).

For fast loading conditions, the "glassy moduli" define the instantaneous, elastic stress responses of the material. Therefore, the "glassy moduli" can be considered as "elastic moduli."

The "rubbery moduli" or "equilibrium moduli" actually define the stress state after all relaxations have died out. Sometimes, the actual time dependent relaxation behavior is not taken into account in simplified simulations. Than, as an approximation of reality, that in few cases might be realistic, only elastic simulations are performed, while considering the "rubbery moduli" or "equilibrium moduli" as "elastic moduli."

Thus, it should be noted, that two different kinds of "elastic moduli" can be considered: The "glassy moduli" and the "rubbery moduli" or "equilibrium moduli." Depending

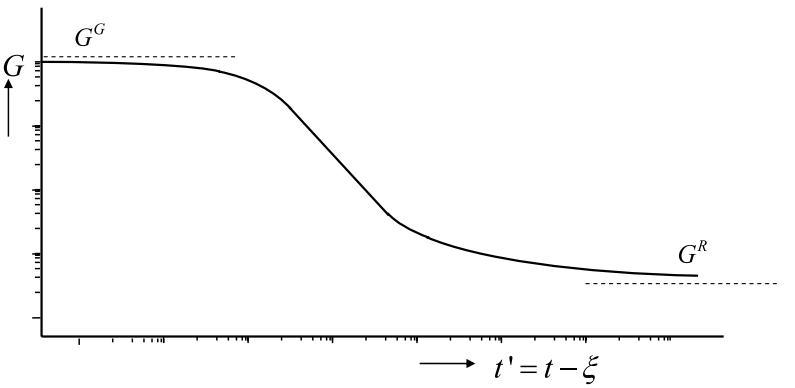


FIGURE 1.5. Typical relaxation curve on a log–log scale.

on the purpose of an elastic simulation (very short time or very large time approximations) the correct choice should be made.

1.4. LINEAR VISCO-ELASTIC MODELING (FULLY CURED POLYMERS)

1.4.1. Introduction

When we exclude moisture influences, for fully cured polymer materials generally the only state parameter (x^l) to be considered is the temperature T. In this case $x^l(\xi) = T(\xi)$ (for $l = 1$). The stress integrals for the "relaxation description" (1.27a, b) thus are simplified to:

$$S_i(t) = \int_{\xi=-\infty}^{t} C_{ij}[T(\xi), (t-\xi)] \cdot \left[\left(E_{j,\xi}\right)_\xi - \left(E^*_{j,\xi}\right)_\xi \right] d\xi, \tag{1.51a}$$

$$\sigma_{ij}(t) = \int_{\xi=-\infty}^{t} C_{ijkl}[T(\xi), (t-\xi)] \cdot \left[\left(\varepsilon_{kl,\xi}\right)_\xi - \left(\varepsilon^*_{kl,\xi}\right)_\xi \right] d\xi \tag{1.51b}$$

with in case of isotropy [see (1.29)]:

$$C_{ij}[T(\xi), (t-\xi)] = K[T(\xi), (t-\xi)] \cdot V_{ij} + G[T(\xi), (t-\xi)] \cdot D_{ij}. \tag{1.52}$$

Equations (1.30b), (1.31) and (1.33) simplify to:

$$\sigma_{ij}(t) = \int_{\xi=-\infty}^{t} \left\{ 2G[T(\xi), (t-\xi)] \cdot \left(\varepsilon^d_{ij,\xi}\right)_\xi + K[T(\xi), (t-\xi)] \cdot \left(\varepsilon^{\mathit{eff}}_{v,\xi}\right)_\xi \right\} d\xi, \tag{1.53}$$

$$\sigma_v(t) = \int_{\xi=-\infty}^{t} \left\{ K[T(\xi), (t-\xi)] \cdot \left(\varepsilon^{\mathit{eff}}_{v,\xi}\right)_\xi \right\} d\xi, \tag{1.54}$$

$$\sigma^d_{ij}(t) = \int_{\xi=-\infty}^{t} \left\{ 2G[T(\xi), (t-\xi)] \cdot \left(\varepsilon^d_{ij,\xi}\right)_\xi \right\} d\xi. \tag{1.55}$$

The "creep formulation" expressions (1.37) and (1.38) are simplified to:

$$\varepsilon_V(t) = \int_{\xi=-\infty}^{t} \left\{ B[T(\xi), (t-\xi)] \cdot \left(\sigma^{\mathit{eff}}_{V,\xi}\right)_\xi \right\} d\xi, \tag{1.56}$$

$$\varepsilon^d_{ij}(t) = \int_{\xi=-\infty}^{t} \left\{ \frac{1}{2} J[T(\xi), (t-\xi)] \cdot \left(\sigma^d_{ij,\xi}\right)_\xi \right\} d\xi. \tag{1.57}$$

1.4.2. Static Testing of Relaxation Moduli

When using the (linear) visco-elastic formulation as given in (1.51–1.55), the material behavior is fully described through the shear- and bulk-relaxation modulus functions,

$G[T(\xi), (t-\xi)]$ and $K[T(\xi), (t-\xi)]$. In selecting experimental characterization methods to establish these functions for a particular polymer, methods from which these functions can straightforwardly be obtained, are in principle most attractive. Depending on the manner of experimental characterization we consider "static testing" and "dynamic testing" procedures. First various "static testing" procedures will be discussed, together with the advantages and disadvantages. Here only tests with relatively steady strain-stress fields will be considered (no bending tests, etc.). Later, in Section 1.4.5 a dynamic procedure that often is referred to as "dynamic mechanical analysis" (DMA) will be discussed.

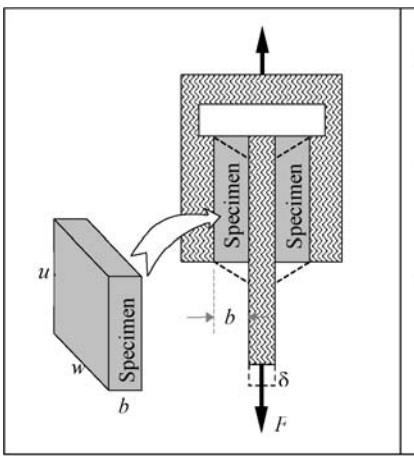
In principle the "static tests" are performed at constant temperature levels and should be repeated at various temperatures to establish the full temperature dependency. However, the number of experiments or the duration can often be reduced by adopting the so-called principle of time–temperature superposition. This method will be discussed in Section 1.4.3.

1.4.2.1. Characterization of the Shear Relaxation Modulus The double simple shear test setup (see Figure 1.6) is suitable for conducting the step-relaxation experiment, as the shear stress and strain distributions are almost constant over the specimen, provided that the specimen thickness (b in Figure 1.6) is small compared to its lateral dimensions (w in Figure 1.6). A (steady) shear deformation is applied while the decaying axial load is measured and used to establish the shear stress relaxation. The relevant description is presented in Figure 1.6.

A disadvantage of this method is that for the required small thickness/lateral dimension ratio, the sample stiffness can be relatively high compared to the tool-machine stiffness. This requires appropriate measures for the control of the constant sample deformation.

Another suitable relaxation experiment can be performed on a cylindrical rod or (thin) cylinder (with circular cross-section), loaded in torsion. A step-torsional deformation is applied while the decaying torsional moment is measured. The relevant description is presented in Figure 1.7.

Simple control of the applied torsional deformation can be performed just by controlling the rotation angles, provided that the sample is relative long such that the torsion stiffness of the sample is low compared to the tool-machine stiffness.



• Step shear deformation:

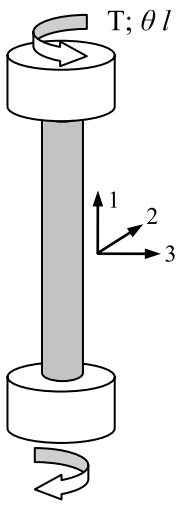
$$\psi_{12} = 2 \cdot \varepsilon_{12}^{d} = 2 \cdot \varepsilon_{12} = \delta / b \qquad (4.8)$$

• Shear stress:

$$\sigma_{12}^{d} = \sigma_{12} = 2 \cdot G[T, t-\xi] \cdot \varepsilon_{12} \qquad (4.9)$$

• Shear load: $F = 2 \cdot u \cdot w \cdot \sigma_{12}$ (4.10)

FIGURE 1.6. Double sandwich simple shear relaxation experiment.



• Step torsional angle: $\theta \cdot l$ (4.11)

• Specific shear angle:

$$\theta = 2 \cdot \varepsilon_{12}^{d} = 2 \cdot \varepsilon_{12} \quad (4.12)$$

• Torsional moment $M_{torsion}$:

$$M_{torsion} = G[T, t-\xi] \cdot \theta \cdot l \cdot I_p \quad (4.13)$$

• Torsional stress near surface:

$$\tau = \sigma_{12}^{d} = \sigma_{12} = G[T, t-\xi] \cdot \theta \cdot r \quad (4.14)$$

FIGURE 1.7. Cylindrical torsional shear relaxation specimen (length l, radius r).

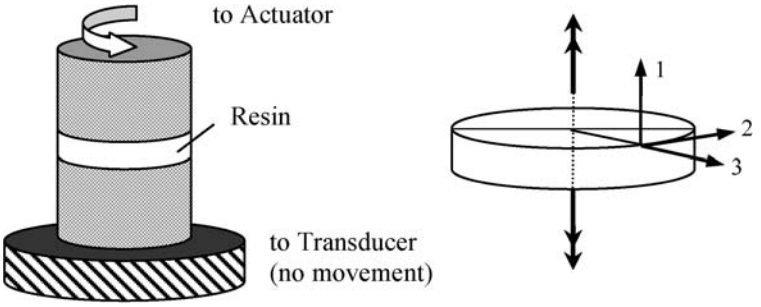


FIGURE 1.8. Coin like torsional shear relaxation specimen (thickness l, radius r).

A variant of the cylindrical rod sample is the coin-like sample as given in Figure 1.8. Here the cylinder length has collapsed to the coin thickness. The torsional stiffness of this sample can be quite high compared to the tool/machine stiffness. Therefore, this kind of sample can only be used if the rotational measurements are performed directly at the transitions between coin sample and tool-plates.

A practical difficulty in executing the relaxation test is that the necessary step deformation generally can not perfectly be realized. Depending on the testing machine and the control, the actual deformation is characterized with an initial transient part before the strain becomes steady. As a result the "theoretical maximum stress" at time ξ is not reached, as illustrated in Figure 1.9. Instead a "stress deviation" occurs at load initiation time ξ. As a consequence the start of the relaxation curve will be inaccurate. In practice this means that the first 0.5 to 1 second of the relaxation curve can not be used.

Another disadvantage is that the duration of relaxation tests can be quite long, in particular for tests at low temperature. However, for so-called "rheologically simple" materials, this problem can be circumvented by just performing short-time testing for various

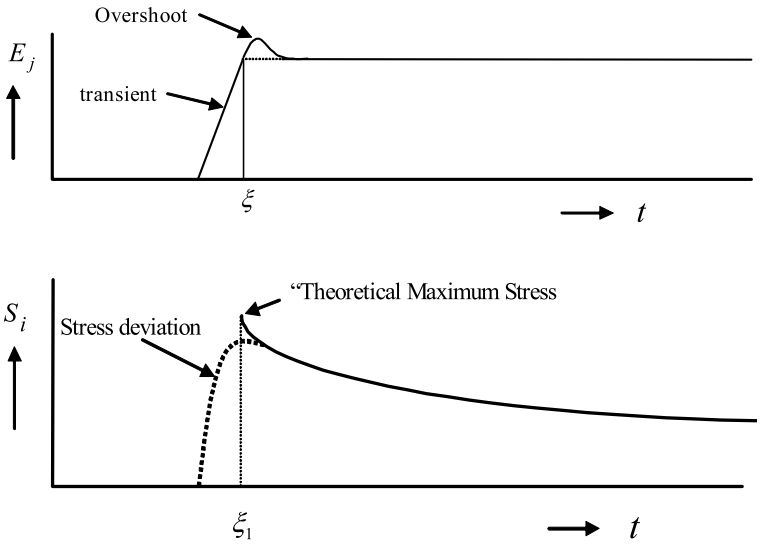


FIGURE 1.9. Step relaxation response.

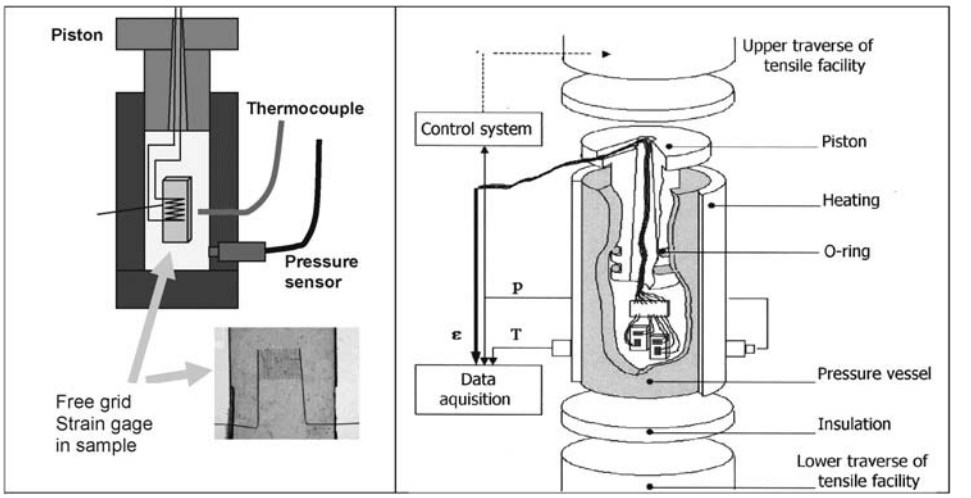
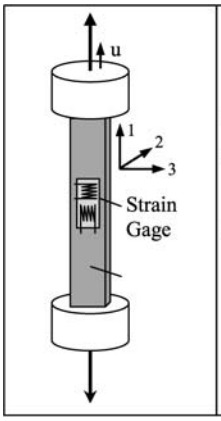


FIGURE 1.10. Pressure cell experiment.

temperatures and applying "time–temperature superposition." This will be explained in Section 1.4.3.

1.4.2.2. Characterization of the Bulk Relaxation Modulus In principle, straightforward measuring the bulk relaxation modulus K could be performed by using a material sample subjected to hydrostatic pressure in a pressure cell (see Figure 1.10) mounted in a tensile testing machine. The strain is measured through free-grid strain gages in the sample. Through a control system the sample deformation should be kept constant, while the hydrostatic stress decay is registered. However, here measurement results could be affected by disturbances due to the seal resistance, when controlling the constant deformation via



• Step axial deformation: u

• Axial strain: $\varepsilon_{11} \approx u/l$ (prescribed, measured with strain gage) (4.15)

• Lateral strains: $\varepsilon_{22}, \varepsilon_{33}$ (ε_{22} measured by strain gage)

• Axial stress: $\sigma_{11} = N/A = E[T, t'] \cdot \varepsilon_{11}$ (measured) (4.16)

FIGURE 1.11. Tensile relaxation test.

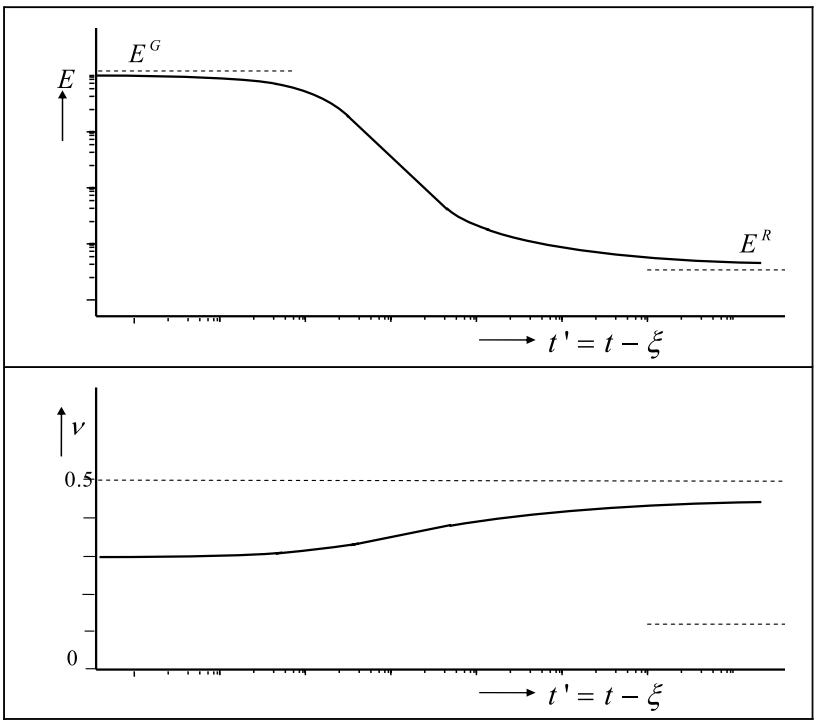


FIGURE 1.12. Schematics of typical relaxation behavior: (a) Young's modulus, (b) Poisson's ratio.

pressure adjustment (via piston movement rather than via adjustment with a pump/valve system). Therefore, in Section 1.4.4.2 the pressure cell experiment is more successfully used in a creep experiment to directly establish the bulk (= volumetric) creep compliance B. Then the bulk relaxation modulus K is indirectly derived by solving the Laplace transform Equation (1.43).

1.4.2.3. Characterization of the Young's Modulus and Poisson's Ratio Straightforward measuring of the relaxation Young's modulus function E and the Poisson's ratio function ν can be performed in a tensile relaxation test. Here a (relatively long) prismatic sample (or dog bone sample) is subjected to a steady elongation, exerting the prescribed longitudinal deformation.

Generally, the axial and lateral strains are measured via (high temperature) strain gages. The axial strain is controlled to be steady. The relaxation curve for the Young's modulus is derived from the measured axial load through (4.16) (Figure 1.11) and has the "general relaxation shape" as discussed in Section 1.3.6. Figures 1.12(a) (b) show the typical relaxation behavior of Young's modulus and Poisson's ratio. The Poisson's ratio ranges from about 0.3 for the glassy state till about 0.5 for the rubbery state.

1.4.3. Time-Temperature Superposition Principle

An important state parameter influencing the viscoelastic behavior is the temperature. The characteristic temperature dependency is best studied by normalizing the relaxation data using the limiting glassy and rubbery modulus values:

$$\hat{C}_{ij}(T, t-\xi) = \frac{C_{ij}(T, t-\xi) - C_{ij}^{R}(T)}{C_{ij}^{G}(T) - C_{ij}^{R}(T)}. \tag{1.58}$$

Figure 1.13 illustrates characteristic relaxation behavior by showing the normalized shear relaxation modulus for various temperatures. For most materials these relaxation curves at different temperatures are all of the same shape and the only effect of temperature turns out to be a shifting along the time axis. When such characteristic behavior appears, the material is said to behave "rheologically simple."

For these materials a so-called "master curve" can be constructed by shifting the individual temperature curves to the curve at a chosen reference temperature. The new time coordinate is denoted as the reduced or effective time and is given as:

$$t_{red}(t') = t' \cdot a_T[T - T_{ref}]. \tag{1.59}$$

Here a_T represents the "temperature shift factor" (as illustrated in Figure 1.14).

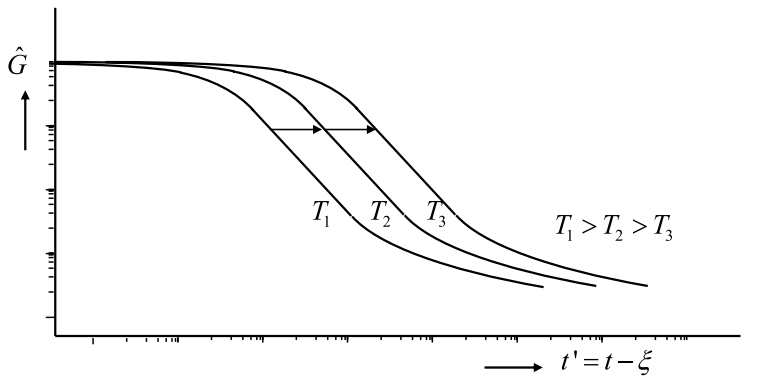


FIGURE 1.13. Schematics of normalized relaxation curves at different temperatures: horizontal shift on (log) time (t') axis.

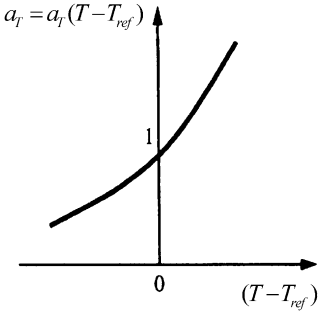


FIGURE 1.14. Schematics of "temperature shift factor."

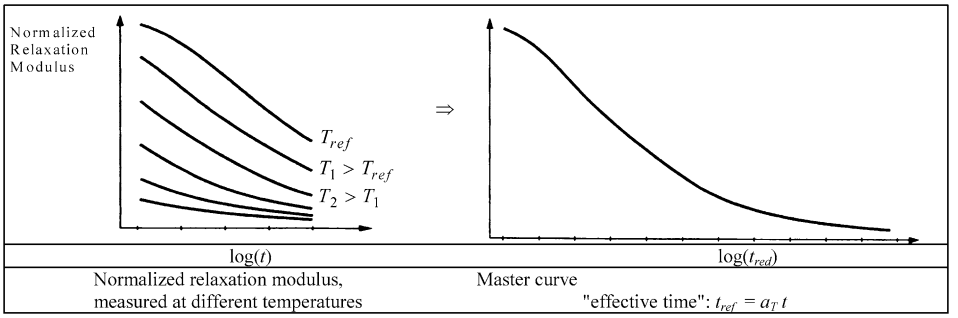


FIGURE 1.15. Schematics of "master curve concept."

A practical advantage of the time–temperature shifting procedure is that it makes it possible to construct the overall master curve from short duration relaxation tests at various temperatures. Each short duration curve is than used to construct part of the master curve by horizontal shifting, as is illustrated in Figure 1.15.

1.4.4. Static Testing of Creep Compliances

When using the viscoelastic Equations (1.35a, b) to describe the creep behavior, the material behavior is fully described through the shear- (= deviatoric) and volumetric creep compliance functions, $J[T(\xi), (t - \xi)]$ and $B[T(\xi), (t - \xi)]$. These compliance functions can straightforwardly be obtained on creep tests with "shear test setups" and a "pressure cell setup" as discussed in Sections 1.4.1 and 1.4.2, for the relaxation experiments. In performing creep tests, the loading will be applied and subsequently kept steady, while the deformations are measured. These creep tests will be discussed in Sections 1.4.4.1 and 1.4.4.2, respectively. In Section 1.4.4.3 the characterization of the "tensile creep compliance" will be discussed. Also the possibility to derive the Young's modulus and the Poisson's ratio from creep test results will be discussed.

1.4.4.1. Characterization of Shear Creep Compliance The (double) simple shear specimen (see Figure 1.6) and the cylindrical torsion specimen (see Figure 1.7) can quite well be used for creep testing to establish the shear- (= deviatoric) creep compliance function,

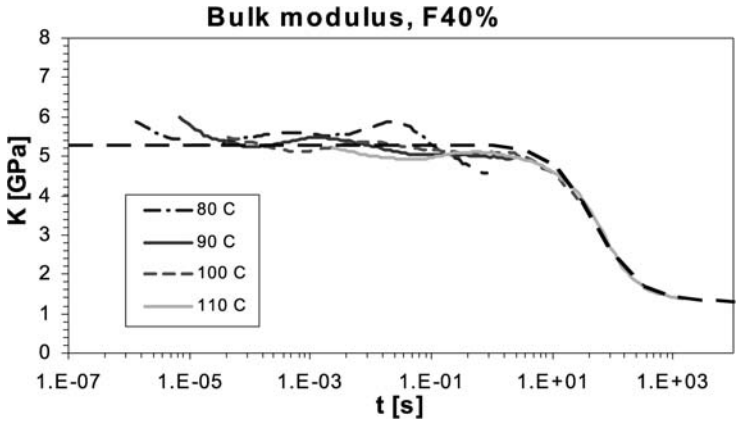


FIGURE 1.16. Bulk-relaxation curves of model molding compound (40% filler) (see Section 1.5.1.2).

$J[T(\xi), (t-\xi)]$. For a creep experiment (rather than a relaxation experiment) the Equations (4.9) and (4.13) are replaced by their creep counterparts:

$$\varepsilon_{12}^{d} = \varepsilon_{12} = \frac{1}{2} \cdot J(T, t-\xi) \cdot \sigma_{12}, \tag{1.60}$$

$$\theta = \frac{1}{2} \cdot J[T, t-\xi] \cdot M_{torsion}/(l \cdot I_p). \tag{1.61}$$

1.4.4.2. Characterization of Volumetric Creep Compliance The pressure cell setup (see Figure 1.10) as described in Section 1.4.2.2 for the characterization of the bulk relaxation modulus is more appropriately used to establish the volumetric creep compliance function $B[T(\xi), (t-\xi)]$. Here the earlier discussed disturbance due to the seal resistance in relaxation testing (when controlling the constant deformation through piston adjustment) is not such an issue here, as now the pressure is controlled to be constant. After having established the volumetric creep compliance function $B[T(\xi), (t-\xi)]$, for various temperatures, the bulk relaxation modulus K can be derived by solving the Laplace transform Equation (1.43). As an example, in Figure 1.15 some bulk-relaxation curves of a model molding compound are shown. Note that the bulk modulus drops from about 5.3 GPa in the glassy state to 1.3 GPa in the rubbery state. This modulus drop is much less that that for the shear and elongation moduli where it can be several orders of magnitude. It should be remarked, that for the curves in Figure 1.16 the Laplace transform relation (1.42) was approximated as:

$$K(t') \cdot B(t') \approx 1. \tag{1.62}$$

1.4.4.3. Characterization of the Tensile Creep Compliance, the Young's Modulus and the Poisson's Effect The creep tensile compliance function $D(\xi, T)$ and the Poisson's ratio function $\nu(\xi, T)$ can also be considered as primary variable functions of isotropic viscoelasticity. These functions can be extracted from a tensile test (as presented in Figure 1.11) however, now loaded under creep conditions. Controlling the axial stress to be steady, the

axial and the lateral deformations, ε_{11} and ε_{22}, are monitored. According to (1.48) the tensile creep compliance is related to the measured longitudinal strain through:

$$D(t') = \varepsilon_{11}(t')/\sigma_0. \tag{1.63}$$

The Poisson's ratio is implicitly related to the experimental strain data, i.e., $\varepsilon_{11}(t')$ and $\varepsilon_{22}(t')$ through (1.49):

$$\varepsilon_{22}(t) = -\varepsilon_{11}(0) \cdot \nu(t) + \int_0^t \nu(\xi) \frac{d\varepsilon_{11}(t-\xi)}{d\xi} d\xi. \tag{1.64}$$

The integral in this expression should be further evaluated to extract the Poisson's ratio function $\nu(t)$. Therefore a numerical interconversion scheme similar to the one described in Section 1.3.5 should be applied. The final result then becomes

$$\nu(t_m) = \Bigg\{ -2\varepsilon_{22}(t_m) + \nu_0[\varepsilon_{11}(t_m - t_1) - \varepsilon_{11}(t_m)] + \sum_{i=1}^{m-1} \nu(t_i) \times [\varepsilon_{11}(t_m - t_{i+1}) - \eta_{11}(t_m - t_{i-1})] \Bigg\} / [\varepsilon_{11}(0) + \varepsilon_{11}(t_m - t_{m-1})]$$

$$(\text{for } m \geq 1). \tag{1.65a}$$

With as a starting value

$$\nu_0 = \nu(0) = \nu(t_0) = \frac{-\varepsilon_{22}(0)}{\varepsilon_{11}(0)}. \tag{1.65b}$$

In [4] the procedure is worked out for an epoxy molding compound. The time–temperature superposition principle is used to construct master curves from isothermal one-day creep experiments at different temperatures (ranging below and above the glass transition temperature of the compound), see Figure 1.17.

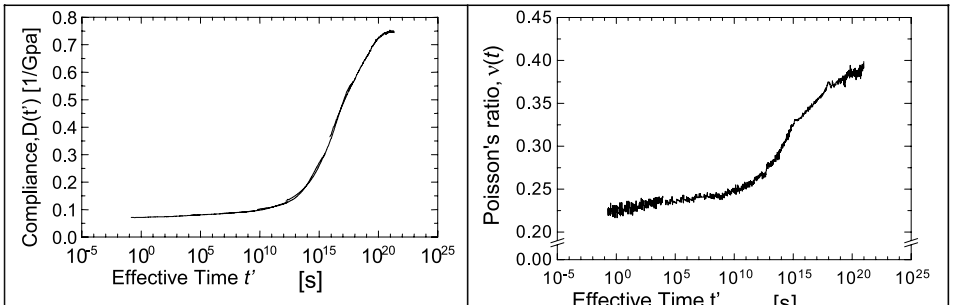


FIGURE 1.17. Typical examples of creep compliance and Poisson's ratio master curves ($T_{ref} = 23°C$).

1.4.5. Dynamic Testing

Another method to experimentally establish the relaxation modulus functions is the so-called "dynamic mechanical analysis" (DMA). Here a sinusoidal deformation is applied to a material sample. After a certain number of deformation periods, the stresses stabilize and than also show a sinusoidal behavior, though shifted in time. The stress amplitude and the time shift appear to be frequency dependent. Applying various test frequencies, a part of the modulus curve can be extracted from the stress amplitudes and time shift, as will be explained in Sections 1.4.5.2 and 1.4.5.3. The type of DMA test depends on the modulus function to be established. Shear DMA tests (torsion or simple shear) are generally used to establish (part of) the shear relaxation modulus curve. Axial DMA testing is often performed to establish the longitudinal relaxation modulus (Young's modulus). Direct DMA tests to establish the relaxation compression modulus curve are hardly feasible. Instead the relaxation compression modulus curve is indirectly obtained from the relaxation Young's modulus and the relaxation shear modulus by applying the interrelations (1.50a, b).

In Section 1.4.5.1 some basics of the theory of dynamic mechanical analyses is discussed. *In the sequence the theory and application is worked out for a simple shear test to establish (part of) the* ***relaxation shear modulus*** *curve. In principle, analogous procedures hold for the* ***longitudinal relaxation modulus****.* Since the applicable frequency range and stiffness of DMA test machines is restricted, generally only parts of relaxation curves can be investigated, which will be discussed in Section 1.4.5.2. For materials that behave "rheologically simple" this problem can be overcome by applying the "frequency temperature superposition" principle. Sufficient data are than obtained from DMA tests with "limited frequency range," but applied for a range of temperatures. This will also be discussed in Section 1.4.5.3.

1.4.5.1. Basic Formulations of Dynamic Mechanical Analysis (DMA) For the application of dynamic mechanical analysis it is necessary to adopt a Prony series approximation of the relaxation modulus functions, such as introduced (1.18). For the linear case and non-varying state parameters the Prony series are:

$$C_{ij}(t-\xi) \approx \sum_{n=1}^{N} C_{ij}^{n} \cdot e^{-(t-\xi)/\tau_n}. \tag{1.66}$$

When using the (linear) visco-elastic formulation as given in (1.51)–(1.55), the material behavior is fully described through the shear- and bulk-relaxation modulus functions, $G(t-\xi)$ and $K(t-\xi)$. Prony series approximations for these functions are:

$$G(t-\xi) \approx \sum_{n=1}^{N} G^{n} \cdot e^{-(t-\xi)/\tau_n}, \tag{1.67}$$

$$K(t-\xi) \approx \sum_{n=1}^{N} K^{n} \cdot e^{-(t-\xi)/\tau_n}. \tag{1.68}$$

Generally, the relaxation times τ_n are appropriately chosen, while the participation factors (or "relaxation strength") are experimentally established, as will be explained subsequently.

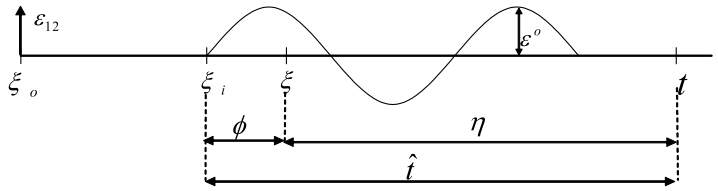


FIGURE 1.18. Sinusoidal shear deformation and definition of various time-variables.

Here the DMA procedure will be fully worked out for simple shear DMA testing, to obtain (part of) the relaxation shear modulus curve. An analogous procedure was worked out for torsional DMA testing in [4].

The applied sinusoidal deformation, with radial frequency ω, started at time ξ_i, (as illustrated in Figure 1.18) is described by:

$$\varepsilon_{12}(\xi,\omega) = \varepsilon_{12}^0 \cdot \sin\{\omega \cdot (\xi - \xi_i)\} \quad \text{for } \xi_i \le \xi \le t. \tag{1.69}$$

Through substitution of the sinusoidal strain (1.69) into (4.9) the corresponding shear stress can be written as:

$$\sigma_{12}(\hat{t},\omega) = 2\varepsilon_{12}^0 \cdot G''(\hat{t},\omega) \cdot \cos(\omega \cdot \hat{t}) + 2\varepsilon_{12}^0 \cdot G'(\hat{t},\omega) \cdot \sin(\omega \cdot \hat{t}), \tag{1.70}$$

where:

$$G'(\hat{t},\omega) = \omega \int_{\eta=0}^{\hat{t}} [G(\eta) \cdot \sin(\omega \cdot \eta)] d\eta, \tag{1.71}$$

$$G''(\hat{t},\omega) = \omega \int_{\eta=0}^{\hat{t}} [G(\eta) \cdot \cos(\omega \cdot \eta)] d\eta. \tag{1.72}$$

Here $\hat{t} = t - \xi_i$ is the "total DMA-time." G' and G'' represent the so-called "storage" and "loss" shear moduli. An equivalent state description is presented below. Here τ_0 and δ are the shear amplitude and phase shift, respectively.

$$\sigma_{12}(\hat{t},\omega) = \tau_0(\hat{t},\omega) \cdot \sin\big[\omega \cdot \hat{t} + \delta(\hat{t},\omega)\big], \tag{1.73}$$

where:

$$\tau_0(\hat{t},\omega) = 2\varepsilon_{12}^0 \cdot \frac{G''(\hat{t},\omega)}{\sin\delta(\hat{t},\omega)} = 2\varepsilon_{12}^0 \cdot \frac{G'(\hat{t},\omega)}{\cos\delta(\hat{t},\omega)}, \tag{1.74}$$

$$\tan\delta(\hat{t},\omega) = \frac{G''(\hat{t},\omega)}{G'(\hat{t},\omega)}. \tag{1.75}$$

The Prony series approximation (1.67) is now substituted into Equation (1.71) for the storage shear modulus. The result can be written as follows:

$$G'(\hat{t},\omega) = \omega \sum_{n=1}^{N} T_n(\hat{t},\omega), \tag{1.76}$$

where T_n represents the following integral expression:

$$T_n(\hat{t},\omega) = \int_{\eta=0}^{\hat{t}} [G^n \cdot e^{-\eta/\tau_n} \cdot \sin(\omega \cdot \eta)] d\eta. \tag{1.77}$$

For steady material properties (T = constant) the stiffness coefficients (= participation factors) G^n can be taken out of the integral such that (1.77) is simplified to:

$$T_n(\hat{t},\omega) \approx G^n(T) \cdot I_n(\hat{t},\omega), \tag{1.78}$$

where I_n represents following integral expression:

$$I_n(\hat{t},\omega) = \int_{\eta=0}^{\hat{t}} [e^{-\eta/\tau_n} \cdot \sin(\omega \cdot \eta)] d\eta. \tag{1.79}$$

This integral can be evaluated into:

$$I_n(\hat{t},\omega) = \frac{\omega \cdot (\tau_n)^2}{\omega^2 \cdot (\tau_n)^2 + 1} \left[-\frac{1}{\omega \cdot \tau_n} \cdot e^{-\hat{t}/\tau_n} \cdot \sin(\omega \cdot \hat{t}) - e^{-\hat{t}/\tau_n} \cdot \cos(\omega \cdot \hat{t}) + 1 \right]. \tag{1.80}$$

According to (1.76) to (1.80) the cure dependent storage shear modulus can be presented by:

$$G'(\hat{t},\omega) = \omega \sum_{n=1}^{N} T_n(\hat{t},\omega) \approx \sum_{n=1}^{N} \left\{ G^n \frac{\omega^2 \cdot (\tau_n)^2}{\omega^2 \cdot (\tau_n)^2 + 1} \cdot [\mathfrak{T}(\hat{t},\omega) + 1] \right\} \tag{1.81}$$

with transient function $\mathfrak{T}(\hat{t},\omega)$:

$$\mathfrak{T}(\hat{t},\omega) = -\frac{1}{\omega \cdot \tau_n} \cdot e^{-\hat{t}/\tau_n} \cdot \sin(\omega \cdot \hat{t}) - e^{-\hat{t}/\tau_n} \cdot \cos(\omega \cdot \hat{t}). \tag{1.82}$$

Similar calculations for the loss shear modulus yield:

$$G''(\hat{t},\omega) = \omega \sum_{n=1}^{N} \overline{T}_n(\hat{t},\omega) \approx \sum_{n=1}^{N} \left\{ G^n \cdot \frac{\omega \cdot \tau_n}{\omega^2 \cdot (\tau_n)^2 + 1} \cdot [\hat{\mathfrak{T}}(\hat{t},\omega) + 1] \right\} \tag{1.83}$$

with transient function $\hat{\mathfrak{T}}(\hat{t},\omega)$:

$$\hat{\mathfrak{T}}(\hat{t},\omega) = \omega \cdot \tau_n \cdot e^{-\hat{t}/\tau_n} \cdot \sin(\omega \cdot \hat{t}) - e^{-\hat{t}/\tau_n} \cdot \cos(\omega \cdot \hat{t}). \tag{1.84}$$

With these (transient) expressions for the storage and loss shear moduli substituted into expression (1.70) the (transient) shear stress, for a relative short vibration is obtained:

$$\sigma_{12}(t) = 2\varepsilon_{12} \cdot \sum_{n=1}^{N} \left\{ G^n \cdot \frac{\omega \cdot \tau_n}{\omega^2 \cdot (\tau_n)^2 + 1} \left[\hat{\mathfrak{T}}(\hat{t}, \omega) + 1 \right] \right\} \cdot \cos(\omega \cdot \hat{t}\,)$$

$$+ 2\varepsilon_{12} \cdot \sum_{n=1}^{N} \left\{ G^n \cdot \frac{\omega^2 \cdot (\tau_n)^2}{\omega^2 \cdot (\tau_n)^2 + 1} \left[\hat{\mathfrak{T}}(\hat{t}, \omega) + 1 \right] \right\} \cdot \sin(\omega \cdot \hat{t}\,). \tag{1.85}$$

For larger DMA times asymptotic values are reached for the transient functions $\mathfrak{T}(\hat{t}, \omega)$ and $\hat{\mathfrak{T}}(\hat{t}, \omega)$:

$$\lim_{\hat{t} \gg \tau_n} [\mathfrak{T}(\hat{t}, \omega)] = \lim_{\hat{t} \gg \tau_n} [\hat{\mathfrak{T}}(\hat{t}, \omega)] = 0. \tag{1.86}$$

Steady state values are thus obtained for the storage and loss shear moduli:

$$G'(\omega) \approx \sum_{n=1}^{N} \left[G^n \cdot \frac{\omega^2 \cdot (\tau_n)^2}{\omega^2 \cdot (\tau_n)^2 + 1} \right], \tag{1.87}$$

$$G''(\omega) \approx \sum_{n=1}^{N} \left[G^n \cdot \frac{\omega \cdot \tau_n}{\omega^2 \cdot (\tau_n)^2 + 1} \right]. \tag{1.88}$$

The corresponding steady state shear stress is thus given by:

$$\sigma_{12}(\hat{t}\,) = 2\varepsilon_{12} \cdot \sum_{n=1}^{N} \left[G^n \cdot \frac{\omega \cdot \tau_n}{\omega^2 \cdot (\tau_n)^2 + 1} \right] \cdot \cos(\omega \cdot \hat{t}\,) + 2\varepsilon_{12} \cdot \sum_{n=1}^{N} \left[G^n \cdot \frac{\omega^2 \cdot (\tau_n)^2}{\omega^2 \cdot (\tau_n)^2 + 1} \right] \cdot \sin(\omega \cdot \hat{t}\,). \tag{1.89a}$$

Or simply:

$$\sigma_{12}(\hat{t}\,) = 2\varepsilon_{12} \cdot \left[G''(\omega) \cdot \cos(\omega \cdot \hat{t}\,) + G'(\omega) \cdot \sin(\omega \cdot \hat{t}\,) \right]. \tag{1.89b}$$

1.4.5.2. Frequency Scanning at Constant Temperature In order to establish the relaxation shear modulus $G(t - \xi)$ at a certain test temperature T through DMA, a Prony series approximation according to (1.67) is adapted:

$$G(t - \xi) \approx \sum_{n=1}^{N} G^n \cdot e^{-(t-\xi)/\tau_n}. \tag{1.90}$$

On adequate experimental data, the relaxation strengths (the participation factors, G^n) and the relaxation times (τ_n) could be fitted. However, it is less complicated just to choose a number of relaxation times and to direct the fitting process on the participation factors G^n only. Of course such a method can only be successful if the "spacing" between the chosen relaxation times is not too large. Generally, the relaxation times are chosen equally spaced on the logarithmic time axes, with 2 sample points per decade.

It should be noted, that according (1.90) all terms decay to zero. To attain a lower limit, being the rubbery modulus G^R, a constant term ($= G^R$) should be added. Alternatively, a Prony term with very large relaxation time, $\tau^N \to \infty$ should be used to simulate

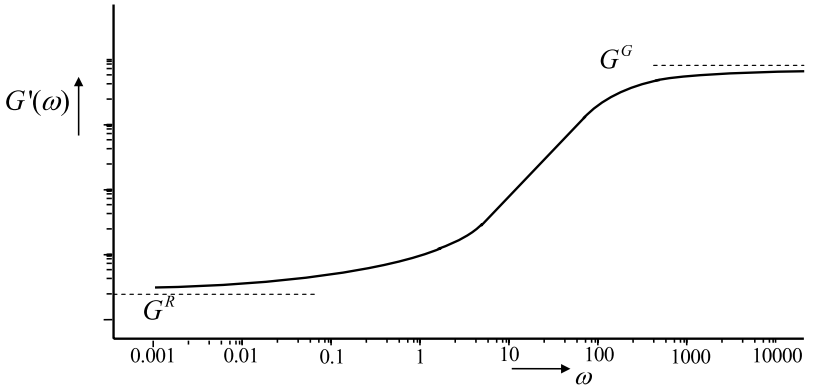


FIGURE 1.19. Typical storage shear modulus curve on a log–log scale.

the constant term. The smallest relaxation time chosen should be such that the glassy state is also involved.

The relaxation strengths also occur as participation factors of the series Equation (1.87) for the stabilized storage shear modulus $G'(\omega)$:

$$G'(\omega) \approx \sum_{n=1}^{N} \left[G^n \cdot \frac{\omega^2 \cdot (\tau_n)^2}{\omega^2 \cdot (\tau_n)^2 + 1} \right]. \tag{1.91}$$

Therefore, in principle they could be obtained through fitting the measured storage shear modulus curve to this series expression. Figure 1.19 shows an illustration of a typical storage shear modulus curve on a log–log scale.

Generally we observe a horizontal plateau for small test frequencies. This can be identified as the "rubbery plateau." For small frequencies the storage shear modulus is getting frequency independent and thus "elastic:"

$$\lim_{\omega \to 0} \left[G'(\omega) \right] = \lim_{\omega \to 0} \left[\sum_{n=1}^{N} G^n \cdot \frac{\omega^2 \cdot (\tau_n)^2}{\omega^2 \cdot (\tau_n)^2 + 1} \right] = G^N \equiv G^R \quad \text{for } \tau^N \to \infty. \tag{1.92}$$

Also for large test frequencies, a horizontal plateau is reached, being the "glassy plateau." For large frequencies the storage shear modulus is getting frequency independent and thus "elastic:"

$$\lim_{\omega \to 0} \left[G'(\omega) \right] = \lim_{\omega \to 0} \left[\sum_{n=1}^{N} G^n \cdot \frac{\omega^2 \cdot (\tau_n)^2}{\omega^2 \cdot (\tau_n)^2 + 1} \right] = \sum_{n=1}^{N} G^n \equiv G^G. \tag{1.93}$$

It should be noted, that in practical applications of dynamic mechanical analysis only a limited part of the storage shear modulus curve is obtained, because the frequency window is limited, such as illustrated in Figure 1.20. Generally, there is an upper frequency limit because of limitations of the DMA test facility (for many DMA facilities the upper limit is between 50–100 Hz). A lower frequency limit is due to practical reasons. For DMA frequency sweeps, including very low frequencies, the total duration of DMA testing would

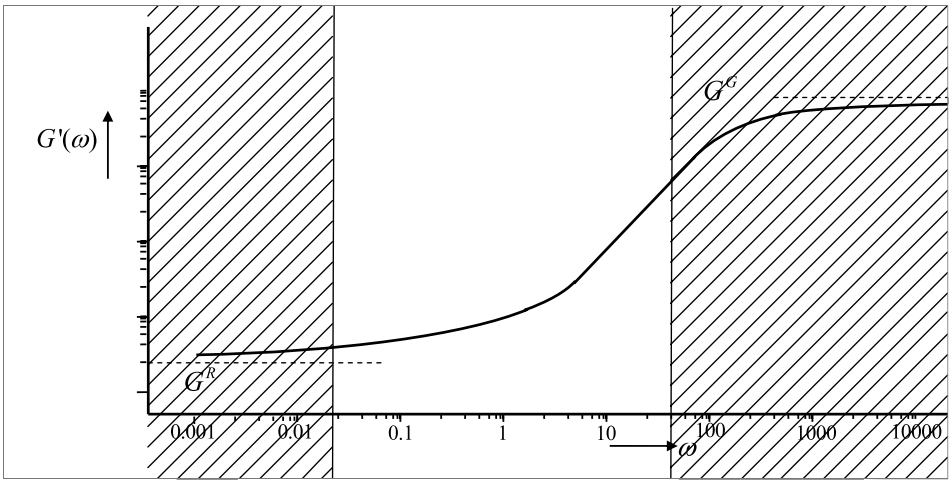


FIGURE 1.20. Limited storage modulus data because of small frequency window.

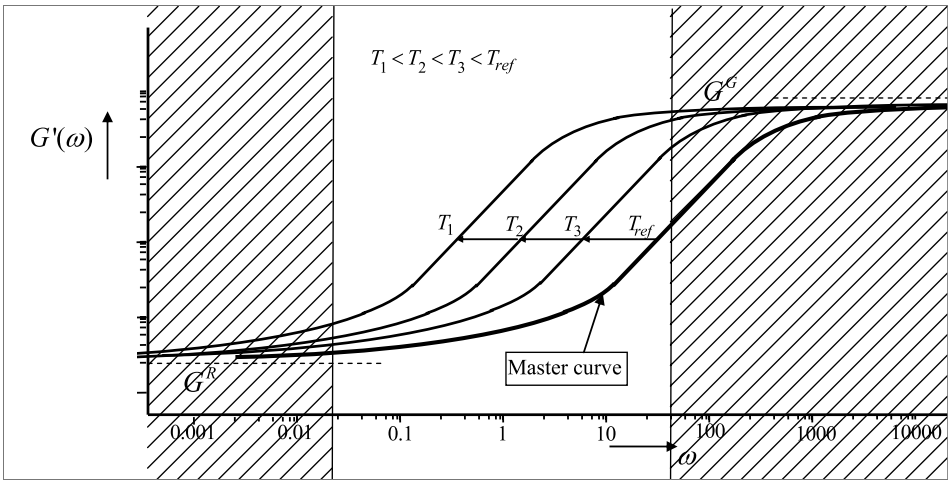


FIGURE 1.21. Schematics of storage modulus curves at different temperatures: horizontal shift on (log) frequency (ω) axis.

be quite long. For materials that behave "rheologically simple" (see also Section 1.3.4) so-called "frequency–temperature shifting" can be applied to virtually enlarge the frequency window. This will be explained in Section 1.4.5.3.

1.4.5.3. Frequency–Temperature Superposition Figure 1.21 illustrates the characteristic dynamic behavior through showing the storage shear relaxation modulus for various temperatures, within a DMA frequency window. For most materials these curves at different temperatures are all of the same shape and the only effect of temperature turns out to be a shifting along the (log) frequency axis. This is related to the earlier discussed "time–temperature" superposition principle (see Section 1.3.4). For these materials a so-called storage modulus "master curve" can be constructed by shifting the individual temperature

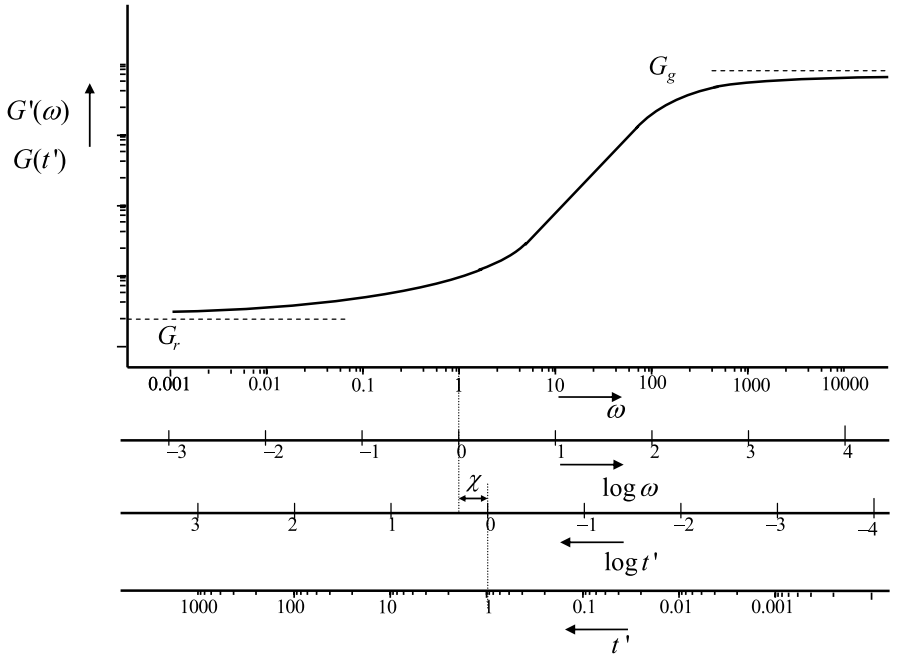


FIGURE 1.22. Typical storage shear modulus curve and relaxation shear modulus curve on a log–log scale.

curves to the curve at a chosen reference temperature. The new frequency coordinate is denoted as the reduced or effective frequency and is given as:

$$\omega_{red} = \frac{\omega}{a_T(T - T_{ref})}. \tag{1.94}$$

Here $a_T = a_T(T - T_{ref})$ is the temperature shift factor. On the basis of (1.94) it can be shown, that this shift factor is equivalent to the shift factor as introduced in Section 1.4.3.

1.4.5.4. Storage Modulus Versus Relaxation Modulus The relaxation strengths G^n can be obtained through fitting Equation (1.91) to the established storage modulus curve, $G'(\omega)$. Subsequently they can be substituted into the Prony series Equation (1.67) to obtain the shear relaxation modulus curve $G(t')$. Beside this fitting procedure, an approximate method to establish the relaxation modulus curve from the storage modulus curve is often applied. This is based on the observation that the relaxation modulus curve has approximately the same shape as the "mirrored" storage modulus curve, when presented on appropriate frequency and/or time scales, such as illustrated in Figure 1.22.

For "mirroring" the following relation between the relaxation time t' and the frequency ω holds:

$$\omega \cdot t' = \chi, \tag{1.95a}$$

or equivalently:

$$\log(\omega) + \log(t') = \log \chi. \tag{1.95b}$$

Here χ is a number that defines the offset between the $\log \omega$ and the $\log t'$ scales as shown in Figure 1.19. In [11] it is shown that the best correspondence between the "mirrored" shear relaxation modulus curve and the storage modulus curve is obtained when:

$$\chi = 0.56. \qquad [11] \tag{1.96}$$

Other published values are:

$$\chi = 1, \qquad [14] \tag{1.97}$$

$$\chi = 2/\pi. \qquad [19] \tag{1.98}$$

1.5. MODELING OF CURING POLYMERS

In electronic packaging, filled thermosetting polymers such as underfills and molding compounds are frequently used for encapsulation. The combined chemical shrinkage and thermal expansion occurring during cure and subsequent cooling down often result into undesirable product stresses and warpage. In order to be able to reduce or prevent these stresses and/or warpage the compound selection and the curing procedure should be optimized within the design process. For this the package behavior should be understood in advance, which in turn requires adequate modeling and characterization of compound behavior during cure.

In this section we will focus on the thermo-mechanical modeling of packaging polymers during cure. Generally, packaging polymers are epoxy resins filled with inorganic (silica) particles, carbon black and processing aids. They show a clear viscoelastic behavior which is not only temperature but also cure dependent. The latter effect is quite large: the position of the viscoelastic transition region shifts for more than 8 decades during cure of the molding compound.

In previous modeling attempts this complex behavior was simplified by assuming the material to be stress free (i.e., infinitely fast and complete stress relaxation) until the end of the molding stage and either elastic or viscoelastic during the subsequent cooling and post curing stages. Since these simplified approaches are unable to accurately predict complex phenomena like warpage of thin packages, a more complete description of the thermo-mechanical behavior of the molding compound during and after cure is required. The present chapter discusses the state of the art of cure and temperature dependent thermal mechanical modeling and characterization of packaging polymers.

First in Section 1.5.1 so-called "partly state dependent" modeling of curing polymers is discussed. Here actually the model implies all changes of state parameters during cure, but the evolution of stress actually is fully based on the state parameters at the moment of load application. The principle of cure-temperature–time superposition is adopted for this type of modeling. The "partly state dependent" model is an approximation of reality that can result into inaccurate results for very fast curing polymers. The various experimental characterizations necessary for this model are discussed in Section 1.5.1.1. For a model-molding compound typical data are presented in Section 1.5.1.2.

In Section 1.5.2 we discuss "fully state dependent" modeling, which requires that the "history of state parameters" is adequately taken into account in the convolution integral expression for the state of stress. In Section 1.5.2.1 an "approximate fully cure dependent" model, which can be seen as an extension of the model presented in Section 1.5.1 is dis-

cussed in Section 1.5.2.1. This is an attractive model that simply can be implemented in standard FEM packages. A more complicated "fully state dependent" model being recently developed by [11] is described in Section 1.5.5.2. Although this model is quite interesting, as it does not pre-assume the validity of cure-temperature time superposition, it is hard to be implemented into a FEM program. Therefore it is used as a reference to compare the model results.

Finally, the chapter is concluded with future trends in cure dependent modeling in Section 1.5.2.3.

1.5.1. "Partly State Dependent" Modeling (Curing Polymers)

In previous work [2] a preliminary cure dependent viscoelastic model was presented and applied to underfill resin. Here the cure dependency was involved in the relaxation strengths (= participation factors of the Prony series description of the relaxation modulus functions). The model could only be applied to constant temperature curing. Some basics of an improved model describing cure and temperature dependent viscoelasticity has recently been published in [22]. Here the theory and characterizations will more extensively be discussed.

The state dependent visco-elastic modeling of packaging polymers starts out from the general relations (1.27) as presented in Section 1.3.3:

$$\sigma_{ij}(t) = \int_{\xi=-\infty}^{t} C_{ijkl}\left[x^{l}(\xi), (t-\xi)\right] \cdot \left[\left(\varepsilon_{kl,\xi}\right)_{\xi} - \left(\varepsilon^{*}_{kl,\xi}\right)_{\xi}\right] d\xi. \tag{1.99}$$

We will assume that the only important state parameters during cure are:

- The degree of cure α: $x^{1}(\xi) = \alpha(\xi)$. (1.100)
- The temperature T: $x^{2}(\xi) = T(\xi)$. (1.101)

(Moisture influences during cure are assumed to be negligible.)

If we assume isotropic material behavior (as discussed in Section 1.3.4) Equation (1.99) can be simplified to:

$$\sigma_{ij}(t) = \int_{\xi=-\infty}^{t} \left\{2G[\alpha(\xi), T(\xi), (t-\xi)] \cdot \left(\varepsilon^{d}_{ij,\xi}\right)_{\xi} + K[\alpha(\xi), T(\xi), (t-\xi)] \cdot \left(\varepsilon^{\mathit{eff}}_{v,\xi}\right)_{\xi}\right\} d\xi, \tag{1.102}$$

where: G = relaxation shear modulus function, K = relaxation bulk modulus function, σ^{d}_{ij} = deviatoric stress [see (1.34a)], ε^{d}_{ij} = deviatoric strain [see (1.34.b)], $\varepsilon^{\mathit{eff}}_{v}$ = effective volumetric strain [see (1.32a)], σ_{v} = volumetric stress [see (1.32b)].

In order to be able to construct adequate models for the description of the relaxation shear modulus function and the relaxation bulk modulus function for curing polymers, we realize that during cure, due to the molecular network formation, the material state changes from liquid into solid, as illustrated in Figure 1.23. At the same time the relaxation modulus functions (G and K) change their behavior. Typical changes of the shear-relaxation modulus function G that generally are observed during cure are illustrated in Figure 1.24.

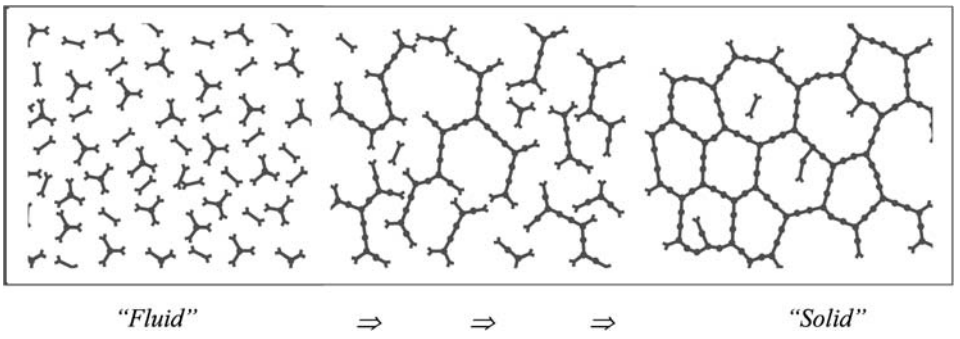


FIGURE 1.23. Illustration of molecular changes during cure.

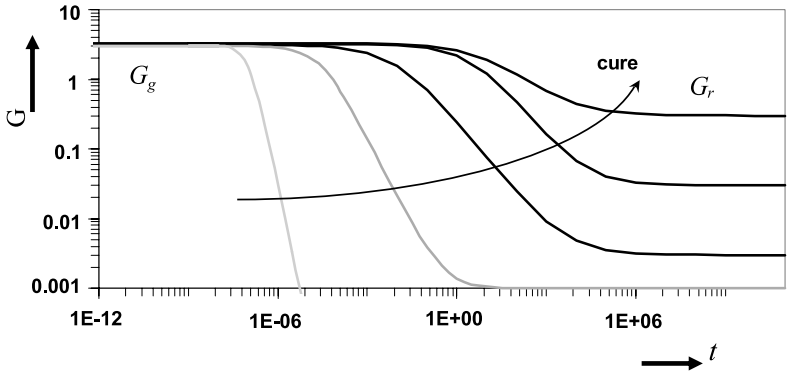


FIGURE 1.24. Illustration of changing shear modulus curves during cure.

The typical changes of the relaxation bulk modulus are analogously, but less pronounced. In the sequence the discussion will be primarily focused on the description of the relaxation shear modulus. An analogous discussion can be given for the relaxation bulk modulus.

Generally we observe that during cure from the liquid state until the fully cured state behavior is reached:

- the "glassy modulus" G_g hardly changes,
- the transition to the rubbery part is delayed,
- the slope of the transient part is reduced,
- and the rubbery modulus G_r increases.

In view of the observed phenomena, the dependency of the modulus functions from the degree of cure is first split off in a transient part and a (more or less) steady part, being the rubbery modulus:

$$G[\alpha(\xi), T(\xi), (t-\xi)] = G_r[\alpha(\xi), T(\xi)] + G_{transient}[\alpha(\xi), T(\xi)]. \tag{1.103}$$

Further, a "time–cure-temperature" superposition principle is adopted (and successfully tested, see Section 1.5.1.2) to describe the dependency of the transient part, from the

degree of cure and the curing temperature. The governing equations for the transient part are thus described through:

$$G_{transient} = \underbrace{\{G_g - G_r[\alpha(\xi), T(\xi)]\}}_{relaxation_amplitude} \cdot \sum_{n=1}^{N} h_n \cdot \exp\left[-\frac{t_{red}(t) - t_{red}(\xi)}{\tau_n^{ref}}\right], \tag{1.104}$$

where: ξ = load setup time, t = current time, τ_n^{ref} = relaxation times at reference state, h_n = normalized relaxation strength, with:

$$\sum h_n = 1, \tag{1.105}$$

t_{red} = "reduced time" defined through:

$$t_{red}(t') = t' \cdot a_\alpha[\alpha(t')] \cdot a_T[T(t')], \tag{1.106}$$

a_α = cure-shift factor, a_T = temperature shift factor.

For the relaxation bulk modulus K an analogous set of governing equations holds:

$$K[\alpha(\xi), T(\xi), (t-\xi)] = K_r[\alpha(\xi), T(\xi)] + K_{transient}[\alpha(\xi), T(\xi)], \tag{1.107}$$

$$K_{transient} = \underbrace{[K_g - K_r[\alpha(\xi), T(\xi)]]}_{relaxation_amplitude} \cdot \sum_{n=1}^{N} k_n \cdot \exp\left[-\frac{t_{red}(t) - t_{red}(\xi)}{\tau_n^{ref}}\right], \tag{1.108}$$

where: τ_n^{ref} = relaxation times at reference state, k_n = normalized relaxation strength, with:

$$\sum k_n = 1. \tag{1.109}$$

1.5.1.1. Experimental Characterizations The viscoelastic characterizations can be performed with various special tests in which an initially fluid sample is subjected to dynamic mechanical analysis during cure or where partly cured samples are characterized in various manners. In the sequence an overview of recently used test methods is given, together with a discussion of the necessary improvements.

In [13] and later in [2] coin like resin samples were successfully used in DMA measurements (illustrated in Figure 1.25) to establish the visco-elastic properties at isothermal cure. The mechanical properties, such as the (curing-) time dependent initial strains and the relaxation shear and bulk moduli were established by means of specially designed DMA measurements. Here the parameter identification process was actually based on one-directional measurement data. These originate from relative thickness changes and relative torsional deformations of the coin shaped specimen. Implicit assumptions on the deformation in the radial direction are involved and bring a restriction to the model accuracy. Because of frequency limitations of the test set-up, in relation to the rate of change of the model parameters, the proposed method was restricted to relatively slowly curing underfill polymers. Both torsion and axial DMA measurements were performed. The coin-shaped resin samples were kept free to shrink in thickness direction except on selected and relatively short periods of time, where the sinusoidal torsional or axial deformations were

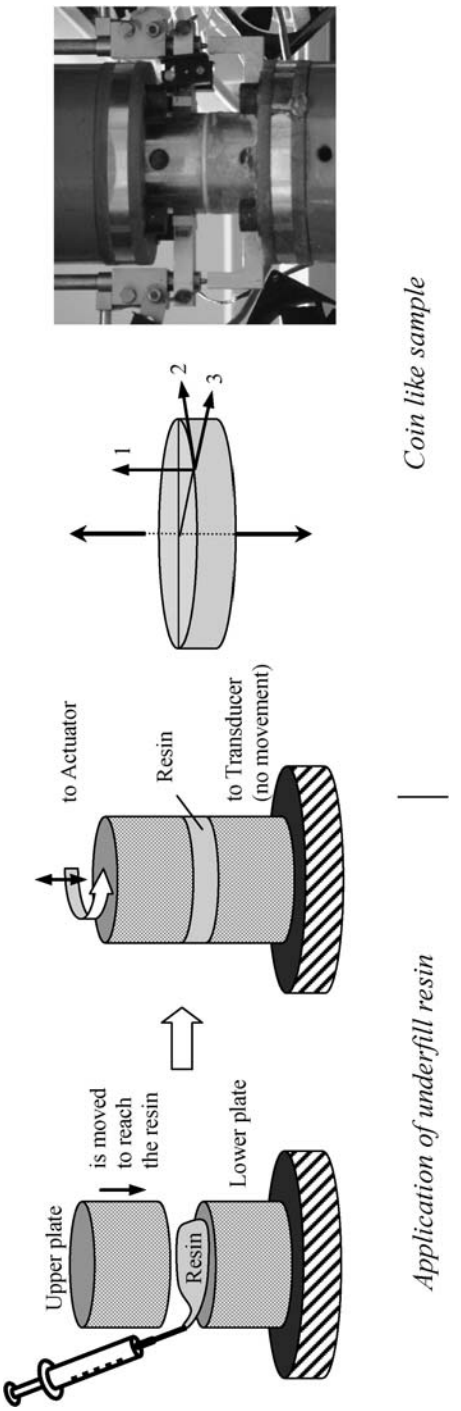


FIGURE 1.25. DMA measurements during cure start out from liquid samples.

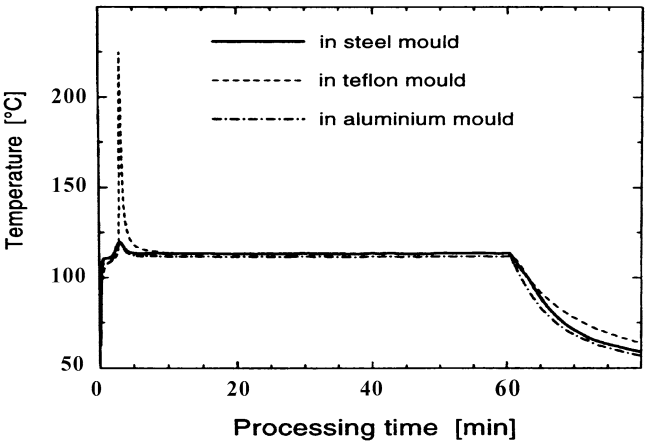


FIGURE 1.26. Temperature transient behavior at beginning of cure due to exothermic reaction.

prescribed. The axial shrinkage was used to establish the volumetric shrinkage under the assumption of plane strain. Measurements of the torque during such intermittent sinusoidal loading yielded data for investigation of the storage shear modulus $G'(\omega, \xi)$. Measurements of the axial load resulted into data to establish the storage longitudinal modulus $M'(\omega, \xi)$. Both data were used to indirectly establish the storage bulk modulus $K'(\omega, \xi)$. It appeared that the procedure to establish the bulk relaxation modulus function in this indirect manner, together with the assumption of plane strain, did not result into a sufficient accurate description. Further, for larger resin stiffness the data appeared to be affected by the insufficient stiffness of the test facilities. Therefore, in [5] the model was improved using additional data from various step relaxation experiments on coin-like samples.

In [2] it was also discussed, that at the beginning of cure the sample temperature can show a peak due to the exothermic reaction (see Figure 1.26). It was found that this peak could quite well be reduced through selection of aluminum as the plate material on both sides of the coin like sample. Further it was observed that at the moment that the peak arises the material still is in the fluidic state (the shear relaxation modulus has a negligible value). This means that for stress simulations the influence could be omitted.

In more recent research (i.e., [22,23]) the torsional DMA test to establish the relaxation shear modulus during cure was replaced with a double simple shear tool (DST) specimen, with optimized dimensions concerning the ratio between sample stiffness and machine stiffness, while the gap between sample holder plates is small enough to ensure that in the fluidic state the resin does not simply leak out. A first test setup is presented in Figure 1.27. Here the samples are free to shrink in thickness direction because of the lateral flexibility of four applied vertical leaf springs.

From various stiffness measurements and dynamic characterization [9,10] of the test facility again it was observed that for high sample stiffness the machine + tool stiffness was insufficient to provide reliable test results. There fore stiffer shear tools were designed, where the lateral shrinkage compensation was abandoned (see Figure 1.28).

It should be noted that with the sample holder as given in Figure 1.28, only the stiffness of the sample holder is improved, while the machine stiffness still could be insufficient. This occurs in some exceptional cases with very high sample stiffness (near the glassy plateau at low temperatures). For these situations in [9,10] a "sandwich beam shear tool" (SBT) was proposed. This newly designed "sandwich beam" specimen (see

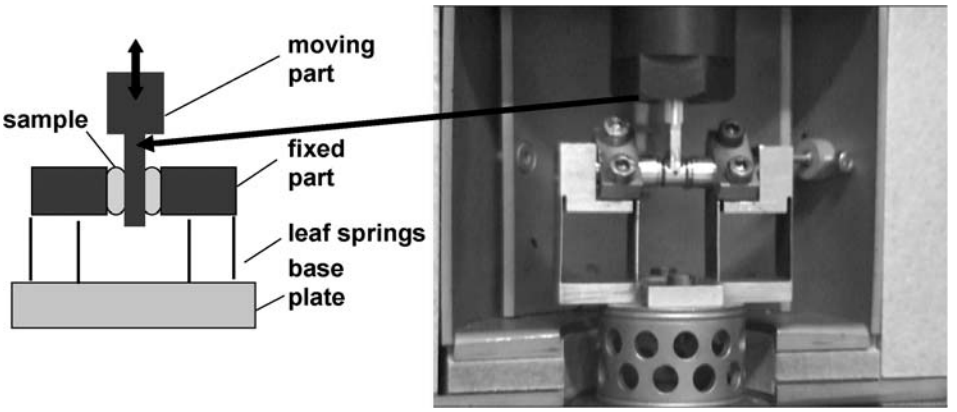


FIGURE 1.27. Double simple shear (SST) specimen with shrinkage compensation.

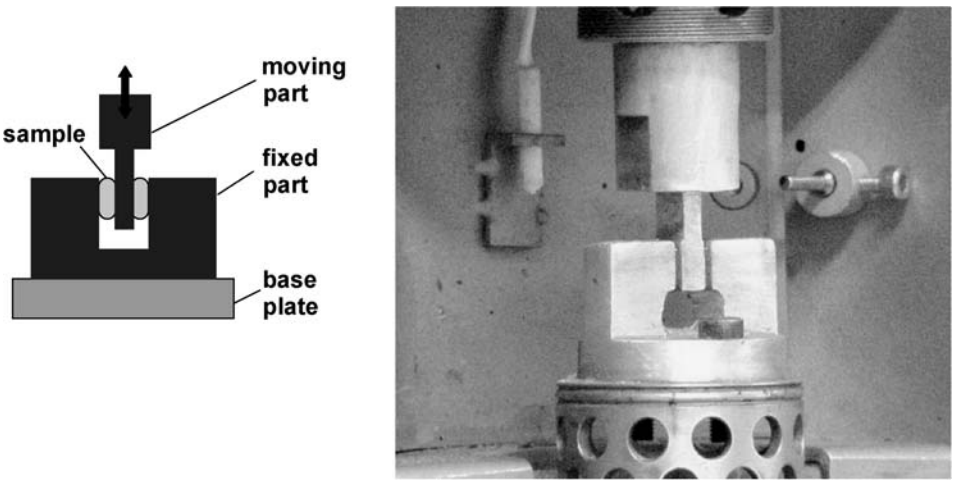


FIGURE 1.28. Double simple shear (DSS) specimen without shrinkage compensation.

Figure 1.29) is appropriate for viscoelastic characterization of the relaxation shear modulus in case of higher stiffness. Figure 1.30 shows storage shear modulus data for 5 frequencies, from a simple shear tool (with shrinkage compensation, see Figure 1.27) and from a sandwich beam tool (see Figure 1.29). It is advised to use combined results: simple shear data for the low modulus range and sandwich beam data for the high modulus range.

With the sandwich beam shear tool also the glassy properties of partly cured, even ungelled resin can be measured. Such specimens are very brittle, i.e., both handling and mainly clamping of other kinds of partly cured samples can be rather problematic. Now, instead of clamping, adhesion to the upper and lower leaf springs is used as fixation.

The sandwich beam shear tool used consists of a symmetric frame, in which two 4 mm wide leaf springs are clamped at both ends, thickness 0.3 mm, with 1 mm thick spacers in between. In the middle of the springs the springs are also rigidly clamped using a spacer. Two slits of $25 \times 4 \times 1$ mm, in between two parallel leaf springs, are obtained in this

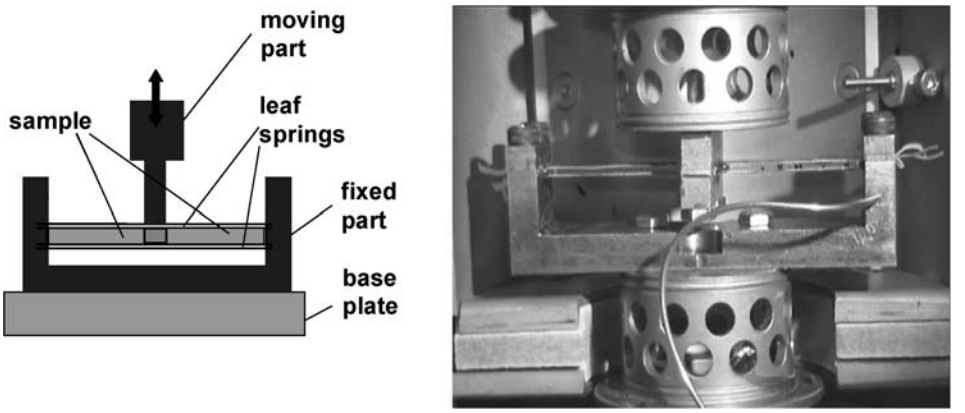


FIGURE 1.29. Sandwich beam shear tool (SBT).

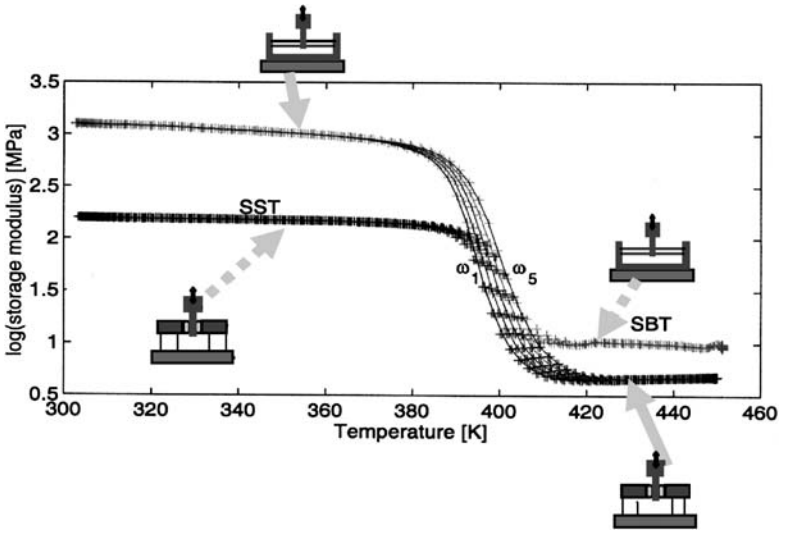


FIGURE 1.30. Combined use of SST and SBT results.

way. The frame of the tool is connected to the force transducer of the DMA analyzer, the middle clamp is connected to the combined excitator and displacement transducer. Resin is applied in liquid form between the two parallel leaf springs; such that a sandwich structure is obtained. Thereby leakage is prevented and the geometry is relatively well-defined. With the stiffness ratio of the leaf springs and the epoxy core the deformation mode of the core is almost purely shear. The tool's stiffness thus increases when the shear modulus of the resin increases. As even with fully cured resin the tool stiffness remains small compared to the machine stiffness the SBT is quite suitable.

The fact that the steel springs are fully clamped results in a complicated shear stress pattern along the sandwich length l. For data processing this is a disadvantage as the mathematical relations between experimental results and material modulus become very complex as the modulus does so. The complexity of the mathematics is reduced by the using approximate equations, based on the assumption of a constant shear stress distribution along the

sandwich length. The "approximate modulus" deviates from the "exact" modulus, in particular in the lower stiffness range (where just the results of the SST are usable). According to [9,10] the approximate solution for the storage shear modulus G' and the loss shear modulus G'' are found as:

$$G' = \frac{-EA \cdot \delta_2 - FG}{E^2 + F^2}, \tag{1.110a}$$

$$G'' = \frac{EG - FA \cdot \delta_2}{E^2 + F^2}. \tag{1.110b}$$

Here: δ_1 = "in phase part" of the measured force signal, δ_2 = "out of phase part" of the measured force signal,

$$A = \frac{12EI}{l^3}, \quad B = \frac{b(h_s + h_k)}{2}, \quad C = \frac{h_s + h_k}{l \cdot h_k}, \quad D = \frac{b \cdot l^2}{6EA \cdot h_k}, \tag{1.110c–f}$$

$$E = (BC + DA)\delta_2, \quad F = (BC + DA)\delta_1 - DP/4, \quad G = A \cdot \delta_1 - P/4. \tag{1.110g–i}$$

Straightforward measuring the relaxation bulk modulus K during cure is performed by applying partly cured samples in a pressure cell as earlier described in Sections 1.4.2.2 and 1.4.4.2 (see Figure 1.10). The material sample is subjected to hydrostatic pressure. The strain is measured through free-grid strain gages in the sample. Actually creep experiments are performed to directly establish the bulk (= volumetric) creep compliance B. The bulk relaxation modulus K is indirectly derived by solving the interconversion Equation (1.43). The method is restricted to material samples with relatively high degree of cure, because of the inherent stiffness of the free grid strain gages. Also the temperature control appears to be difficult because of the energy release through adiabatic compression during load application.

Another method to indirectly establish the relaxation bulk modulus is through longitudinal DMA testing. This can be performed on partly cured samples (see Figure 1.30). With the strip DMA test the relaxation Young's modulus $E(t - \xi)$ is first established. With known relaxation Young's modulus and known relaxation shear modulus, the relaxation bulk modulus $G(t - \xi)$ can be obtained by solving the Laplace interrelation (1.50a). Because of the handling and clamping of the partly cured samples this method is restricted to samples with a relatively high degree of cure. This problem is circumvented through longitudinal DMA testing on initially fluid "dog bone like" samples, being molded in a specially shaped mold (see Figure 1.31), when mounted in a testing machine. The shape of the mold is such designed that after initial chemical shrinkage; the sample is released from the mold.

An important parameter is the combined chemical shrinkage and thermal expansion evolution during cure. This is investigated by monitoring the density changes during cure of material samples immersed in silicon oil (see Figure 1.32). The effective volumetric strain is derived from the difference in weight of the free and the immersed sample (applying the Archimedes principle).

Independent kinetic measurements are required to supply the relation between cure time, temperature and conversion level. In the present work the kinetic data was obtained from DSC (differential scanning calorimetry) measurements. Here some basic expressions related to the cure kinetics are summarized:

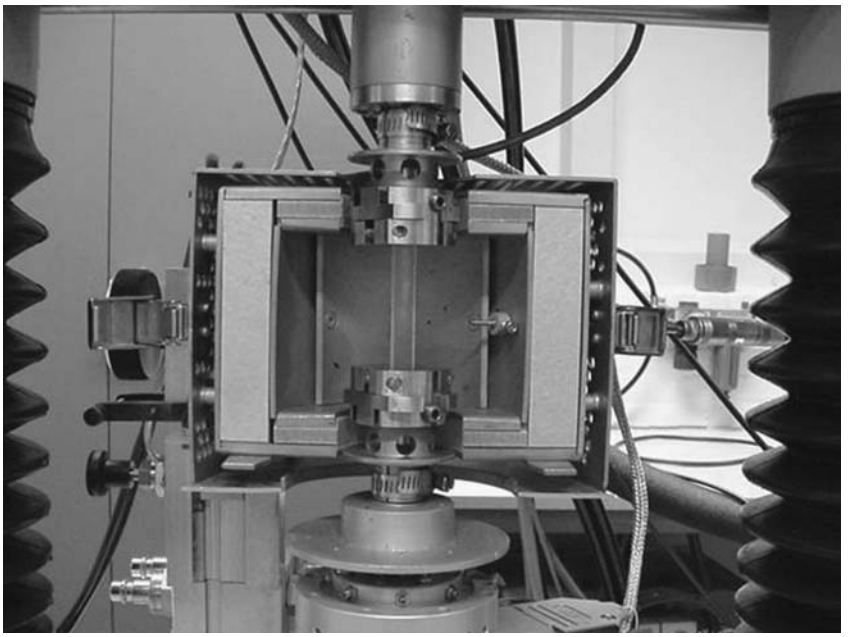


FIGURE 1.31. Longitudinal DMA testing on partly cured samples.

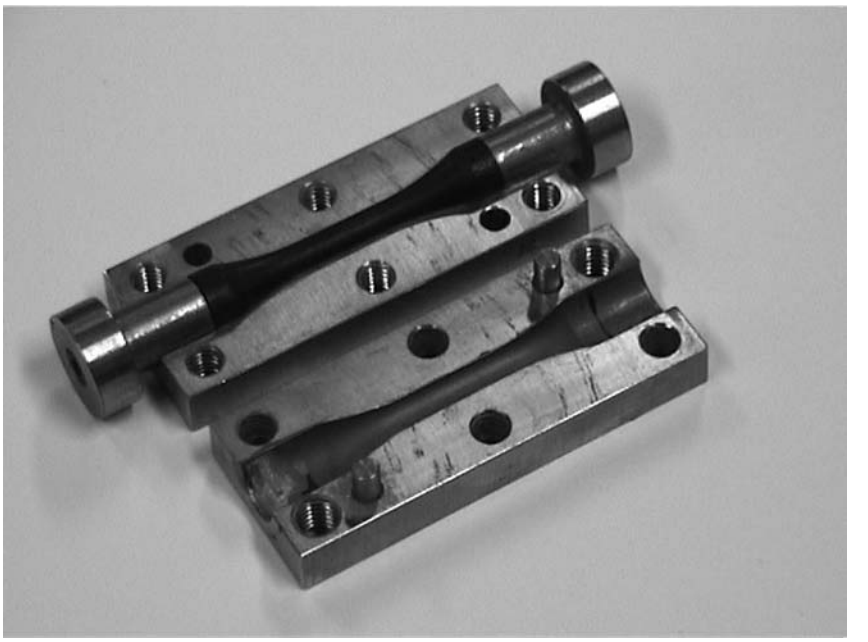


FIGURE 1.32. Initially fluid sample, molded and tested in a tensile test facility.

Degree of cure *(conversion)*:

$$\alpha = 1 - \Delta H_R / \Delta H_T, \tag{1.111}$$

with: ΔH_R = residual heat of reaction (from partly to fully cured), ΔH_T = total heat of reaction (for full cure).

Relationship between T_g ***and*** α:

$$\frac{T_g - T_{g^0}}{T_{g\infty} - T_{g^0}} = \frac{\lambda\alpha}{1 - (1-\lambda)\alpha} \quad \text{(DiBenedetto equation)}, \tag{1.112}$$

where: λ = material constant, $T_{g^0} = T_g$ of uncured resin $T_{g\infty} = T_g$ of fully cured resin.

Cure kinetics model:

$$\frac{d\alpha}{dt} = k_T \alpha^m (1-\alpha)^n \quad \text{(Kamal and Sourour Equation)}, \tag{1.113}$$

where: m and n represent the reaction orders, k_T = reaction rate:

$$k_T = k_0 \exp\left(-\frac{E}{RT}\right), \tag{1.114}$$

k_0 = a constant, E = activation energy, R = gas constant, T = absolute temperature.

1.5.1.2. Experimental Data for Model Molding Compounds In order to illustrate the experimental characterizations for curing polymers, various experimental investigations as performed for model molding compounds will be presented. These molding compounds are created on the basis of an epoxy resin filled with various filler loads. The epoxy resin used is novolac epoxy (EPN 1180) with an equivalent weight of epoxy groups equal to 175–182 g/eq. Triphenylphosphine (TPP) is used as an accelerator for the curing of the epoxy resin. The hardener is bisphenol-A with an equivalent epoxy weight of 114.1 g/eq. Fused silica spheres are used as filler. The filler has a median diameter of 15 μm and a density of 2.20 g/cm^3. Stoichiometric amounts of novolac epoxy and bisphenol-A are mixed at 150°C. After cooling down to 85°C, the accelerator is added and the mixture again is fully stirred. Afterwards the mixture is put under vacuum at 75°C for about 20 minutes, in order to degas. Afterwards, the silica particles are dispersed in the mixture and well mixed in a special mixer. After mixing, the mixture is immediately cooled down and stored in a refrigerator at −15°C for later use.

1.5.1.2.1. Results from DSC Experiments. DSC measurements were performed both in isothermal and in dynamic heating conditions. For dynamic cure, the samples are heated at six heating rates (0.5, 1, 2, 5, 10, 20°C/min). For isothermal cure, the samples are cured for a certain time, cooled quickly to −40°C, and subsequently scanned at a heating rate of 10°C/min to 300°C. Some typical DSC results are presented in subsequent graphs. Figure 1.33 shows the relationship between T_g and conversion. The T_g of the uncured sample ($\alpha = 0$) is determined from the mixed resin without adding catalyst. The other measured data is acquired from isothermal curing and subsequent heat scanning. The continuous lines are obtained by fitting the DeBenedetto Equation (1.112). The fit parameters are given in the table.

The degree of cure α versus temperature can be calculated from the dynamic cure data through fitting the Kamal and Sourour Equation (1.113). Figure 1.34(a–c) shows the results for compounds with various filler % and various curing temperatures.

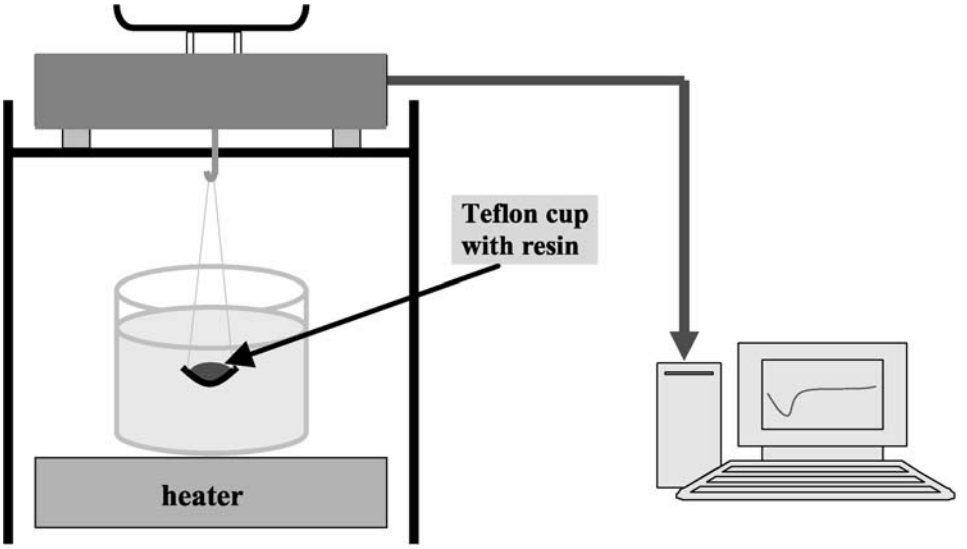
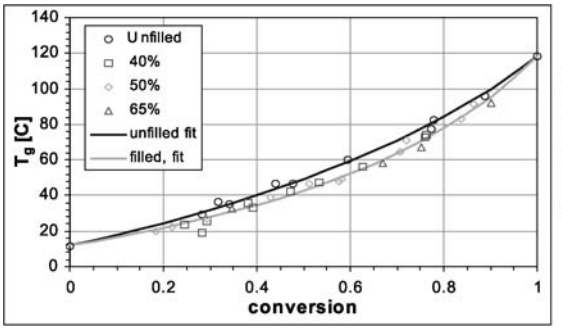


FIGURE 1.33. Archimedes principle set-up.



wt% filler	0%	40%	50%	65%
T_{g0} [C]	12	12	12	12
T_{gI} [C]	118	118	118	118
λ	0.54	0.41	0.41	0.41

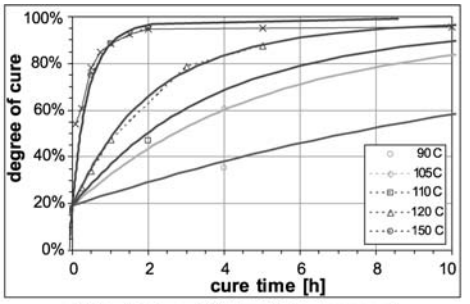
Fit parameters of the DiBenedetto equation for compounds with various filler %

FIGURE 1.34. T_g versus α for various filler %.

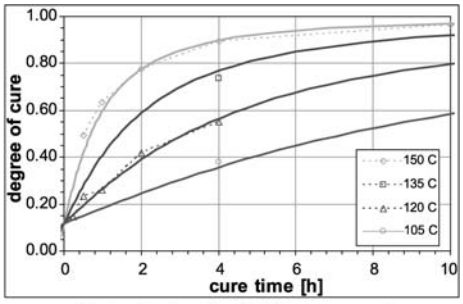
1.5.1.2.2. Results from DMA Experiments. Originally the DMA measurements were performed by applying continuous frequency sweeps at various (fixed) temperatures (as illustrated in Figure 1.35). In that way the frequency dependent modulus is scanned during cure. However, at high temperatures the continuing cure could affect the (low frequency) DMA results. Therefore, in [11] an alternative procedure was introduced, where intermittent cure is performed. In between the cure intervals the material is cooled down below T_g such that the curing is virtually stopped. DMA measurements are than performed during cooling down or heating up (as illustrated in Figure 1.35).

To illustrate the procedure, for the unfilled material, the evolution of the storage shear moduli G′ (for various frequencies) with ongoing cure is presented in Figure 1.36. The low frequency data (around 10^{-3} Hz) are obtained through so-called ultra low frequency DMA measurements and denote the increase of the rubbery modulus with ongoing cure.

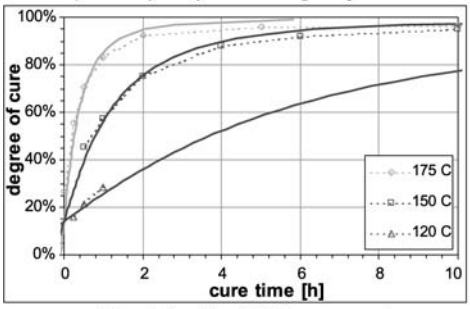
In accordance with (1.103), the "transient parts" of the storage shear moduli are obtained by subtracting the rubbery modulus. In Figure 1.37 these transient parts are shown as functions of the applied frequencies. Then the mastercurve for the fully cured material is



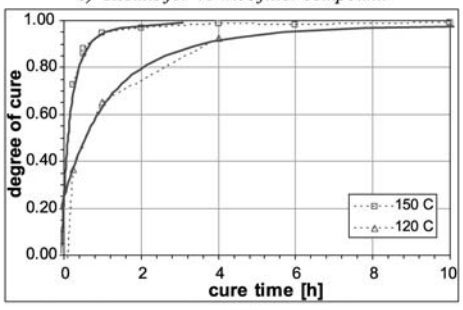
a) Results for unfilled molding compound.



b) Results for 40 wt% filler compound



c) Results for 50 wt% filler compound

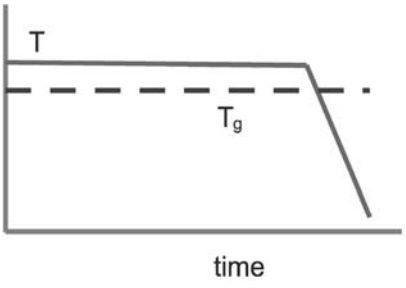


d) Results for 65 wt% filler compound

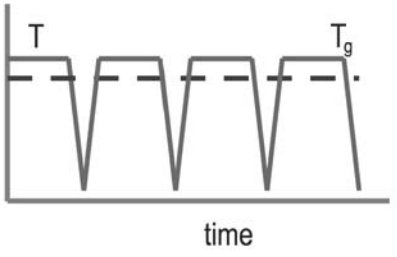
wt% filler	0%	40%	50%	65%
E [kJ/mol]	76.81	69.4	70.66	76.81
k_0 [1E5/s]	280	1.40	1.55	51.1
m [-]	0.205	0.21	0.208	0.205
n [-]	1.69	1.63	1.452	1.69
α_0	0.17	0.11	0.14	0.24

e) Parameters of the kinetic equation, Eq. 5.15

FIGURE 1.35. Conversion α versus curing time [full lines are fits to Equation (1.113)].



a) Steady Temperature during DMA
For curing temperature above "ultimate T_g"



b) DMA at intermittend cooling down
For curing temperature above "ultimate T_g"

FIGURE 1.36. Temperature profiles during DMA.

added (thick full line in Figure 1.37). This mastercurve was obtained by cooling down each sample (at 1°C/min) after the isothermal cure measurements and continuing the frequency

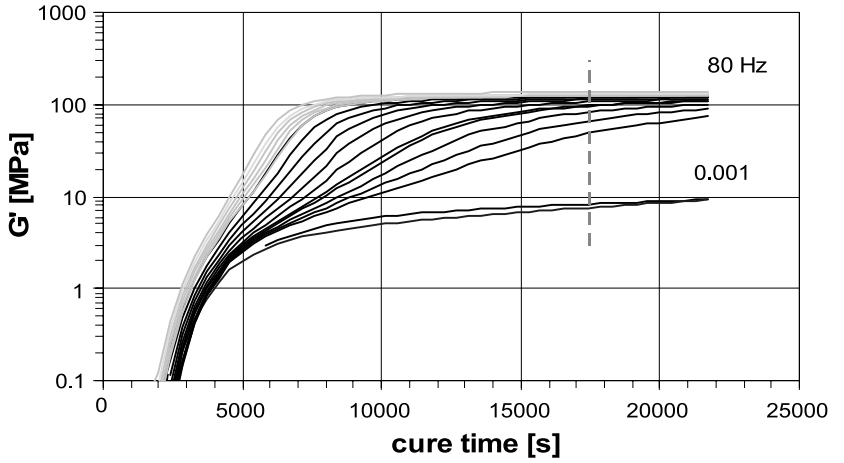


FIGURE 1.37. Storage modulus evolution (0% filled).

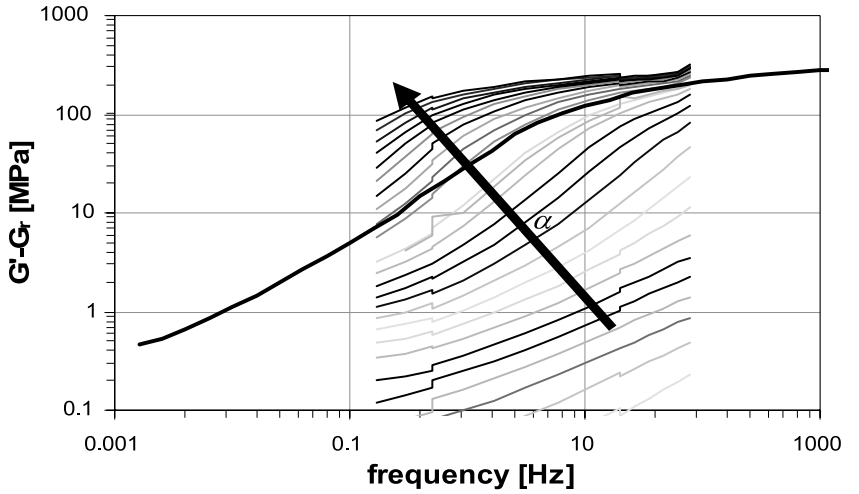


FIGURE 1.38. Transient parts of storage moduli ($G' - G_r$) versus frequency for various degree of cure and $T_{cure} = 150°$C. (Thick full line: "fully cured" mastercurve ($T_{ref} = 120°$C).)

sweeps. The resulting data is then interpolated to equidistant temperature levels and shifted using the regular time–temperature superposition technique.

The "cure shift factor" a_α is obtained after shifting the individual curves to match with the "fully cured" mastercurve. The procedure is repeated for various temperatures and for compounds with various filler %. The final mastercurves for the relaxation shear modulus G for various compounds (with different filler %) are presented in Figure 1.38. The conversion shift factors are presented in Figure 1.39(a). The (inverse) temperature shift factor for a 65% filled compound is presented in Figure 1.39(b). For the 65% material finally the evolution of the rubber modulus during cure is presented in Figure 1.40.

1.5.1.2.3. Results from Pressure Cell Experiments. Mastercurves for the bulk relaxation modulus as obtained from pressure cell experiments were previously presented in Sec-

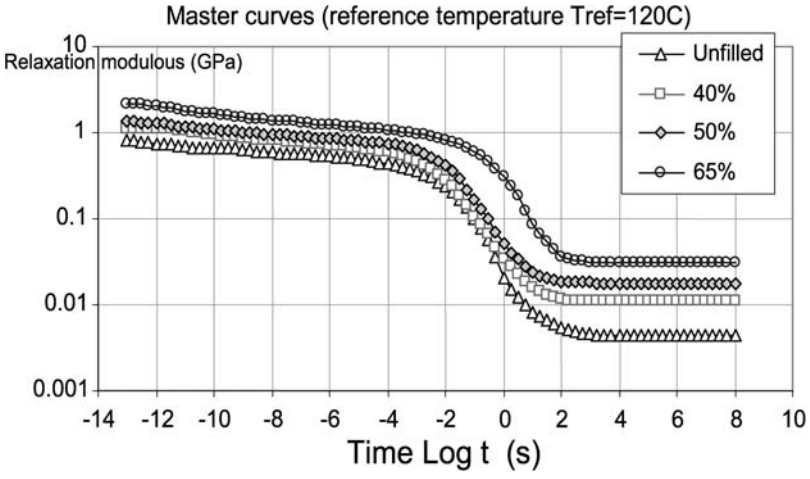
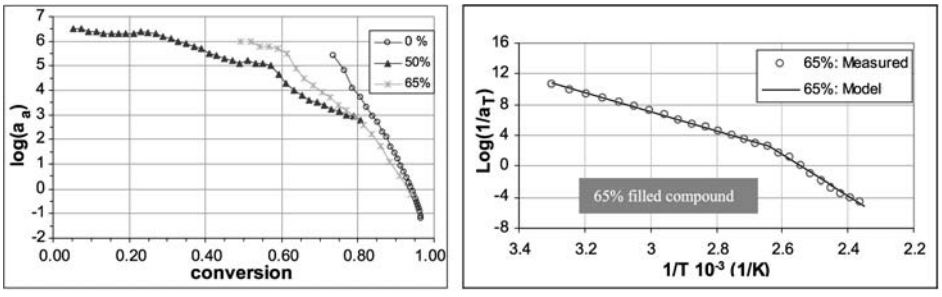


FIGURE 1.39. Master curves of relaxation shear modulus G for various filler %.



a) Conversion shift factors for various filler %.

b) (Inverse) temperature shift factor versus (inverse) temperature (65% filled compound)

FIGURE 1.40. Conversion- and temperature-shift factors.

tion 1.4.4.2, Figure 1.16. for compounds with 40% filler. In principle, the cure- and temperature-shift factors [Figures 1.39(a, b)] also apply to the relaxation bulk moduli.

1.5.1.2.4. Results from Density Measurements. Figure 1.41 shows the density evolution of a 65% filled compound for isothermal cure at 120°C. In Figure 1.42 the result is presented against the degree of cure. Also the result for isothermal cure at 140°C is presented. It can be seen that the density increases linearly with the degree of cure. Therefore, following linear cure and temperature dependent density model can be applied:

$$\rho(T,\alpha) = \rho_{ref}[1 - 3\beta(T - T_{ref}) + 3\gamma(\alpha - \alpha_{ref})], \tag{1.115a}$$

with: T_{ref} = reference temperature, α_{ref} = reference conversion, ρ_{ref} = density at T_{ref} and α_{ref}, β = linear coefficient of thermal expansion (CTE), γ = linear cure shrinkage parameter.

The effective volumetric shrinkage is thus obtained as:

$$\varepsilon_v^{eff} = \frac{\Delta V}{V} = \frac{\rho}{\Delta\rho} \simeq 3 \cdot \beta \cdot (T - T_{ref}) - 3 \cdot \gamma \cdot (\alpha - \alpha_{ref}). \tag{1.115b}$$

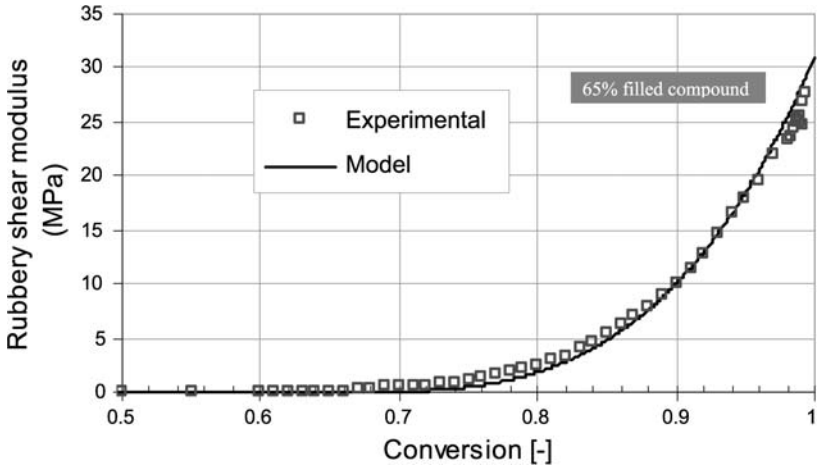


FIGURE 1.41. Evolution of rubbery shear modulus G_r during cure (65% filled compound).

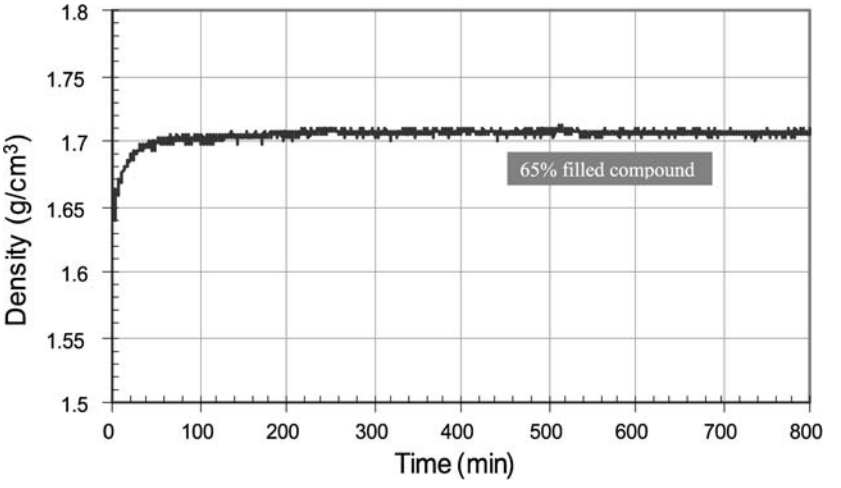
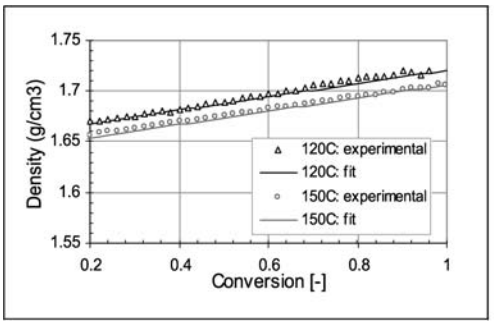


FIGURE 1.42. Density evolution during cure (65% filled compound, $T_{cure} = 120°C$).

1.5.2. "Fully State Dependent" Modeling (Curing Polymers)

Although the previously discussed model (Section 1.5.1) has successfully been applied for various thermal mechanical simulations of packages with relatively slowly curing resins (i.e., [20–23]) it should be realized that the state dependent (cure + temperature) relaxation curves actually are for materials with "fixed states," connected to the "load set-up time" ξ, as illustrated in Figure 1.43.

In reality the material is changing continuously during cure, such that the actual relaxation behavior might be deviating from the "fixed state" behavior, as is illustrated in Figure 1.44. In particular, for fast curing systems such a deviation could have significant influence to the stress/strain solution. Therefore so-called "fully cure dependent" visco-



T_{ref} [°C]	α_{ref}	ρ_{ref} [g/cm^3]	$\beta\ (T<T_g)$ [10^{-6} K^{-1}]	$\beta\ (T>T_g)$ [10^{-6} K^{-1}]	γ [-]
118	1.00	1.7215	26	92	0.0127

Table of fit parameters for expr. (5.17)

FIGURE 1.43. Density evolution versus degree of cure (65% filled compound, $T_{cure} = 120°C$ and $T_{cure} = 150°C$).

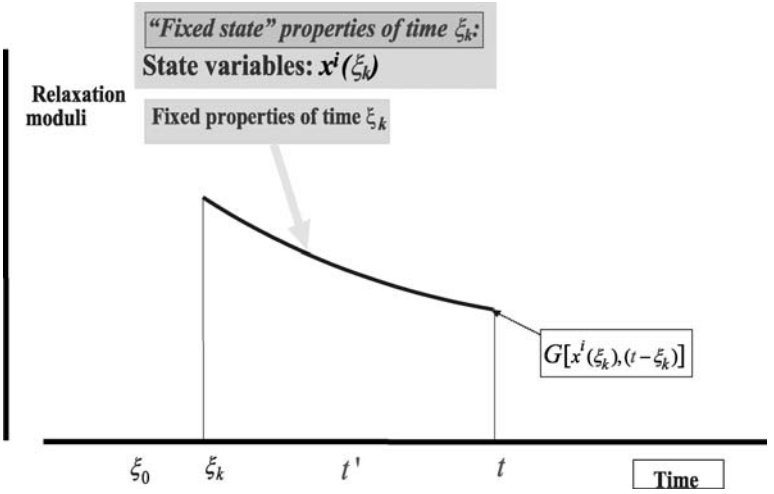


FIGURE 1.44. Illustration of "fixed state" relaxation curve.

elastic models are being developed. In such models the history of state parameters should be appropriately be included in the expressions for the relaxation moduli:

$$G(t) = G\left\{\mathcal{H}ist\left[x^i(\xi_k \to t)\right], t, \xi_k\right\}, \tag{1.116a}$$

$$K(t) = K\left\{\mathcal{H}ist\left[x^i(\xi_k \to t)\right], t, \xi_k\right\}. \tag{1.116b}$$

A model, based on assumptions of the changes in molecular mobility during cure is presented in [11]. This model can not simply be implemented into a finite element package and therefore will be used as a reference, to test approximate fully cure dependent models. Such an approximate model, which involves the history of state parameters, but requires only moderate adaptations of standard FEM packages, is presented in Section 1.5.2.1. The fully cure dependent model from [11] is summarized in Section 1.5.2.2. Finally Section 1.5.2.3 discusses future developments necessary for fast and non-isothermal curing.

1.5.2.1. Approximate Fully Cure Dependent Modeling In creating a "fully cure dependent" visco-elastic description, we realize that for thermo rheologically simple materials

the dependency of the moduli to changes in temperature are similar to changes in the degree of cure. For changing temperatures much research has already been done and merely resulted into an integral description for the reduced time [15]:

$$t_{red}(t') = \int_{-\infty}^{t'} a_T[T(s)]ds. \tag{1.117}$$

Assuming that modulus curves shift similar, due to cure and due to temperature, for curing materials and in analogy to (1.117) the following convolution integral expression is now proposed to replace Equation (1.106):

$$t_{red}(t') = \int_{-\infty}^{t'} a_T[T(s)] \cdot a_\alpha[\alpha(s)]ds. \tag{1.118}$$

If we (again) assume that for the materials in consideration the glassy plateau is hardly affected by the degree of cure or the temperature, we could continue to use Equations (1.104) and (1.108), except that we now use the alternative definition (1.118) for the reduced time. However, we should realize that the rubbery moduli are dependent on the degree of cure as well as on the temperature and thus also some history functionals should be used here. As an approximation of reality we abandon the creation of such history integrals. Instead, in the rubbery modulus the temperature dependency $T(\xi)$ will be replaced by $T(t)$, such that the "approximate fully cure dependent" viscoelastic description is given by:

$$G(t,\xi) = G\left\{\mathcal{H}ist\left[x^i(\xi_k \to t)\right], t, \xi_k\right\} = G_r[\alpha(\xi), T(t)] + \underbrace{\{G_g - G_r[\alpha(\xi), T(t)]\}}_{relaxation_amplitude} \cdot \sum_{n=1}^{N} h_n \cdot \exp\left[-\frac{t_{red}(t) - t_{red}(\xi)}{\tau_n^{ref}}\right], \tag{1.119}$$

$$K(t,\xi) = K\left\{\mathcal{H}ist\left[x^i(\xi_k \to t)\right], t, \xi_k\right\} = K_r[\alpha(\xi), T(t)] + \underbrace{\{K_g - K_r[\alpha(\xi), T(t)]\}}_{relaxation_amplitude} \cdot \sum_{n=1}^{N} k_n \cdot \exp\left[-\frac{t_{red}(t) - t_{red}(\xi)}{\tau_n^{ref}}\right]. \tag{1.120}$$

It should be noted, that this since this formulation starts out from the assumption that the glassy modulus G_g hardly changes during the evolution of the state parameters. Therefore only a horizontal shift on the time axes is sufficiently to account for the evolution of the state parameters (Figure 1.45). For the model molding compound used, the difference between the "fully state dependent model" from [11] and the "partly state dependent model" of the previous section, appeared to be less than 5%. The difference with the above "approximate state dependent model" was about 1%. However, these percentages were observed for slowly curing systems. There are indications, that the deviations for fast curing systems can be more significant.

1.5.2.2. Fully Cure Dependent Modeling In [11] the development of a "fully cure dependent" viscoelastic model is presented. In this work it is shown that the stress evolution is during cure is appropriately described through the following convolution integral expressions:

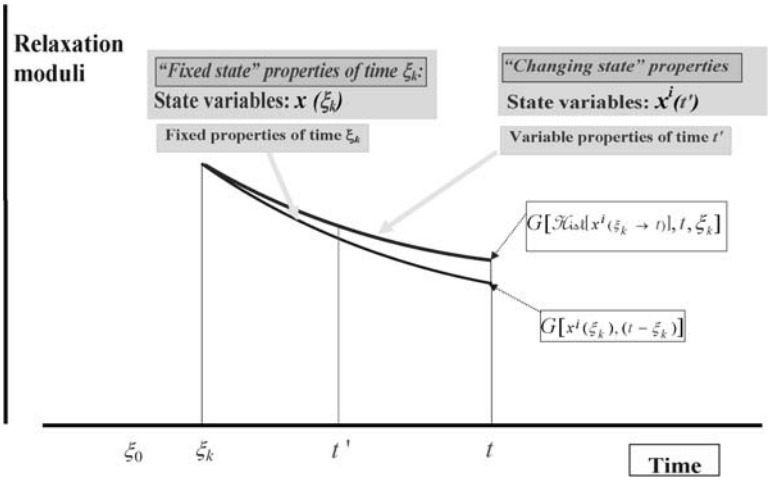


FIGURE 1.45. Comparison of "changing state" versus "fixed state" relaxation curves.

$$\sigma^d(t) = 2\int_{\xi=t_0}^{t}\left\{G_\infty(\xi) + \sum_{n=1}^{N}\left[G_n(\xi) + \int_{t'=\xi}^{t} G_n(t')\frac{\partial}{\partial t'}\exp\left[-\int_{t''=\xi}^{t'}\frac{dt''}{\tau_n(t'')}\right]dt'\right]\right\}\frac{d\varepsilon^d(\xi)}{d\xi}d\xi, \tag{1.121}$$

$$\sigma^v(t) = \int_{\xi=t_0}^{t}\left\{K_\infty(\xi) + \sum_{p=1}^{P}\left[K_p(\xi) + \int_{t'=\xi}^{t} K_p(t')\frac{\partial}{\partial t'}\exp\left[-\int_{t''=\xi}^{t'}\frac{dt''}{\tau_p^B(t'')}\right]dt'\right]\right\}\frac{d\varepsilon^v(\xi)}{d\xi}d\xi, \tag{1.122}$$

where: $G_\infty(\xi) = G_\infty[\alpha(\xi)] =$ (shear) rubber modulus, $K_\infty(\xi) = K_\infty[\alpha(\xi)] =$ (bulk) rubber modulus, $G_n(t') = G_n[\alpha(t')] =$ (shear) relaxation strength ($n = 1, \ldots, N$), $K_p(t') = K_p[\alpha(t')] =$ (bulk) relaxation strength ($p = 1, \ldots, P$), $\tau_n(t') = \tau_n[\alpha(t'), T] =$ (shear) relaxation times ($n = 1, \ldots, N$), $\tau_p^B(t') = \tau_p^B[\alpha(t'), T] =$ (bulk) relaxation times ($p = 1, \ldots, P$).

The relaxation strength G_n and K_p represent the participation functions of Prony series representations of "fixed degree of cure" expressions for the relaxation moduli:

$$G[\alpha(t'), T, \theta] = G_\infty[\alpha(t')] + \sum_{n=1}^{N} G^n[\alpha(t')]\cdot e^{-\theta/\tau_n[\alpha(t'),T]}, \tag{1.123}$$

$$K[\alpha(t'), T, \theta] = K_\infty[\alpha(t')] + \sum_{p=1}^{P} K^n[\alpha(t')]\cdot e^{-\theta/\tau_p^B[\alpha(t'),T]}. \tag{1.124}$$

Here: $\theta = t - \xi$.

In order to be able to use these expressions as a basis for the "fully cure dependent" Equations (1.121) and (1.122), the fitting process on "fixed degree of cure" data has been "conditioned" based on basic ideas of how stress relaxation passes through the different

relaxation curves at different conversion. The details of this process are beyond the scope of the present work. They will be published separately.

1.5.2.3. Future Trends in Cure Dependent Modeling In the previous sections, thermal mechanical models for the simulation of polymers during cure have been discussed. A so-called "partly cure dependent" visco-elastic model has first been introduced together with the experimental characterization of the model parameter functions (see Section 1.5.1). The effect of cure on the position of the viscoelastic transition region was modeled by a cure dependent shift factor. A separate term was used to account for the increase of the rubbery modulus during cure. Subsequently, in Section 1.5.2.1, an "approximate fully cure dependent" viscoelastic model was presented. For this no additional data is required. The history of state parameters is accounted for through redefining the reduced time through an appropriate integral expression. The models will be used in future verification studies. It is expected that the "approximate fully cure dependent" viscoelastic model is advantageous for application to fast curing compounds. Finally, in Section 1.5.2.2, the "fully cure dependent" model from [11] was briefly discussed. As this model can not simply be implemented in a standard finite element scheme, this model is merely considered as a reference.

An interesting question is the correct modeling for compounds where the glassy modulus G_g is changing with the evolution of the state parameters. Also, it is known that the rubbery modulus G_r slightly increases with temperature. So far, for both cases we have assumed constant plateaus and have based our models on this assumption. In case of fast changing temperatures during cure, or in case of very fast curing (snap curing), the validity of the presented models still has to be tested. This will be subject to continuing research.

1.6. PARAMETERIZED POLYMER MODELING (PPM)

The viscoelastic properties of thermosets are quite sensitive to relatively small changes in chemistry. Batch to batch variations in average monomer functionality or mixing ratio may therefore cause severe changes in product performance. For critical applications as in the electronics industry, the change in mechanical properties may affect the product reliability. In order to anticipate to these problems it is desirable to develop a fundamental understanding of what parameters influence the viscoelastic properties of thermosets. A quantitative model, such as will be presented in this chapter, has the additional advantage that it can be used within a package optimization process through virtual prototyping, where the polymer material parameters are (among other parameters) taken into account as design variables.

The main problem with thermoset polymers is the resin shrinkage during and after cure. This shrinkage results into stresses and deformations and as a consequence eventually in low yield. However, stress simulations for packages that include these materials, are complicated by the time-dependent (viscoelastic) nature of the mechanical properties: stresses partly relax after initiation and the relaxation rate depends on temperature and the degree of cure.

At present no models or theories exist which describe how the relaxation curves change with variations in monomer properties or with changes in the degree of cure. Here we present a new model based on the idea that the network (or crosslink) density is the most important parameter, which determines the thermoset viscoelasticity, and that all other issues are of minor importance. This model was tested by preparing epoxy systems in which the network density was varied in three, independent ways:

(1) by changing the conversion level,
(2) by changing the monomer functionality,
(3) by changing the mixing ratio.

From each of these systems the viscoelastic master curves were determined and compared to the model predictions.

1.6.1. PPM Hypotheses

As previously discussed (see Section 1.3.6), relaxation modulus curves consist of a glassy plateau at short loading times (low temperatures), a rubbery plateau at long loading times (high temperatures) and a transition region in between, such as illustrated in Figure 1.46.

Here the main features of an arbitrary relaxation curve are conveniently captured with the stretched exponential function (1.125) which contains four characteristic measures: the glassy and rubbery moduli, E_g and E_r, the relaxation time constant τ (position of the transition region) and power-law parameter n (steepness of the transition region). More detailed features of relaxation curves such as the small changes of the plateau moduli with temperature or second order transitions are disregarded in this study.

$$E' = E_r + (E_g - E_r)\exp[-(t/\tau)^n]. \tag{1.125}$$

In general, these four characteristic parameters may depend on chemical details such as the polarity and bulkiness of the groups and physical quantities such as the crosslink density, the average distance between crosslinks and the number of loosely attached chains or ring formation in the network. We now state that from all these issues the only important quantity is the crosslink density (the number of crosslinks per unit volume) and will show its validity below. According to the theory of rubber elasticity this crosslink density, ν_c, is directly related to the rubbery modulus

$$E_r = 3A\nu_c RT. \tag{1.126}$$

Here A is a constant (equal to unity), $R = 8.314$ J/mol/K is the gas constant and T is the absolute temperature. The glassy modulus, E_g, is proportional to the cohesive energy density and turns out to be roughly 3 GPa for all unfilled polymer resins. The relaxation

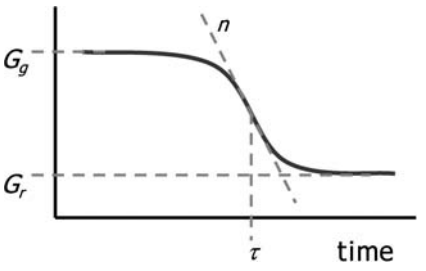


FIGURE 1.46. Illustration of commonly observed relaxation behavior with characteristic measures.

time constant τ depends strongly on the distance between the glass transition temperature T_g and the reference temperature T_{ref}:

$$\tau(T_{ref}) = \tau_{Tg} \frac{a_T(T_g)}{a_T(T_{ref})}. \tag{1.127}$$

Here a_T stands for the time-temperature shift factor. The glass transition temperature on the other hand, is known to vary linearly with the crosslink density [7]:

$$T_g = T_g^{lin} + B\nu_c. \tag{1.128}$$

Here T_g^{lin} is the T_g of the uncross-linked, linear polymer and B is a constant. For the dependency of parameter n on the crosslink density no a priori relation is known. In the sequence we will therefore investigate experimentally whether the supposed dependency on the crosslink density holds and how the individual parameters vary with this ν_c.

1.6.2. Experimental Characterizations

1.6.2.1. Materials A series of model epoxy systems was formulated with properties similar to the molding compounds used for chip encapsulation, and similar to the system used in Section 1.5. The resins consisted of an epoxy novolac (Araldite EPN series, ex Vantico), bisphenol-A as hardener and TPP (triphenyl phosphate) as catalyst. The epoxy monomers only differed in the number of epoxy groups (see Figure 1.47).

The monomer characteristics are listed in Table 1.1. All materials were used without further purification.

1.6.2.2. Sample Preparation The bisphenol-A was first dissolved at 165°C in half the amount of required epoxy (+15% extra to account for losses). This mixture was cooled

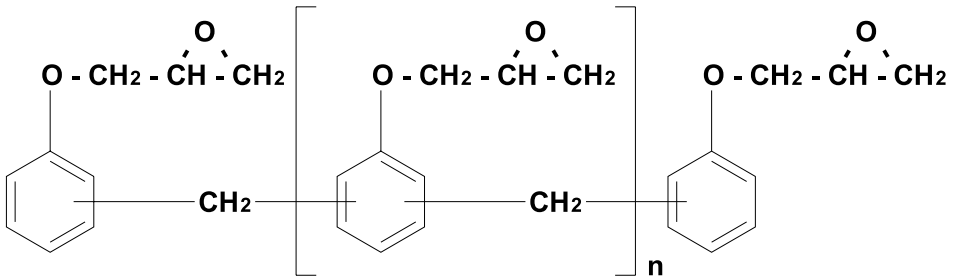


FIGURE 1.47. General formula of epoxy monomers ($n = 1.6$, 0.8 or 0.2).

TABLE 1.1.
Monomer properties.

	EPN 1180	EPN 1179	EPN 1178	Bisphenol-A
Epoxy equiv (g/eq)	175–182	172–179	169–179	114.15
Average functionality	3.6	2.8	2.2	2.0
M_n (g/mol)	534	398	380	228.29
M_w (g/mol)	761	464	440	

TABLE 1.2.
Properties and cure schedules of the three epoxy series.

Series	Epoxy EPN	mix ratio, r	α_{gel} Equation (1.129)	Cure schedule	T_g^{DSC} (°C)	Conversion*
1a	*1180*	*1.0*	*0.62*	*8h@150°C*	*112.7*	*0.97*
1b	1180	1.0	0.62	2.5h@150°C	105.5	0.93
1c	1180	1.0	0.62	1h@150°C	100.6	0.90
1d	1180	1.0	0.62	3h@120°C	94.5	0.87
1f	1180	1.0	0.62	1.5h@130°C	85.8	0.81
2a	*1180*	*1.0*	*0.62*	*8h@150°C*	*112.7*	*0.97*
2b	1179	1.0	0.75	8h@150°C	92.3	
2c	1178	1.0	0.91	8h@150°C	89.3	
3a	1180	0.6	0.48	8h@150°C	82.3	
3b	1180	0.8	0.55	8h@150°C	92.3	
3c	*1180*	*1.0*	*0.62*	*8h@150°C*	*112.7*	*0.97*
3d	1180	1.2	0.68	8h@150°C	106.6	
3e	1180	1.4	0.73	8h@150°C	106.1	

*The conversion levels are determined from DSC measurements, see [12]. Standard conditions are shown in italics.

down to 90°C. The catalyst (0.5 g TPP/ 100 g epoxy) was then dissolved in the other half of the epoxy (at 90°C) and mixture A was added (ensuring that the extra 15% remains in the bottle). The mixture was stirred for 5 minutes at 90°C and slowly cast into metal molds which were pretreated with Lossing-A50 wax (ex Poly-Service BV) and cured according to the schedule in Table 1.2. The nominal size of the rectangular test specimen was $55 \times 10 \times 2$ mm^3.

The conversion at gelation, α_{gel}, is calculated from the Flory relation

$$\alpha_{gel} = \left[\frac{r}{(f_{ep} - 1)(f_{BA} - 1)} \right]^{0.5}. \tag{1.129}$$

Here r is the stoichiometric (or mixing) ratio (hydroxy groups/epoxy groups) and f_{ep} and f_{BA} are the epoxy and bisphenol-A functionality (number of reactive groups per monomer).

1.6.2.3. Equipment and Test Procedures A TA-Instruments DSC 2920 (differential scanning calorimeter) was used to measure both the glass transition and the degree of cure (conversion). The standard heating rate was 10 K/min.

Dynamic mechanical elongation tests (see Figure 1.30) were performed in a Metravib VA4000 viscoanalyzer with continuous frequency sweeps (0.3 to 60 Hz) during a temperature scan at 1 K/min. The relatively low amount of catalyst ensured that the reaction of the partly cured samples (1b–1f) did not proceed significantly during the DMA temperature scan. The data of each DMA experiment was automatically shifted along the frequency axis according to the time-temperature superposition principle to produce a master curve and a shift factor curve. A non-linear least square fit was then applied to the master curve data to generate the fit parameters used in Equation (1.125).

1.6.2.4. Discussion of Results The PPM theory was tested using three series of epoxy novolac resins in which the conversion, the epoxy functionality (number of reacting groups per molecule) and the mixing ratio was varied systematically. In all three series the crosslink density varied between 300 and 1500 mol/m^3 whereas the chemical and physical parameters change considerably. In the mixing ratio series, for example, the nature of

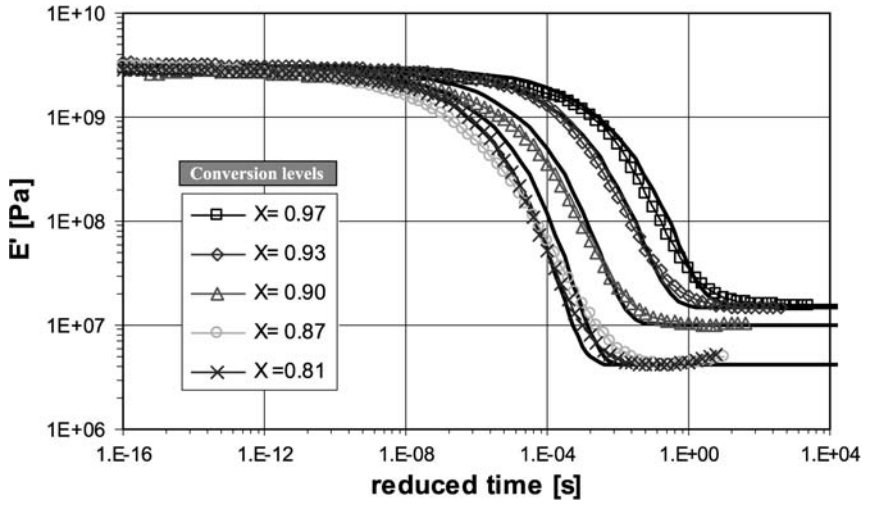


FIGURE 1.48. Master curves for epoxies with different conversion levels $T_{ref} = 120°C$. [Full lines are obtained from a fit to Equations (1.125–1.128)].

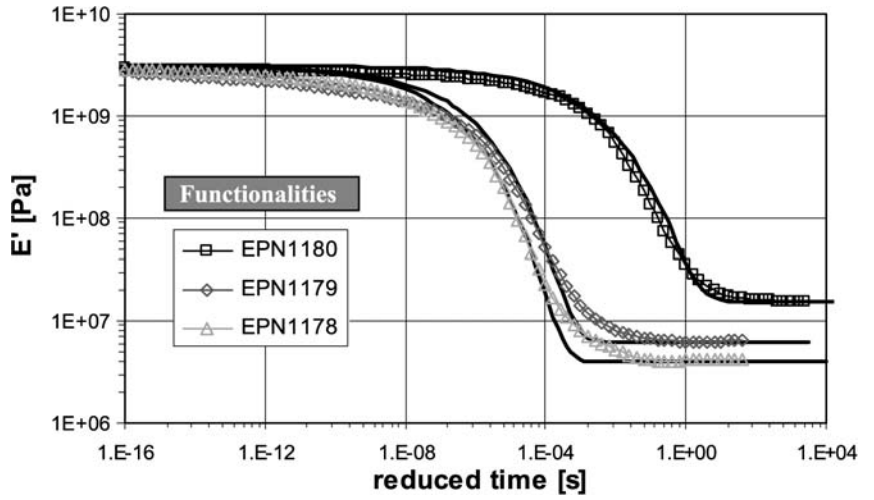


FIGURE 1.49. Master curves for epoxy series with different functionalities $T_{ref} = 120°C$. [Full lines obtained from a fit to Equations (1.125–1.128)].

the unreacted chain ends change from mainly epoxide groups to hydroxy groups as r increases. In the conversion series with $r = 1.0$ there is always a balance between the number of unreacted epoxide and hydroxy groups, while in the series with changing functionality the crosslink density is changed without increasing the number of unreacted groups.

The master curves for the three different series are shown in Figures 1.48–1.50. All curves show a clear glassy plateau and a rubbery plateau which changes with changing conversion, functionality and mixing ratio. The glass transition region is seen to shift towards shorter times as the rubbery plateau decreases. More quantitative data of these changes can be found in Table 1.3, where we listed the fit parameters. The full lines in Figures 1.47–1.49

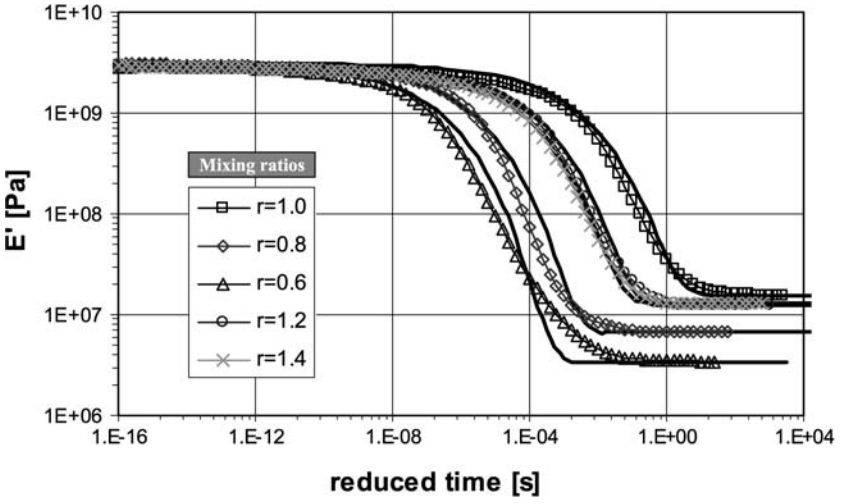


FIGURE 1.50. Master curves for epoxies with different mixing ratios. [Full lines obtained from a fit to Equations (1.125–1.128).]

TABLE 1.3.
Viscoelastic parameters for the three epoxy series obtained from individual fits. (The crosslink density is calculated from E_r using Equation (1.126)).

Series	ν_c mol/m^3	E_g GPa	E_r MPa	$\tau(T_{ref})$ s	$\tau(T_g^\delta)$ 10^{-3} s	n	T_g^δ (°C)
1a	1479	3.09	15.6	$5.1\ 10^{-2}$	1.32	0.276	120.9
1b	1412	3.50	14.9	$6.2\ 10^{-3}$	1.12	0.274	116.6
1c	964	2.81	10.3	$3.7\ 10^{-4}$	1.04	0.282	106.9
1d	403	3.14	4.3	$3.9\ 10^{-5}$	0.84	0.245	96.4
1f	393	3.05	4.1	$3.9\ 10^{-3}$	0.87	0.307	93.4
2a	1479	3.09	15.6	$5.1\ 10^{-2}$	1.32	0.276	120.9
2b	594	2.74	6.3	$2.6\ 10^{-5}$	0.90	0.275	100.9
2c	384	3.13	4.2	$1.4\ 10^{-5}$	0.89	0.284	96.5
3a	323	2.92	3.5	$9.2\ 10^{-6}$	1.07	0.254	89.3
3b	627	3.31	6.7	$4.6\ 10^{-5}$	1.26	0.296	102.0
3c	1479	3.09	15.6	$5.1\ 10^{-2}$	1.32	0.276	120.9
3d	1168	3.10	12.2	$2.6\ 10^{-3}$	1.17	0.275	113.7
3e	1208	3.15	12.8	$1.5\ 10^{-3}$	1.15	0.281	112.3

are fits using the measured rubbery modulus to obtain the crosslink density. The rest of the fit parameters were kept constant for all curves. A closer inspection shows that the shapes of curves with similar crosslink density (like samples 1d, 1f and 2c or 3d and 3e) are almost indistinguishable. Note also that a small excess of epoxy ($r < 1.0$ in Figure 1.50) has a much larger effect on the relaxation curve than a small excess of hydroxy groups ($r > 1.0$).

The T_g^δ listed in Table 1.3 is the T_g obtained from the maximum in phase angle δ at a frequency of 1 Hz. A detailed analysis showed that this mechanical glass transition

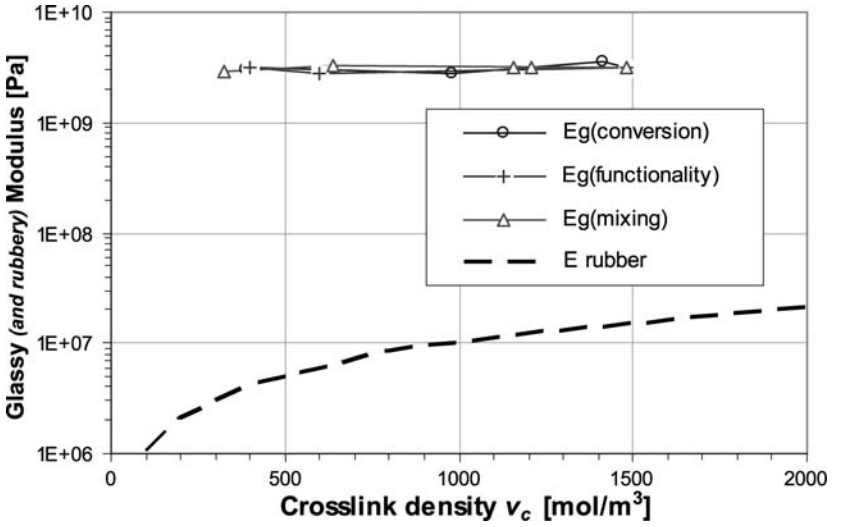


FIGURE 1.51. Change of glassy modulus with increasing crosslink density. The rubber modulus is given for comparison.

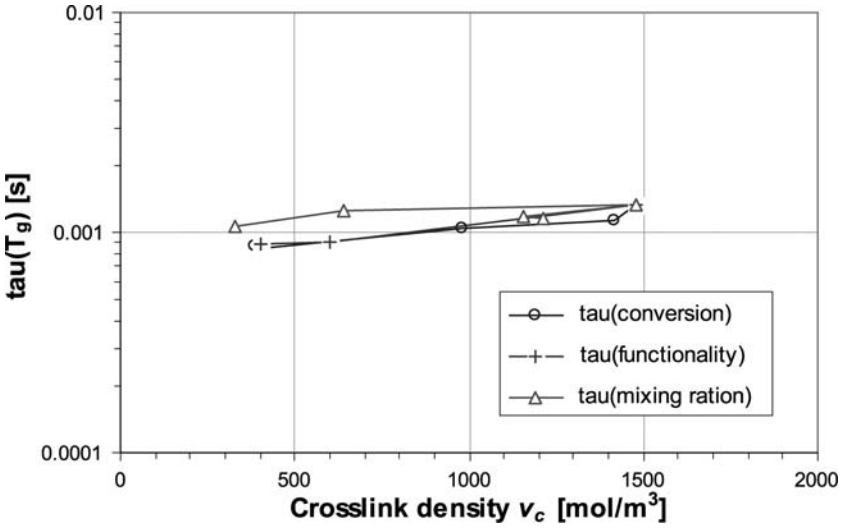


FIGURE 1.52. Change of relaxation time constant (with $T_{ref} = T_g^{\delta}$) versus crosslink density for the three data series.

temperature is related to the applied frequency and the DSC glass transition temperature (10 K/min) as:

$$T_g^{\delta} = T_g^{DSC} + 5.67 * \log(freq) + 7.6. \tag{1.130}$$

The effects of the individual fit parameters are shown in Figures 1.51–1.53. Figure 1.51 clearly shows that the glassy modulus is indeed independent of the crosslink density. Its average value amounts to 3.1 GPa. The relaxation time constant at a fixed

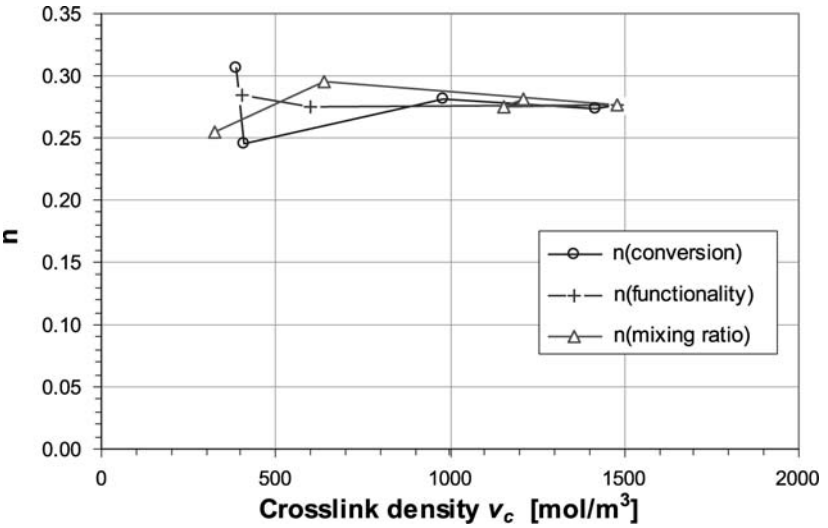


FIGURE 1.53. Effect of crosslink density on the power-law coefficient n.

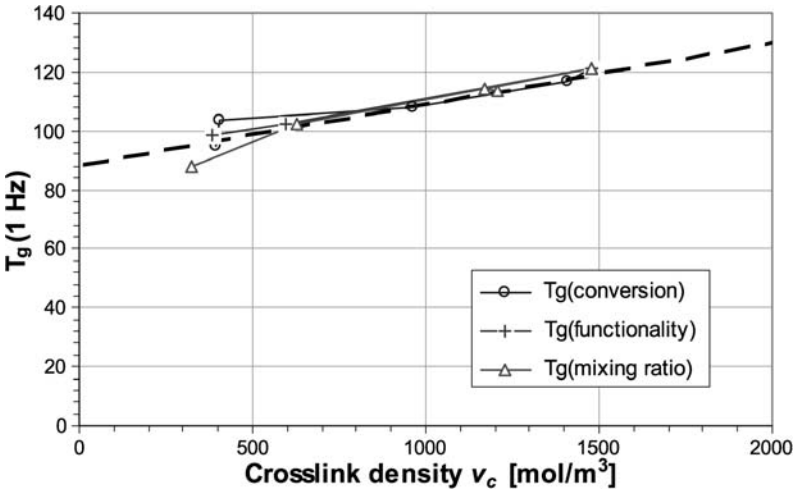


FIGURE 1.54. Glass transition as a function of crosslink density. Dashed line: fit to Equation (1.128) with $T_g^{lin} = 88.0°\text{C}$ and $B = 0.021°\text{C m}^3/\text{mol}$.

$T_{ref} = 120°\text{C}$ changes by four orders of magnitude (see Table 1.3). The relaxation time constant at the glass transition temperature, however, turns out to be fairly constant and equal for all epoxy series (Figure 1.52). Also the power-law coefficient n is more or less constant for all materials (Figure 1.53). The glass transition temperature (Figure 1.54) increases linearly with increasing conversion, as was expected from Equation (1.128).

Up to this point we can thus conclude that the crosslink density is indeed the scaling parameter for the relaxation curves and that (for the epoxy systems used in this study) the parameters E_g, τ_{Tg} and n turn out to be independent of the crosslink density. The last item required for the PPM model is an expression for the shift factor a_T.

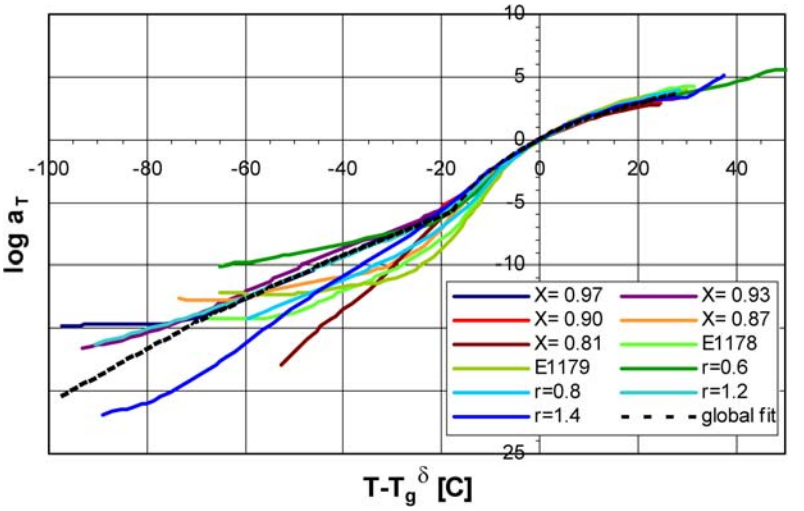


FIGURE 1.55. Shift factor curves for all data series.

In Figure 1.55 all individual shift factor curves with T_g^δ as a reference temperature are plotted versus $T - T_g^\delta$. It can be seen that in the glass transition region (from T_g −10°C to T_g +30°C) all curves coincide and that in the glassy plateau region the shift factor curves show a wide variation. The latter, however, does not have a large effect on the master curves since in the glassy region the relaxation behavior is almost frequency and time independent. We therefore simply fitted the shift factor of the standard epoxy system with a WLF equation for the rubbery part and an Arrhenius equation for the glassy part:

$$\log a_T = \frac{C_1(T - T_{ref})}{C_2 + T - T_{ref}} \quad T > T_c, \tag{1.131}$$

$$\log a_T = \frac{-H}{2.303R}\left(\frac{1}{T} - \frac{1}{T_0}\right) \quad T < T_c, \tag{1.132}$$

with $C_1 = 9.87$, $C_2 = 47.9$ K, $T_c = T_g^\delta - 18$ and $H = 391.4$ kJ/mol. T_0 is determined such that Equations (1.131) and (1.132) match at $T - T_c$. The resulting fit is shown in Figure 1.55 as the dashed line.

As a next step we determined the average fit parameters (resulting in $E_g = 3.1$ GPa, $\tau_{Tg} = 1.11 \times 10^{-3}$ s, $n = 0.27$, $T_g^{lin} = 88$°C and $B = 0.021$°C m^3/mol) and used these, together with the shift factor equations, to test the suitability of Equations (1.125–1.128) for parameterized modeling. The results are shown in Figures 1.48–1.50 as the full lines. It can be seen that the model predictions agree fairly well with the experimental relaxation curves. We therefore conclude that the present approach works (at least for this simple epoxy system) and that it now becomes possible to predict the full viscoelastic behavior of a thermoset resin from a limited set of parameters only. Moreover, the results clearly show that for the viscoelastic behavior of a material the crosslink density is the only important parameter. Details of how this crosslink density is obtained appear not to be important at all.

1.6.3. PPM Modeling in Virtual Prototyping

The parameterized polymer modeling (PPM) approach as presented in the previous sections, not only enables us to generate realistic relaxation curves which can be used in optimization procedures. It gives a realistic cure and temperature dependent description, where only four parameters act as polymer related design variables. These are: the glassy and rubbery plateau moduli and the position and width of the transition region. These parameters were seen to depend only on the network density. Changes in the chemistry of a resin system such as the mixing ratio, conversion or monomer functionality affected the relaxation curve only through changes in this network density and did not show any other dependencies.

It is a future challenge to apply this model in virtual prototyping of microelectronics components, in order to create optimal reliability and to minimize package warpage by optimal choice of process parameters. Results of simulations will be separately published.

ACKNOWLEDGMENTS

The authors acknowledge the European Community for their financial support (Mevipro project GRD1-2001-40296), Denka for the silica fillers and Vantico for the supply of epoxy resin and the corresponding molecular weight data.

REFERENCES

1. H.J.L. Bressers, P. Beris, J. Caers, and J. Wondergem, Influence of chemistry and processing of flip chip underfills on reliability, Adhesives in Electronics'96, June 3–5, Stockholm, Sweden, 1996, p. 306.
2. L.J. Ernst, C. van 't Hof, D.G. Yang, M.S. Kiasat, G.Q. Zhang, H.J.L. Bressers, J.F.J. Caers, A.W.J. den Boer, and J. Janssen, Mechanical modeling and characterization of the curing process of underfill materials, ASME Journal of Electronic Packaging, Transactions of ASME, 124(2), pp. 97–105 (2002).
3. L.J. Ernst, K.M.B. Jansen, C. van 't Hof, D.G. Yang, G.Q. Zhang, and H.J.L. Bressers, Recent developments in thermal mechanical modeling of the curing process of filled thermoset polymers, Proceedings of the 10th Int Conference on Mechanics and Technology of Composite Materials, Sofia, September 2003, pp. 38–44.
4. L.J. Ernst, K.M.B. Jansen, G.Q. Zhang, and R. Bressers, Time and temperature dependent thermo-mechanical modeling of a packaging molding compound and its effect on packaging process stresses, ASME Journal of Electronic Packaging, Transactions of ASME, 125(Dec.), pp. 539–548 (2003), ISSN 1043-7398.
5. L.J. Ernst, D.G. Yang, K.M.B. Jansen, C. van 't Hof, G.Q. Zhang, and W.D. van Driel, On the effect of cure-residual stress on flip chip failure prediction, Proceedings of the 4th Electronics Packaging Technology Conference, Singapore, December 2002, 2003, pp. 398–403, ISBN 0-7803-7435-5.
6. X.J. Fan, G.Q. Zhang, W. van Driel, and L.J. Ernst, Mechanism-based delamination prediction during reflow with moisture preconditioning, in L.J. Ernst, G.Q. Zhang, P. Rodgers, and O. de Saint Leger, Eds., Thermal and Mechanical Simulation and Experiments in Micro-Electronics and Micro-Systems, IEEE Press, 2004, pp. 329–336, ISBN 0-7803-8420-2.
7. Fox, T.G. and S. Loshaek, J. Polym. Sci., 15, p. 371 (1955).
8. M.A.J. Gils, W.D. van Driel, G.Q. Zhang, H.J.L. Bressers, and J.H.J. Jansen, Virtual qualification of moisture induced failures of advanced packages, in L.J. Ernst, G.Q. Zhang, P. Rodgers, and O. de Saint Leger, Eds., Thermal and Mechanical Simulation and Experiments in Micro-Electronics and Micro-Systems, IEEE Press, 2004, pp. 157–162, ISBN 0-7803-8420-2.
9. C. van 't Hof, P. Mohanty, and D.J. Rixen, Testing a dynamic mechanical analyzer: influence of the measuring column dynamics, Proceedings of IMAC-XXI, A Conference on Structural Dynamics, The Hyatt Orlando, Kissimee, FL, February 3–6, 2003.
10. C. van 't Hof, L.J. Ernst, K.M.B. Jansen, D.G. Yang, H.J.L. Bressers, and G.Q. Zhang, A novel tool for cure dependent viscoelastic characterization of packaging polymers, in L.J. Ernst, G.Q. Zhang, P. Rodgers, and O. de Saint Leger, Eds., Proceedings of EuroSIME 2004 (Thermal and Mechanical Simulation and

Experiments in Micro-Electronics and Micro-Systems), May 2004, Brussels, Belgium, IEEE Catalog Number 04EX831, ISBN (book) 0-7803-8420-2, Library of Congress Number 2004102577, IEEE Press, 2004, pp. 385–390.
11. C. van 't Hof, Mechanical characterization and modeling of curing thermosets, Ph.D. Thesis, Delft University of Technology, 2005.
12. K.M.B. Jansen, L. Wang, D.G. Yang, C. van 't Hof, L.J. Ernst, H.J.L. Bressers, and G.Q. Zhang, Cure, temperature and time dependent constitutive modeling of moulding compounds, in L.J. Ernst, G.Q. Zhang, P. Rodgers, and O. de Saint Leger, Eds., Proceedings of EuroSIME 2004 (Thermal and Mechanical Simulation and Experiments in Micro-Electronics and Micro-Systems), pp. 581–586, May 2004, Brussels, Belgium, IEEE Catalog Number 04EX831, ISBN (book) 0-7803-8420-2, Library of Congress Number 2004102577, IEEE Press, 2004.
13. M.S. Kiasat, Curing shrinkage and residual stresses in viscoelastic thermosetting resins and composites, Ph.D. thesis, Delft University of Technology, Delft, The Netherlands, 2000.
14. J. Lange, et al., Polymer, 37(26), pp. 5859–5868 (1996).
15. R.A. Schapery, in G.P. Sendeckyj, Ed., Composite Materials, Vol. 2, Academic Press, New York, 1974, pp. 97–111.
16. E.G. Septanika and L.J. Ernst, Application of the network alteration theory for the modeling the time-dependent constitutive behaviour of rubbers, Part 1, General theory, Mechanics of Materials, 30(4), pp. 253–263 (1998).
17. E.G. Septanika and L.J. Ernst, Application of the network alteration theory for the modeling the time-dependent constitutive behaviour of rubbers, Part 2, Experimental verification, Mechanics of Materials, 30(4), pp. 265–273 (1998).
18. E.G. Septanika and L.J. Ernst, Hysteresis and time-dependent constitutive modeling of filled vulcanized rubber, J. Phys. IV France, 9, pp. 63–72 (1999).
19. N.W. Tschoegl, The Phenomenological Theory of Linear Viscoelasticity, A Introduction, Springer-Verlag, Berlin, 1989.
20. D.G. Yang, L.J. Ernst, G.Q. Zhang, W.D. van Driel, and J.H.J. Janssen, Parameter sensitivity study of cure-dependent underfill properties on flip chip failures, in Proc. of the 52th Electronic Components and Technology Conference (ECTC2002), San Diego, May 28–May 31, 2002, ISBN 0-7802-7430-4, 2002, pp. 865–871.
21. D.G. Yang, L.J. Ernst, K.M.B. Jansen, C. van 't Hof, G.Q. Zhang, and W.D. van Driel, On the effect of cure-residual stress on flip chip failure prediction, in Proceedings of the 4th Electronics Packaging Technology Conference, Singapore, December 2002, ISBN 0-7803-7435-5, 2002, pp. 398–403.
22. D.G. Yang, K.M.B. Jansen, L.J. Ernst, G.Q. Zhang, W.D. van Driel, and H.J.L. Bressers, Modeling of cure-induced warpage of plastic IC packages, in L.J. Ernst, G.Q. Zhang, P. Rodgers, and O. de Saint Leger, Eds., Thermal and Mechanical Simulation and Experiments in Micro-Electronics and Micro-Systems, ISBN 0-7803-8420-2, IEEE Press, 2004, pp. 33–40.
23. D.G. Yang, K.M.B. Jansen, L.J. Ernst, G.Q. Zhang, W.D. van Driel, H.J.L. Bressers, and X.J. Fan, Prediction of process-induced warpage of IC packages encapsulated with thermosetting polymers, Proceedings of the 54rd Electronic Components and Technology Conference, Las Vegas, USA, June 1–4, 2004.
24. L. Zhang, L.J. Ernst, and H.R. Brouwer, A study of nonlinear viscoelasticity of an unsaturated polyester resin, part 1, uniaxial model, part 2, 3D-model, Mechanics of Materials, 26(3), pp. 141–195 (1997).
25. L. Zhang, L.J. Ernst, and H.R. Brouwer, Transverse behavior of unidirectional composite (glass fibre reinforced polyester), Part 1, Fibre packing geometry influence, Part 2, Initial strain influence, Mechanics of Materials, 27, pp. 13–61 (1998).
26. G.Q. Zhang, A. Tay, and L.J. Ernst, Virtual thermo-mechanical prototyping of electronic packaging—Bottlenecks and solutions of damaging modelling, 3rd Electronic Packaging Technology Conference (EPTC), Singapore, 2000.
27. G.Q. Zhang, L.J. Ernst, S. Liu, Z.F. Qian, A.A.O. Tay, H.J.L. Bressers, and J. Janssen, Virtual thermo-mechanical prototyping of electronic packaging-challenges in material characterization and modeling, Proc. of the 51th Electronic Components and Technology Conference (ECTC2001), Orlando, May 29–June 1, 2001, IEEE, Piscataway, NJ, 2001, ISBN: 0-7803-7038-4, ISSN: 0569-5503, 2001, pp. 1479–1486.
28. G.Q. Zhang, The challenges of virtual prototyping and qualification for future microelectronics, J. Microelectronics Reliability, 43, pp. 1777–1785 (2003).
29. R. Zwiers, H.J.L. Bressers, B. Ouwehand, and D. Baumann, Development of a new low stress Hyperred LED encapsulant, IEEE Transactions on Components, Hybrids and Manufacturing Technology, 12(3) (1989).
30. E.H. Wong, et al., Comprehensive treatment of moisture induced failure-recent advances, IEEE Transactions on Electronic Packaging Manufacture, 25, pp. 223–230 (2002).

2

Thermo-Optic Effects in Polymer Bragg Gratings

Avram Bar-Cohen, Bongtae Han, and Kyoung Joon Kim

University of Maryland, College Park, Maryland, USA

2.1. INTRODUCTION

In spite of their relatively high light absorption rate, polymer materials provide a potent alternative to conventional optical materials due to low-cost, ease of fabrication and assembly, and compatibility with other materials. The recent literature reveals a rapidly increasing interest in the use of polymer components in photonics systems. While polymer waveguides are currently receiving much of the attention, it is to be expected that signal management requirements will lead to progressively greater efforts in gratings, mirrors, and lenses.

A Bragg grating (BG) in a light transmitting waveguide produces a very narrow band of reflected optical energy, with a maximum reflectivity at the characteristic wavelength of the grating, called the *Bragg wavelength*, as illustrated in Figure 2.1. Unlike the conventional glass fibers, the index of refraction in light-transmitting polymers typically varies inversely with the temperature, leading to negative thermo-optic coefficients that are 10 to 30 times greater than the positive thermo-optic coefficient of conventional silica glass [1]. This strong negative thermo-optical characteristic imbues polymer BGs, generally packaged on low thermal expansion substrates, with precise wavelength discrimination when used as tuning filters.

Another important characteristic of polymeric optical materials is their relatively high light absorption rates, at approximately 0.2 dB/cm, as compared to 0.2 dB/km for glass fibers at wavelengths of 1550 nm. However, as shown in Figure 2.2 for the acrylic-based polymers of allied signal [1], the absorption rates of polymeric optical materials strongly depend on the wavelength. For example, in acrylic with full CH content the absorption rate (0.5 dB/cm) at 1.55 μm is 25 times greater than that at 0.8 μm.

Intrinsic self-heating, resulting from these high absorption rates, can produce considerable temperature changes within polymer waveguides and gratings. Such self-heating, as well as the induced temperature gradients and possible changes in the ambient temperature, can cause undesirable shifts in the Bragg wavelength and changes in the reflectivity/transmissivity of the grating. To facilitate the selection of polymeric optical materials

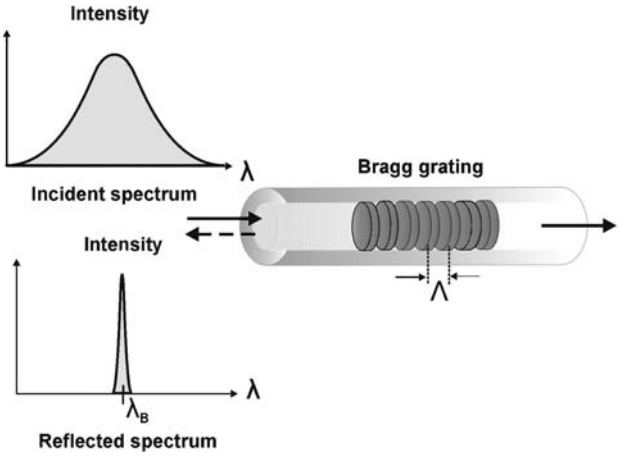


FIGURE 2.1. Fundamental characteristics of Bragg grating.

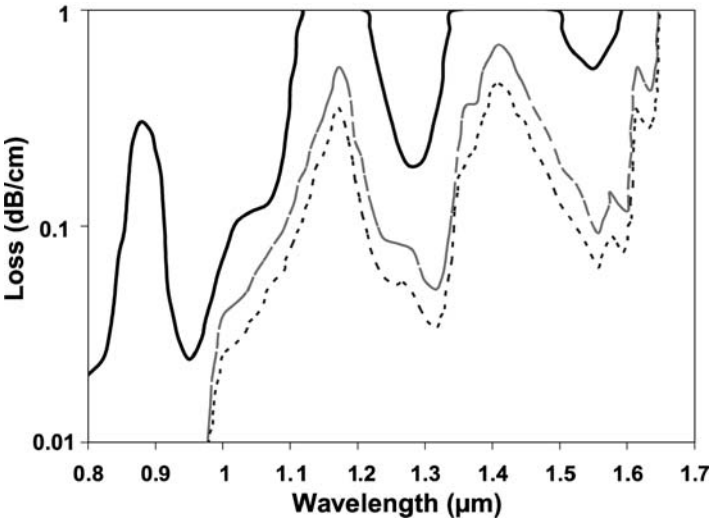


FIGURE 2.2. Light absorption in polymer systems (allied signal acrylic: solid line—100% CH; dashed line—30% CH; dotted line—20% CH) [1].

and the rational design of polymer BGs, a complete understanding of thermo-optic behavior of polymer BGs is required.

This chapter, thus, focuses on the derivation and exploration of a thermo-optical model that can be used to characterize the thermally-induced optical behavior of a polymer fiber Bragg grating (PFBG). The results and implications of an extensive parametric study, using two distinct light sources, are presented and discussed.

2.2. FUNDAMENTALS OF BRAGG GRATINGS

2.2.1. Physical Descriptions

A Bragg grating (BG) can be formed by index modulations in the waveguide material, induced by exposure to ultraviolet (UV) light and/or doping [2–4], or by trenching (interrupting) of the waveguide material with etchants and/or focused laser beams [5]. A BG formed by UV exposure and/or doping, without physical trenches, is referred to as a *bulk index grating*, and a BG formed by physical deformation is referred to as a *surface relief grating*. Bragg gratings (BGs) can be found in numerous photonic components and systems, including distributed feedback laser diodes (DFB LD), distributed Bragg reflector laser diodes (DBR LD), optical fibers, and planar waveguides.

Two most popular forms of BGs are illustrated in Figures 2.3 and 2.4; a fiber Bragg grating (FBG) and a planar Bragg grating (PBG). A FBG is a BG in an optical fiber, and a PBG is a BG in a planar waveguide. A FBG is usually a bulk index grating, while both surface relief and bulk indexing are used to create a PBG.

Figure 2.5 illustrates the most widely used *phase-mask* technique to inscribe a BG in an optical fiber [3], using a diffractive optical component (a transmission diffraction grating) as a phase-mask. The phase mask is especially fabricated to suppress the zeroth order diffraction, while the ± 1st diffraction orders interfere to produce a sinusoidal intensity distribution in space. A doped photosensitive optical fiber is placed in the region where the two beams intersect and a BG is imprinted on a glass or polymer fiber. Inscription of a BG in a polymer optical fiber requires more rigorous fabrication techniques because of its much higher UV absorption rate [6].

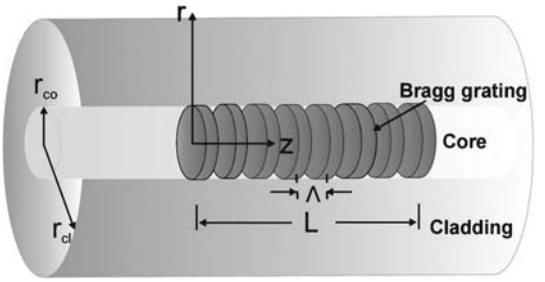


FIGURE 2.3. Fiber Bragg grating structure.

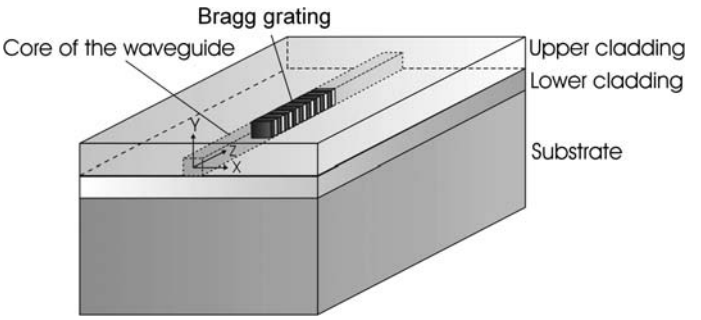


FIGURE 2.4. Packaged planar Bragg grating structure.

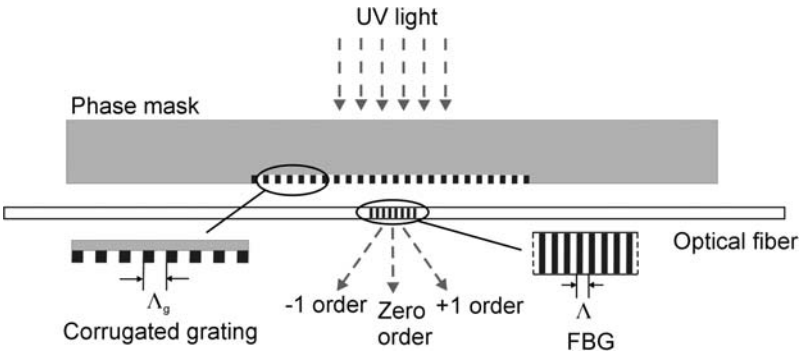


FIGURE 2.5. Phase-mask technique for fiber Bragg grating fabrication.

Since formation of a permanent BG in a glass optical fiber was first demonstrated by Hill et al. [4], BGs integrated fibers and other waveguides have been among the key components in photonic systems. They have been used widely for filtering, switching and stabilizing optical signals in telecommunication applications such as Wavelength Division Multiplexing/Demultiplexing (WDM) and Optical Add-Drop Multiplexing (OADM) [5]. With their simplicity and unique filtering characteristics, BGs have also been used as narrow band reflectors in wavelength stabilized lasers, fiber lasers, and pump amplifiers [2,5]. More recently, FBG sensors have been proposed as future medical sensor systems to measure temperatures during medical treatments [7–11]. A FBG is dielectric, and thus unlike thermocouples and thermistors, or other conventional temperature measuring electronic devices, it is immune to electromagnetic interference.

Polymer BGs have been proposed as passive filters [12–14], tuning filters [15,16], WDM systems [1,17–19] and couplers [20]. A very narrow spectral bandwidth of ∼0.2 nm at 1550 nm was demonstrated using passive polymer BG filters with uniform grating periods [12], and a even narrower band of 0.03 nm at 1290 nm was demonstrated using phase shifted BG filters [13]. Polymer BGs have also been demonstrated as filters in WDM systems for short haul data transmission with a wavelength tolerance of 0.2 to 0.5 nm/K [1,15]. As a future solution for an economical multiplexing system, polymer BG-based optical add/drop multiplexers (OADM's) were introduced with a channel spacing of 400 GHz (3.2 nm) and a thermal stability of 0.04 nm/K [19].

2.2.2. Basic Optical Principles

The axial refractive index modulation in a BG is typically represented in sinusoidal form [21], as illustrated in Figure 2.6 and expressed in Equation (2.1)

$$\delta n_{eff}(z) = \overline{\delta n_{eff}(z)}\left[1 + \cos\left(\frac{2\pi z}{\Lambda} + \phi(z)\right)\right], \tag{2.1}$$

where $\delta n_{eff}(z)$ is the modulation of the effective refractive index, $\overline{\delta n_{eff}(z)}$ is the average modulation of the effective refractive index, Λ is the grating period, and $\phi(z)$ is a grating chirp, i.e., the inherent or fabricated non-uniformity in the grating period.

As light is incident on the fiber, entering from the left in Figure 2.6, the refractive index modulation results in the reflection of a narrow bandwidth of light, centered around

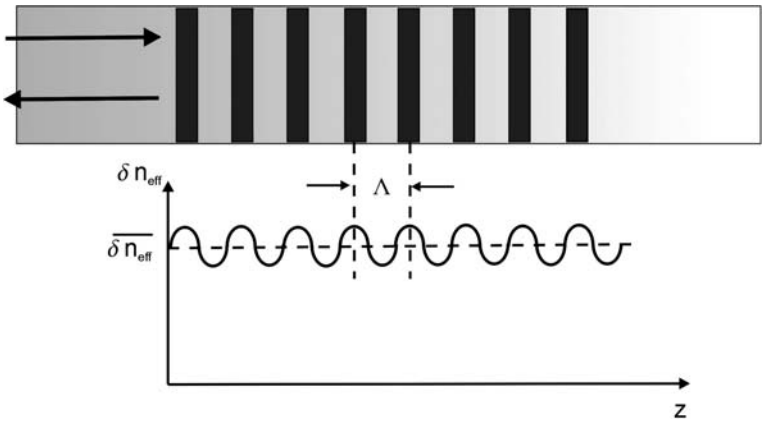


FIGURE 2.6. Refractive index modulation along Bragg grating.

the Bragg wavelength, λ_B. This characteristic wavelength is determined by the effective refractive index and the grating period and can be defined as [21]

$$\lambda_B = 2n_{eff}\Lambda, \tag{2.2}$$

where n_{eff} is the effective refractive index.

The intrinsic optical parameters of FBGs include the refractive indices of the fiber core and the cladding (n_{co} and n_{cl}), the index modulation ($\overline{\delta n_{eff}(z)}$), and the grating period (Λ). From these intrinsic parameters, working through several defined optical parameters, the effective refractive index (n_{eff}), Bragg wavelength (λ_B), coupling coefficient (κ) (relating to the interaction between the incident and reflected waves in the BG), and maximum reflectivity ($\rho_{\max}$) can be determined.

For single mode operation, the Numerical Aperture, *NA* is defined [22] as

$$NA = \sin\phi_{\max} = \sqrt{n_{co}^2 - n_{cl}^2}, \tag{2.3}$$

where $\phi_{\max}$ is the critical angle (or the maximum incident angle for total internal reflection).

For single mode operation, the generalized frequency, V, defined in Equation (2.4), must be maintained below 2.4048 [22], i.e.,

$$V = \frac{2\pi r_{co}}{\lambda} NA < 2.4048, \tag{2.4}$$

where r_{co} is the core radius.

This leads a relation for the generalized guide index, b, given by Equation (2.5)

$$b = \left(1.1428 - \frac{0.996}{V}\right)^2. \tag{2.5}$$

Reflecting the combined effects of the numerical aperture, guide index, and cladding refractive index, the effective index of refraction, n_{eff}, can be expressed as [22]

$$n_{eff} = \sqrt{b \cdot (NA)^2 + n_{cl}^2}. \tag{2.6}$$

Substituting Equation (2.2) through (2.5) into Equation (2.6) and rearranging the terms, the n_{eff} relation takes the following form

$$\left[1 - \frac{0.996^2}{(\pi r_{co})^2}\Lambda^2\right] n_{eff}^2 + 2.2765\frac{\Lambda \cdot NA}{\pi r_{co}} n_{eff} - n_{cl}^2 - (1.1428)^2 NA^2 = 0, \tag{2.7}$$

where single mode operation dictates that the solution of Equation (2.7) for the effective index must satisfy the following relationship

$$n_{eff} > \frac{\pi \cdot r_{co} NA}{2.4048 \cdot \Lambda}.$$

Using the intrinsic parameters (r_{co}, n_{co}, n_{cl}, and Λ) of a FBG, the effective index of the FBG can be evaluated using Equation (2.7).

For a single mode BG [23,24], the coupling coefficient, κ is defined as

$$\kappa = \frac{\pi}{\lambda}\overline{\delta n_{eff}}. \tag{2.8}$$

Then, the maximum reflectivity, produced at the Bragg resonance condition, is

$$\rho_{\max} = \tanh^2(\kappa_B L), \tag{2.9}$$

where L is the total grating length (Figure 2.3) and κ_B is the coupling coefficient at the Bragg wavelength, λ_B.

2.3. THERMO-OPTICAL MODELING OF POLYMER FIBER BRAGG GRATING

This section presents a methodology for thermo-optical modeling of a polymer fiber Bragg grating (PFBG) associated with the intrinsic absorption of light energy. An analytical formulation based on a modified coupled-mode theory is described, which determines the power variation induced by the coupling between counter-directional light waves within the PFBG. An analytical description for absorption-induced heat generation is provided and a semi-numerical thermo-optical model, using the modified Transfer Matrix Method (TMM), along with a simple analytical thermo-optical model, is also given.

2.3.1. Heat Generation by Intrinsic Absorption

2.3.1.1. Power Spectra of Light Sources The radial variation of the power irradiance, $I(r)$, of a light source, can be assumed to follow a Gaussian profile, and to be given by [22]

$$I(r) = \frac{2P_{inc}}{\pi w^2} e^{-2r^2/w^2}, \tag{2.10}$$

where P_{inc} is the incident optical power and w is the beam radius. The beam radius, w, can be defined as [25]

$$w = \sqrt{\frac{-2r_{co}^2}{\ln(1-\Gamma)}}, \tag{2.11}$$

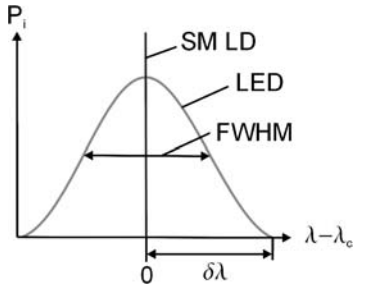


FIGURE 2.7. Power spectra of Light Emitting Diode and Single Mode Laser Diode.

where r_{co} is the radius of the core of the fiber and Γ is the power confinement factor; which determines the fraction of the incident optical power which actually enters the core of the fiber. In this study, the power confinement factor, Γ, is defined as [25]

$$\Gamma = \frac{P_{co}}{P_{tot}}, \tag{2.12}$$

where P_{co} is the power in the core, and P_{tot} is the total power illuminating the fiber. The confinement factor [25] is governed by various optical parameters. Using a Bessel function formulation, the confinement factor is found to be [25]

$$\Gamma = b\left[1 - \frac{J_0^2(V\sqrt{1-b})}{J_1(V\sqrt{1-b})J_{-1}(V\sqrt{1-b})}\right], \tag{2.13}$$

where V and b are the previously defined generalized frequency [see Equation (2.4)] and generalized guide index [see Equation (2.5)] of the fiber.

Engineering data [26,27] for typical light sources of interest suggests that the power spectra of both Single Mode Laser Diodes (SM LD's) and Light Emitting Diodes (LED's) often follow a Gaussian distribution. The normalized spectral power density, $\overline{P(\lambda)}$, for these light sources can, thus, be expressed as

$$\begin{aligned} \overline{P(\lambda)} &= Be^{4\ln 0.5((\lambda-\lambda_c)/FWHM)^2}, \\ B &= \frac{1}{2\int_0^{\delta\lambda} e^{4\ln 0.5((\lambda-\lambda_c)/FWHM)^2} d(\lambda)}, \end{aligned} \tag{2.14}$$

where $\delta\lambda$ is half of the total spectral bandwidth and FWHM is the "full width half maximum," which defines the spectral bandwidth at half of the maximum power. The central wavelength of the light source is given by λ_c. Figure 2.7 illustrates the power spectra of a typical LED and a SM LD.

2.3.1.2. Power Variation Along a PFBG The optical characteristics of the BG, including the reflection spectrum and the reflectivity, can be explained well by the coupled-mode theory [28–30]. Wave propagation in the BG can be described by the scalar wave equation [28] as

$$\frac{\partial^2 E}{\partial z^2} + k_c^2 E = 0, \tag{2.15}$$

where E is the complex amplitude of the electrical field and k_c is the wave number in a lossy medium [28], as follows

$$k_c^2 = \beta^2 - i \cdot \hat{a}\beta + 4\kappa\beta \cos(2\beta_B z), \tag{2.16}$$

where the propagation constant β [28] is defined as

$$\beta = \frac{2\pi n_{eff}}{\lambda}. \tag{2.17}$$

In Equation (2.16), β_B is the propagation constant at the Bragg condition, $\hat{a}$ is the intrinsic absorption coefficient, and κ is the coupling coefficient. It is to be noted that the absorption coefficient, $\hat{a}$ (m^{-1}), used in the present analysis is related to the more common dB-based absorption coefficient, a (dB/cm) by

$$\hat{a} = 10 \cdot a \ln(10). \tag{2.18}$$

For example, an absorption coefficient of 11.513 m^{-1} is equal to 0.5 dB/cm, and is represents about 10% of optical power loss along 1 cm of propagation.

Following the approach proposed by Kogelnik [28], the electrical field in a FBG can be written as

$$E(z) = R(z)e^{-i\beta_B z} + S(z)e^{i\beta_B z}, \tag{2.19}$$

where $R(z)$ and $S(z)$ are the amplitudes of the forward traveling and backward traveling waves, respectively. Substituting Equations (2.16), (2.17) and (2.19) into (2.15), we obtain

$$\begin{aligned}\frac{\partial^2 E}{\partial z^2} + k_c^2 E = {} & e^{-i\beta_B z}R'' - \left(2i\beta_B e^{-i\beta_B z}\right)R' - R\left[e^{-i\beta_B z}\left(-\beta^2 + i \cdot \hat{a}\beta + \beta_B^2\right)\right. \\ & \left. - 2\kappa\beta\left(e^{i\beta_B z} + e^{-i\cdot 3\beta_B z}\right)\right] + e^{i\beta_B z}S'' + \left(2i\beta_B e^{i\beta_B z}\right)S' \\ & + S\left[e^{i\beta_B z}\left(-\beta^2 - i \cdot \hat{a}\beta + \beta_B^2\right) + 2\kappa\beta\left(e^{-i\beta_B z} + e^{i\cdot 3\beta_B z}\right)\right] = 0. \end{aligned} \tag{2.20}$$

Assuming the amplitudes (R and S) vary slowly, R'' and S'' can be neglected in Equation (2.20) [28]. Furthermore, as can be seen in Equation (2.17), β is extremely large, and thus oscillating terms with much higher frequency than the Bragg frequency, i.e., $e^{i\cdot 3\beta_B z}$ and $e^{-i\cdot 3\beta_B z}$ can be also neglected [28]. Equation (2.20) can then be rewritten as

$$\begin{aligned} & -\left(2i\beta_B e^{-i\beta_B z}\right)R' - R\left\lfloor e^{-i\beta_B z}\left(-\beta^2 + i \cdot \hat{a}\beta + \beta_B^2\right) - 2\kappa\beta e^{i\beta_B z}\right\rfloor \\ & \quad + \left(2i\beta_B e^{i\beta_B z}\right)S' + S\left[e^{i\beta_B z}\left(-\beta^2 - i \cdot \hat{a}\beta + \beta_B^2\right) + 2\kappa\beta e^{-i\beta_B z}\right] = 0. \end{aligned} \tag{2.21}$$

Comparing terms with equal exponentials in Equation (2.21), the following coupled differential equations are obtained

$$\begin{aligned} \frac{dR}{dz} &= -\left(\frac{\beta}{\beta_B}\frac{\hat{a}}{2} - \frac{\beta^2 - \beta_B^2}{i \cdot 2\beta_B}\right)R + \frac{\beta}{i\beta_B}\kappa S, \\ \frac{dS}{dz} &= \left(\frac{\beta}{\beta_B}\frac{\hat{a}}{2} - \frac{\beta^2 - \beta_B^2}{i \cdot 2\beta_B}\right)S - \frac{\beta}{i\beta_B}\kappa R. \end{aligned} \tag{2.22}$$

Assuming departures from Bragg wavelength are very small, i.e., $(\beta^2 - \beta_B^2)/(2\beta_B) \approx \beta - \beta_B$ and $\beta/\beta_B \approx 1$, Equation (2.22) can be rewritten as the following coupled differential equations, which are frequently referred to as the "*coupled-mode*" equations, [21,28]

$$\frac{dR}{dz} = -\left(\frac{\hat{a}}{2} + i\delta\right)R(z) - i\kappa S(z),$$
$$\frac{dS}{dz} = \left(\frac{\hat{a}}{2} + i\delta\right)S(z) + i\kappa R(z), \tag{2.23}$$

where δ is the detuning value, which represents the departure of the propagation constant from the Bragg resonance and it is defined as [21]

$$\delta = \beta - \beta_B = \frac{2\pi n_{eff}}{\lambda} - \frac{\pi}{\Lambda}. \tag{2.24}$$

The detuning parameter, δ, is a measure of the spectral proximity of the incident light to the Bragg wavelength. For small values of δ much of the incident light will be reflected by the BG, creating a backward moving wave and transferring a significant amount of energy into the reflected wave, while for large detuning values nearly all the incident light will pass through the waveguide with little energy exchange between the counter directional waves.

The closed form solutions of the coupled-mode equations can be written as

$$R(z) = r_1 e^{mz} + r_2 e^{-mz},$$
$$S(z) = s_1 e^{mz} + s_2 e^{-mz}, \tag{2.25}$$

where the Eigen value of the coupled-mode equations is

$$m = \sqrt{\left(\frac{\hat{a}}{2} + i\delta\right)^2 + \kappa^2}. \tag{2.26}$$

In solving these equations use is made of the boundary conditions, which set the amplitude of the forward wave at the inlet to unity and the amplitude of the backward moving wave at the outlet to zero ($R(0) = 1$ and $S(L) = 0$). A detailed mathematical procedure to obtain the four coefficients in Equation (2.25), using the above boundary conditions, is shown in Appendix A.

Using the DeMoivre's formula [31], the Eigen value can be expressed as a complex variable, i.e.,

$$m = m_1 + im_2, \tag{2.27}$$

where m_1 and m_2 are real numbers defined as

$$m_1 = \sqrt{\frac{1}{2}\left\{\sqrt{\left[\left(\frac{\hat{a}}{2}\right)^2 - \delta^2 + \kappa^2\right]^2 + (\hat{a}\delta)^2} + \left[\left(\frac{\hat{a}}{2}\right)^2 - \delta^2 + \kappa^2\right]\right\}}, \tag{2.28}$$

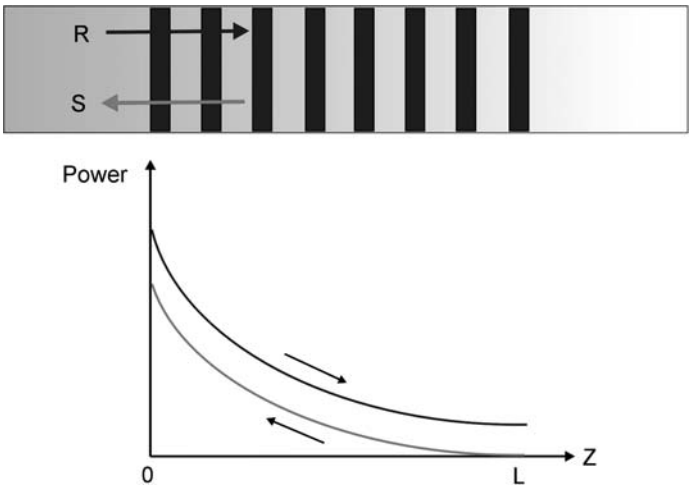


FIGURE 2.8. Axial power variations of incident and reflected light waves in Bragg grating.

$$m_2 = \sqrt{\frac{1}{2}\left\{\sqrt{\left[\left(\frac{\hat{a}}{2}\right)^2 - \delta^2 + \kappa^2\right]^2 + (\hat{a}\delta)^2} - \left[\left(\frac{\hat{a}}{2}\right)^2 - \delta^2 + \kappa^2\right]\right\}}. \tag{2.29}$$

Substituting Equation (2.27) into Equation (2.25), the amplitudes of the two waves (i.e., general solutions of the coupled-mode equations) can be rewritten as

$$R = 2r_1 e^{m_2 z}[\sinh(m_1 z)] + e^{-m_1 z}\left(r_1 e^{im_2 z} - r_1 e^{-im_2 z} + e^{-im_2 z}\right), \tag{2.30}$$

$$S = 2\left(e^{im_2 z} s_1 e^{m_1 L}\right)\sinh[m_1(z-L)] + s_1 e^{-m_1 z}\left(e^{im_2 z} - e^{im_2(2L-z)}\right)e^{2m_1 L}. \tag{2.31}$$

By multiplying the conjugates of the complex amplitudes of two waves, the axial power of the forward and backward moving waves, respectively, can be expressed as

$$|R|^2 = \frac{c_1 e^{2m_1 z} + c_2 e^{2m_1(2L-z)} + c_3\cos[2m_2(z-L)] + c_4\sin[2m_2(z-L)]}{c_1 + c_5 + c_2 e^{4m_1 L}}, \tag{2.32}$$

$$|S|^2 = \frac{c_6\{e^{2m_1 z} + e^{2m_1(2L-z)} - 2e^{m_1 L}\cos[2m_2(z-L)]\}}{c_2\kappa^2 e^{4m_1 L} + c_1\kappa^2 + c_7}. \tag{2.33}$$

The complete mathematical descriptions of the coefficients can be also found in Appendix A.

Figure 2.8 illustrates the power variations of forward and backward traveling waves propagating through a PFBG at the Bragg condition. The exponentially decaying power of the forward moving wave is mainly due to the strong coupling (or transfer) to the backward moving light wave and results in the exponentially increasing power of the reflected wave, traversing the waveguide in the opposite direction.

However, the coupled-mode equations [Equation (2.23)]—in either numerical or closed form—can not be used to determine the power variations through the BG for wavelengths that depart even modestly from the Bragg resonance. This limitation of the coupled-mode equations is mainly due to the fundamental assumptions; i.e., small departures from

Bragg resonance and negligible second derivatives of the wave amplitudes; these assumptions were used for the "coupled mode equation" solution obtained by Kogelnik [28] and the present study follows his approach.

Although the literature does not appear to contain analyses that remove either of these limitations, it is possible to obtain an approximate solution of the coupled mode equations by recognizing and exploiting physical characteristics of the forward and backward moving waves in the BGs at wavelengths that are somewhat removed from the Bragg wavelength, but still within the bandwidth of the grating. Due to the energy transfer between the two light waves, the power of both the incident and reflected waves must decrease along the axis of the BG. Therefore, the local axial gradients of the optical power of the two waves can be assumed to be negative, i.e.,

$$\frac{d(|R|^2)}{dz} < 0 \quad \text{and} \quad \frac{d(|S|^2)}{dz} < 0. \tag{2.34}$$

In addition, the fundamental assumption underpinning the coupled-mode equations, i.e., slowly varying amplitudes of the two waves, suggests that the local gradients of the optical powers vary smoothly along the grating. This implies that the second derivatives of the optical powers should be small, i.e.,

$$\frac{d^2\{(|R|)^2\}}{dz^2} \approx 0 \quad \text{and} \quad \frac{d^2\{(|S|)^2\}}{dz^2} \approx 0. \tag{2.35}$$

Modifying Equations (2.32) and (2.33) to satisfy the conditions of Equations (2.34) and (2.35) and applying the previously stated power boundary conditions at the inlet and outlet of the grating, it is possible to estimate the axial variation in the forward moving and backward moving wave as

$$|R|^2 = \frac{c_1 e^{2m_1 z} + c_2 e^{2m_1(2L-z)} + c_3\left[\dfrac{(1-\cos 2m_2 L)}{L} z + \cos 2m_2 L\right] + c_4\left[\dfrac{(\sin 2m_2 L)}{L} z - \sin 2m_2 L\right]}{c_1 + c_5 + c_2 e^{4m_1 L}}, \tag{2.36}$$

$$|S|^2 = \frac{c_6\left\{e^{2m_1 z} + e^{2m_1(2L-z)} - 2e^{2Lm_1}\left[\dfrac{(1-\cos 2m_2 L)}{L} z + \cos 2m_2 L\right]\right\}}{c_2\kappa^2 e^{4m_1 L} + c_1\kappa^2 + c_7}. \tag{2.37}$$

For light wavelengths sufficiently removed from λ_B, the BG reverts to a simple waveguide, producing no backward moving wave and no reflection. Under these circumstances, the axial power variation of the wave incident on the grating approaches that associated with Beer's law. Consequently, the power of the forward moving wave at the end of the grating can be expressed as

$$|R|^2|_{z=L} \approx e^{-\hat{a}L}. \tag{2.38}$$

By setting the value of $|R|^2$ from Equation (2.36) to that given by Equation (2.38), simple criteria for determining the bandwidth of the grating can be obtained. For wavelengths greater (and smaller) than the λ needed to match these two values, Beer's law can be em-

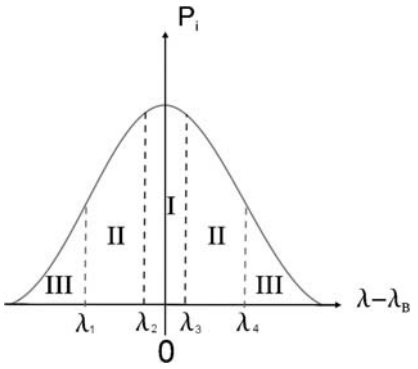


FIGURE 2.9. Light propagation zones in Bragg grating.

ployed to describe the optical power variations through the grating. In this region, the optical power of the two waves becomes

$$|R|^2 = e^{-\hat{a}z}, \tag{2.39}$$

$$|S|^2 = 0. \tag{2.40}$$

Figure 2.9 illustrates the three zones of power variations along BGs: a central zone at and close to the Bragg wavelength (I)—given by the solutions of the coupled-mode equations [Equations (2.32) and (2.33)] the grating bandwidth zone (II)—given approximately by the modified solutions of the coupled-mode equations [Equations (2.36) and (2.37)] and a third zone in which the grating behaves simply as a waveguide (III)—for which Beer's law [Equations (2.39) and (2.40)] can be used. It should be noted that the existence of these three zones in any particular application depends on the bandwidth of the light source; while a broad-band LED centered on λ_B could be expected to give rise to all three zones of behavior, a typical narrow-band SM LD may operate entirely in the domain of the coupled-mode equations (central zone I).

To illustrate the above criteria and thus to concretize the spectral behavior of a PFBG, it is instructive to consider a specific case, for which λ_B is equal to 1550 nm, and κ and $\hat{a}$ are 144.4 m^{-1} and 11.5 m^{-1}, respectively. The axial power of the forward moving wave was calculated using Equations (2.32) and (2.33) and the results are shown in Figure 2.10. It is clear from Figure 2.10 that, for values of detuning greater than 0.23×10^{-6} nm^{-1}, the axial gradient becomes positive and the power of the forward moving wave exceeds unity. Taking n_{eff} of 1.5, Equation (2.24) can be used to convert this "bounding" detuning value to a limiting wavelength of 0.06 nm, which defines the transition from Zone I to Zone II.

The grating bandwidth—obtained by allowing $|R|^2$ from Equation (2.36) to approach within 1% of the value given by Beer's law [Equation (2.38)]—can be numerically found and it lies in the range of $|\delta| < 2.3 \times 10^{-6}$ nm^{-1}. The corresponding wavelength range is $|\lambda - \lambda_B| < 0.6$ nm. Beyond this range, light propagates without interference from the BG and this range defines the outer boundary of Zone II.

2.3.1.3. Heat Generation in a PFBG The distribution of the optical power, $P(\lambda, r, z)$, can be expressed as

$$P(\lambda, r, z) = I(r)\overline{P(\lambda)}\left[|R(\lambda, z)|^2 + |S(\lambda, z)|^2\right]. \tag{2.41}$$

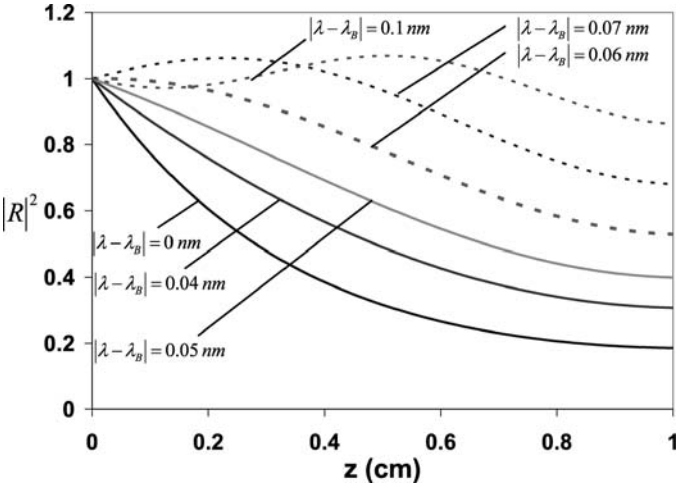


FIGURE 2.10. Axial power variation of incident light in typical polymer fiber Bragg grating.

Accordingly, heat generation at each wavelength, induced by the intrinsic absorption of light energy, can be expressed as

$$q_G(\lambda, r, z) = P(\lambda, r, z) \cdot \hat{a}(\lambda, r, z) \cong P(\lambda, r, z) \cdot \hat{a}, \tag{2.42}$$

where the absorption coefficient is assumed to be constant along the length of the grating and across the wavelength range of interest. Heat generation produced by the full spectrum of incident light can be obtained by integrating the spectral heat generation over the three zones of grating behavior, spanning the full spectral bandwidth, as

$$\begin{aligned} q_G(r, z) &= \left[\int_I q_G(\lambda, r, z)d\lambda + \int_{II} q_G(\lambda, r, z)d\lambda + \int_{III} q_G(\lambda, r, z)d\lambda\right] \\ &= \hat{a} \cdot \left\{\int_I I(r)\overline{P(\lambda)}\left[|R(\lambda, z)|^2 + |S(\lambda, z)|^2\right]d\lambda + \int_{II} I(r)\overline{P(\lambda)}\left[|R(\lambda, z)|^2\right.\right. \\ &\quad \left.\left. + |S(\lambda, z)|^2\right]d\lambda + \left[I(r)e^{-\hat{a}z}\right]_{III}\right\}. \end{aligned} \tag{2.43}$$

It is to be noted again that the modified solution [Equations (2.36) and (2.37)] of the coupled-mode equations should be used to determine $|R(\lambda, z)|^2 + |S(\lambda, z)|^2$ for zone II.

A broadband light source, such as an LED with a central wavelength of 1550 nm, typically displays 50–100 nm of FWHM [27], some 50–100 times larger than the grating bandwidth of the specific case considered in the previous section; i.e., Zone III is much greater than Zones I and II. Consequently, the contribution of the reflected (or backward moving) wave is very small and heat generation in a PFBG is primarily associated with absorption across the full bandwidth of the source, and can be expressed in a simple form as

$$q_G(r, z)|_{LED} = I(r)e^{-\hat{a}z} \cdot \hat{a}. \tag{2.44}$$

A typical SM LD has a narrow spectral bandwidth, which is almost always less than 0.3 nm [26]. Thus, for the specific case cited above, the spectrum of such a SM LD would be wholly or substantially contained within the grating bandwidth of the BG. Consequently, much of the light energy entering a PFBG in the forward moving wave will be transferred to the backward moving wave, leading to strong and spectrally complex axial variations in the propagating waves. Determination of the spatial variation of heat generation in a PFBG, under these circumstances, requires a full integration of the intensity relation [Equation (2.42)], repeated here

$$q_G(r,z)|_{SMLD} = \hat{a} \cdot \int P(\lambda, r, z) d\lambda \tag{2.45}$$

2.3.2. *Analytical Thermal Model of PFBG*

A schematic of the domain and symbols used for the analytical thermal model are shown in Figure 2.11, where r_{cl} is the radius of the entire fiber including its cladding, and r_o is the radius of the volume which contains the total incident optical power. Determination of the steady-state temperature field in the intrinsically-heated PFBG requires a solution of the heat conduction equation with non-uniform heat generation. In cylindrical coordinates, the governing equation for the temperature field can be expressed as [32]

$$\frac{\partial^2\theta}{\partial r^2} + \frac{1}{r}\frac{\partial\theta}{\partial r} + \frac{\partial^2\theta}{\partial z^2} + \frac{1}{k} q_G(r,z) = 0, \tag{2.46}$$

where θ is the excess temperature above ambient, and k is the thermal conductivity. The BG axial dimension is much larger than it core diameter, with a typical aspect ratio greater than 100:1. It is thus reasonable to assume uniform heat generation in the radial direction and reduce the above equation to a one-dimensional heat conduction equation.

Inserting the relation for the internal heat generation [Equation (2.44)] for the broadband LED light source, the one-dimensional heat conduction equation takes the form of

$$\frac{d^2\theta}{dz^2} - p^2\theta = \frac{-P_{inc}}{k\pi r_o^2} \cdot e^{-\hat{a}z} \cdot \hat{a}, \tag{2.47}$$

where $p = \sqrt{\frac{2h}{kr_o}}$.

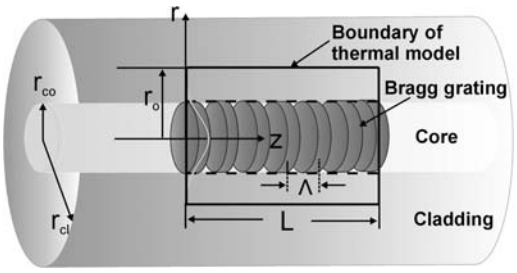


FIGURE 2.11. Geometry of polymer fiber Bragg grating thermal model.

In the above equation, h is the effective heat transfer coefficient on the outer surface of the modeled volume, reflecting the resistance to heat flow through the cladding and then into the ambient, which can be expressed as

$$h = \frac{1}{\frac{r_o}{k}\ln\frac{r_{cl}}{r_o} + \frac{r_o}{h_{cl}r_{cl}}} \cong \frac{r_{cl}}{r_o}h_{cl}, \tag{2.48}$$

where h_{cl} is the total combined (convection and radiation) heat transfer coefficient at the optical fiber surface. For the dimensions and conductivity of typical polymer optical fibers, the conductive term $(r_o/k \cdot \ln(r_{cl}/r_o))$ can be neglected relative to the convective term $(r_o/(h_{cl}r_{cl}))$ in the denominator of Equation (2.48), yielding the approximation shown. For example, when r_o, r_{cl}, and k are 7 μm, 50 μm, and 0.2 W/m K, $r_o/(h_{cl}r_{cl})$ $(= 14 \times 10^{-3})$ is about two orders of magnitude greater than $r_o/k \cdot \ln(r_{cl}/r_o)$ $(= 7 \times 10^{-5})$. For purposes of this analysis, the outer surface of the fiber was assumed to be passively cooled by natural convection and radiation, with an approximate heat transfer coefficient of 10 W/m^2 K.

The general solution of Equation (2.47) can be written as

$$\theta = g_1 e^{pz} + g_2 e^{-pz} + \theta_s. \tag{2.49}$$

The particular solution can be expected to take the form of

$$\theta_s = De^{-\hat{a}z}. \tag{2.50}$$

The procedure used to solve this equation and the derivation of the particular solution are presented in Appendix B.1. The solution is based on the assumptions that all heat loss occurs from the surface of the fiber, and that heat loss at the fiber BG ends can be neglected, i.e., assuming that both ends of the PFBG are adiabatic, i.e.,

$$\left.\frac{d\theta}{dz}\right|_{z=0} = 0 \quad \text{and} \quad \left.\frac{d\theta}{dz}\right|_{z=L} = 0. \tag{2.51}$$

The heat conduction model associated with a moderately narrow-band SM LD, for which light propagation is expected to occur primarily within zone I and II (ruled by the coupled mode equations) requires use of the more complex internal heat generation relation [Equation (2.45)], i.e.,

$$\frac{d^2\theta}{dz^2} - p^2\theta + \frac{1}{k}\cdot\frac{P_{inc}}{\pi r_o^2}\cdot\int_{\lambda_1}^{\lambda_4}\overline{P(\lambda)}\left[|R(\lambda,z)|^2 + |S(\lambda,z)|^2\right]d\lambda\cdot\hat{a} = 0. \tag{2.52}$$

Applying the convective boundary condition on the surface of the fiber, the general solution of Equation (2.52) can be expressed as

$$\theta = d_1 e^{pz} + d_2 e^{-pz} + \theta_p, \tag{2.53}$$

where θ_p is a particular solution for the above conduction equation. The particular solution follows as

$$\theta_p = \int_{\lambda_1}^{\lambda_4} F(\lambda)e^{2m_1(\lambda)z}d\lambda + \int_{\lambda_1}^{\lambda_4} G(\lambda)e^{-2m_1(\lambda)z}d\lambda + \int_{\lambda_2}^{\lambda_3} H(\lambda)\cos[M(\lambda)(z-L)]d\lambda$$

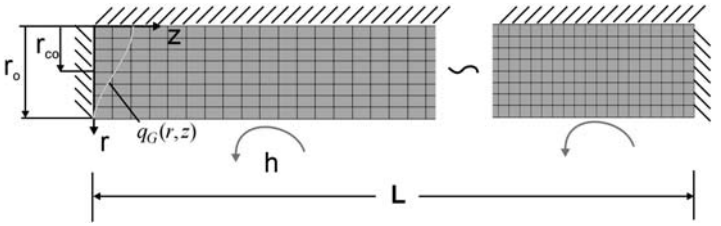


FIGURE 2.12. Geometry of finite element model for polymer fiber Bragg grating.

$$+\int_{\lambda_2}^{\lambda_3} N(\lambda)\sin[U(\lambda)(z-L)]d\lambda+\int_{\lambda_1}^{\lambda_2} V(\lambda)z+W(\lambda)d\lambda$$
$$+\int_{\lambda_3}^{\lambda_4} V(\lambda)z+W(\lambda)d\lambda+\int_{\lambda_1}^{\lambda_2} X(\lambda)z+Y(\lambda)d\lambda$$
$$+\int_{\lambda_3}^{\lambda_4} X(\lambda)z+Y(\lambda)d\lambda. \tag{2.54}$$

It should be noted that the above particular solution is for a SM LD, whose spectral window ($\lambda_1 \le \lambda \le \lambda_4$) lies within zones I and II. More specifically, $\lambda_2 \le \lambda \le \lambda_3$ defines zone I, where the closed-form solution of the power variations [Equations (2.32) and (2.33)] should be used. However, in the spectral window, $\lambda_1 \le \lambda < \lambda_2$ or $\lambda_3 < \lambda \le \lambda_4$, which lies within zone II, the modified solutions of the power variations [Equations (2.36) and (2.37)] should be employed. Appendix B.2 shows a detailed procedure to obtain the two coefficients (d_1 and d_2) for the general solution, using the above boundary conditions [see Equation (2.51)] as well as the particular solution.

2.3.3. FEA Thermal Model of PFBG

The foregoing analytical thermal model served to capture the salient features of the absorption-induced temperature profile in the polymer fiber BG, but neglected any radial temperature variations that might develop. A Finite Element thermal analysis was conducted to address this issue. Figure 2.12 shows a segment of the solution domain for the two-dimensional axi-symmetric finite-element model. The dimensions and the boundary conditions are the same as those used in the analytical model. The model had approximately 40,000 elements and the volumetric heat sources, $q_G(r, z)$, induced by light absorption in the grating, were applied to the model. In recognition of the high aspect ratio of the fiber and in order to simplify the model, the small heat losses from the ends of the PFBG were ignored, and these surfaces were considered as adiabatic. The numerical simulation was conducted with the PCG solver of ANSYS 8.0 [33].

2.3.4. Thermo-Optical Model of PFBG

Thermally-induced changes of the index and the grating period in the PFBG induce a shift in the Bragg wavelength of the grating. For an isotropic PFBG with a small uni-

form temperature change, these individual effects can be superimposed to yield the Bragg wavelength shift [34]

$$\frac{\Delta\lambda_B}{\lambda_B} = \frac{\Delta n_{eff}}{n_{eff}} + \frac{\Delta\Lambda}{\Lambda}, \tag{2.55}$$

where $\Delta n_{eff}/n_{eff}$ represents the change in the index of refraction resulting from a change in temperature and $\Delta\Lambda/\Lambda$ the contribution of thermal expansion.

By similarity to a glass optical fiber, it can be expected that—despite the small differences in the index of refraction—the core and cladding would be similar in other optical properties. Consequently, the derivative of the effective index of a PFBG with respect to temperature can be taken equal to the dn/dT of the fiber core material and the change in the effective index of refraction be expressed as

$$\Delta n_{eff} = \frac{dn}{dT}\Delta T, \tag{2.56}$$

where dn/dT is the thermo-optical coefficient, and ΔT is the temperature change of the PFBG relative to an appropriate reference temperature. Typical values for dn/dT for glassy polymers range from -100 to -200×10^{-6}/K [35].

The relative change of the grating period due to thermal expansion is given as

$$\frac{\Delta\Lambda}{\Lambda} = \alpha\Delta T. \tag{2.57}$$

For typical optical polymer materials, the thermal expansion coefficient can be expected to range from 60 to 80×10^{-6}/K [35]. It may, thus, be seen that for polymer optical fibers, the thermo-optic effect and the grating period effect on the Bragg wavelength shift are relatively large and of the same order of magnitude, but opposite in sign. Since the net Bragg wavelength shift is, thus, the difference of two relatively large numbers, care must be taken in evaluating each of the two terms appearing in Equation (2.55). Furthermore, the large magnitude of these terms, as well as the complex and generally non-uniform temperature field induced by the variations in the optical power in the grating close to the Bragg wavelengths, may necessitate use of a more exact relation for $\Delta\lambda_B/\lambda_B$ than offered by the approximate, linear superposition implicit in Equation (2.55). Succeeding paragraphs will provide more rigorous thermo-optical models for the effect of the temperature field on the optical characteristics of a PFBG.

Dealing first with the potentially large magnitude of the Bragg wavelength shift, it is appropriate to return to Equation (2.2) and observe that the shifted Bragg wavelength, λ_{B2}, can be written as

$$\lambda_{B2} = \lambda_{B1} + \Delta\lambda_B = 2(n_{eff_1} + \Delta n_{eff})(\Lambda_1 + \Delta\Lambda), \tag{2.58}$$

where subscript "1" denotes the initial state before the temperature change.

Substituting Equations (2.56) and (2.57) into (2.58), one can get

$$\lambda_{B2} = 2\left(n_{eff_1} + \frac{dn}{dT}\Delta T\right)(\Lambda_1 + \Lambda_1\alpha\Delta T). \tag{2.59}$$

Subtracting the expression for λ_{B1}, the analytical relation for the normalized Bragg wavelength shift, $\Delta\lambda_B$, can be expressed as

$$\frac{\Delta\lambda_B}{\lambda_{B1}} = \left(\frac{1}{n_{eff_1}}\frac{dn}{dT} + \alpha\right)\Delta T + \left(\frac{1}{n_{eff_1}}\frac{dn}{dT}\cdot\alpha\right)\Delta T^2, \tag{2.60}$$

where $\Delta\lambda_B = \lambda_{B2} - \lambda_{B1}$. It is to be noted that Equation (2.60) provides a more precise relation for the normalized Bragg wavelength shift than offered by Equation (2.55), adding a quadratic dependence on the temperature change to the earlier described linear term. Due to the relatively large thermal expansion coefficient, α, of the PFBG and the relatively small difference between the negative thermo-optic effect and positive grating period effect in the linear ΔT term, this quadratic term can not generally be neglected in the analysis of PFBG. Alternatively, for a glass FBG, the second term in Equation (2.60),

$$\left(\frac{1}{n_{eff_1}}\frac{dn}{dT}\cdot\alpha\right)\Delta T^2$$

can be ignored because the product of $1/n_{eff_1}\cdot dn/dT(6.7\times10^{-6})$ and $\alpha(0.55\times10^{-6})$ is six orders of magnitude smaller than their sum. For this reason, the thermo-optical dependence of the Bragg wavelength shift of a PFBG is far more complex than that of a glass FBG, which is generally found to vary linearly with temperature.

Several methods for determining the optical field in a non-uniform BG can be found in the literatures [21,36–38]. They include the Transfer Matrix Method (TMM) [21,36,37], direct numerical method [21], and Rouard Method [38]. These methods were developed originally to analyze mechanically chirped gratings, in which the grating period varies along the optical path. However, under the influence of non-uniform light absorption, and the consequent non-uniform heating, a uniform grating period may become "chirped," thus making these classical approaches suitable for the thermo-optic analysis of a PFBG. Of the choices available, the Transfer Matrix Method was chosen for implementation since it is relatively simple but appears to offer the desired accuracy.

The Transfer Matrix Method employs an analytical matrix solution to relate any two pairs of intensities (e.g., incident and reflected at the entrance) to any other two intensities (e.g., the transmitted and entering at the outlet), and is most often used to determine the reflectivity of a specified BG. As illustrated in Figure 2.13, the optical characteristics of

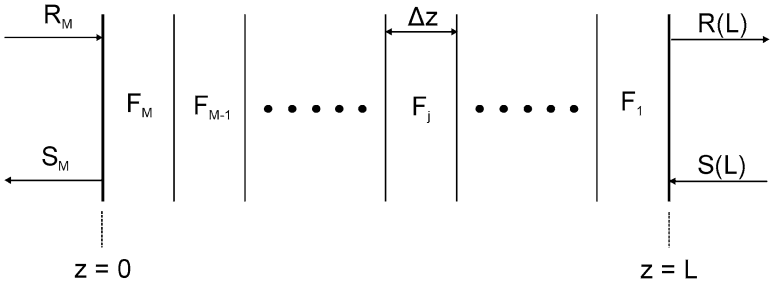


FIGURE 2.13. Illustration of Transfer Matrix Method (TMM).

each small segment, F_j, of a BG can be linearly-coupled to the next segment and can be represented in a matrix form as

$$\begin{bmatrix} R_M(\lambda) \\ S_M(\lambda) \end{bmatrix} = F(\lambda) \begin{bmatrix} R(L) \\ S(L) \end{bmatrix}, \quad F(\lambda) = F_M(\lambda) \cdot F_{M-1}(\lambda) \cdot \cdot F_j(\lambda) \cdot \cdot F_1(\lambda). \tag{2.61}$$

Each of the divided segments, $F_j(\lambda)$, with a grating length of Δz, must satisfy the coupled-mode theory, which results in the following mathematical description [36].

$$F_j(\lambda) = \begin{bmatrix} \cosh[\psi_j(\lambda)\Delta z] - i\dfrac{\hat{\sigma}_j(\lambda)}{\psi_j(\lambda)} \sinh[\psi_j(\lambda)\Delta z] & -i\dfrac{\kappa(\lambda)}{\psi_j(\lambda)} \sinh[\psi_j(\lambda)\Delta z] \\ i\dfrac{\kappa(\lambda)}{\psi_j(\lambda)} \sinh[\psi_j(\lambda)\Delta z] & \cosh[\psi_j(\lambda)\Delta z] + i\dfrac{\hat{\sigma}_j(\lambda)}{\psi_j(\lambda)} \sinh[\psi_j(\lambda)\Delta z] \end{bmatrix}, \tag{2.62}$$

where the previously defined coupling coefficient, κ, is equal to $\kappa(\lambda) = (\pi/\lambda)\delta n_{eff}$ [Equation (2.8)].

A typical index modulation for a BG, δn_{eff}, is approximately 0.01% of the refractive index of the fiber core [6]. In addition, in a bulk index grating, the materials of the alternating layers differ by little else than the small difference in index, and the thermo-optic coefficients, dn/dT, of the two layers can be expected to be nearly identical. Hence, it is reasonable to assume that the index modulation is invariant with temperature and that any possible change in the modulation is negligible compared to the thermally-driven change of the fiber core index material.

Considering the thermally-driven index shift, the "new" index of refraction for each segment of the grating, $n_{eff_{2,j}}$ can be expressed as

$$n_{eff_{2,j}} = n_{eff_1} + \frac{dn}{dT}\Delta T_j. \tag{2.63}$$

Similarly, the "new" period of the grating element caused by thermal expansion is

$$\Lambda_{2,j} = \Lambda_1(1 + \alpha\Delta T_j). \tag{2.64}$$

Using Equations (2.63) and (2.64), the "dc" coupling coefficient for a specified element, $\hat{\sigma}_j$, can be written as

$$\hat{\sigma}_j(\lambda) = \frac{2\pi n_{eff_j}}{\lambda} - \frac{\pi}{\Lambda_j} + i\frac{\hat{a}}{2} = \frac{2\pi}{\lambda}\left(n_{eff_1} + \frac{dn}{dT}\Delta T_j\right) - \frac{\pi}{\Lambda_1(1+\alpha\Delta T_j)} + i\frac{\hat{a}}{2}, \tag{2.65}$$

and the parameter, ψ_j, can be expressed as

$$\begin{aligned} \psi_j(\lambda) &= \sqrt{\kappa^2(\lambda) - \hat{\sigma}_j^2(\lambda)} \\ &= \sqrt{\left(\frac{\pi}{\lambda}\delta n_{eff}\right)^2 - \left[\frac{2\pi}{\lambda}\left(n_{eff_1} + \frac{dn}{dT}\Delta T_j\right) - \frac{\pi}{\Lambda_1(1+\alpha\Delta T_j)} + i\frac{\hat{a}}{2}\right]^2}. \end{aligned} \tag{2.66}$$

Substituting $\hat{\sigma}_j$ and ψ_j into Equation (2.62), defines all the terms of the matrix, F_j.

Using the boundary conditions ($R(0) = 1$ and $S(L) = 0$) [36], the amplitude of the forward moving wave, $R_M(\lambda)$, and backward moving wave, $S_M(\lambda)$, at the inlet of the PFBG can be obtained. The reflectivity spectrum can then be found as

$$\rho(\lambda) = \frac{|S_M(\lambda)|^2}{|R_M(\lambda)|^2}. \tag{2.67}$$

With the reflectivity determined, it is possible to obtain the reflected power spectrum, $P_{ref}(\lambda)$, as the product of the incident optical power and the reflectivity. Mathematically, it can be expressed in a simple form as

$$P_{ref}(\lambda) = P_{inc}\overline{P(\lambda)}\rho(\lambda). \tag{2.68}$$

2.4. THERMO-OPTICAL BEHAVIOR OF PMMA-BASED PFBG

To illustrate and concretize the thermo-optic behavior analyzed and described in Section 2.3, the present section will focus on the specific characteristics of a PFBG fabricated in polymethylmethacrylate (PMMA), and illuminated by a broad-band LED and a narrow-band SM LD, respectively. Subsection 2.4.1 describes the properties of the selected PMMA-based PFBG and the light sources. Subsection 2.4.2 presents the axial optical power variation, determined by the modified coupled-mode equations, through the PMMA PFBG. Subsections 2.4.3 and 2.4.4 apply the previously derived thermo-optic relations to the determination of the thermally-driven optical characteristics of the PMMA PFBG with LED and SM LD illumination. The subsequent subsection discusses the thermo-optical performance of the PFBG illuminated with other light sources.

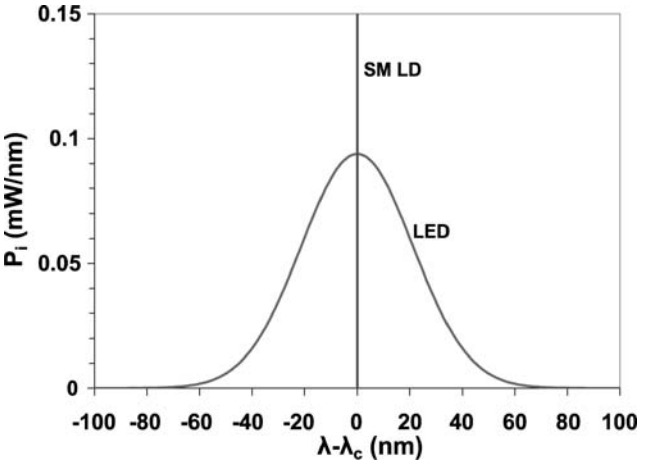
TABLE 2.1.
Properties and geometry of PMMA-based polymer fiber Bragg grating.

Parameter	Symbol	Value
Radius of core	r_{co}	3.5 um
Length	L	1 cm
Grating period	Λ	530.7 nm
Thermal conductivity	k	0.2 $\mathrm{W\,m^{-1}\,K}$
Coefficient of thermal expansion	α	$73 \times 10^{-6}\ \mathrm{K^{-1}}$
Absorption coefficient	$\hat{a}$	11.513 $\mathrm{m^{-1}}$
Thermo-optical coefficient	dn/dT	$-1.1 \times 10^{-4}\ \mathrm{K^{-1}}$
Refractive index of core	n_{co}	1.49
Refractive index of cladding	n_{cl}	1.48
Index modulation	δn_{eff}	7.244×10^{-5}
Effective refractive index	n_{eff}	1.4853
Bragg wavelength	λ_B	1576.5 nm
Coupling coefficient at Bragg wavelength	κ_B	144.36 $\mathrm{m^{-1}}$
Maximum reflectivity	$\rho_{\max}$	0.8
Generalized frequency	V	2.404
Generalized guide index	b	0.5307
Power confinement factor	Γ	0.8265
Beam radius	w	3.7402 um

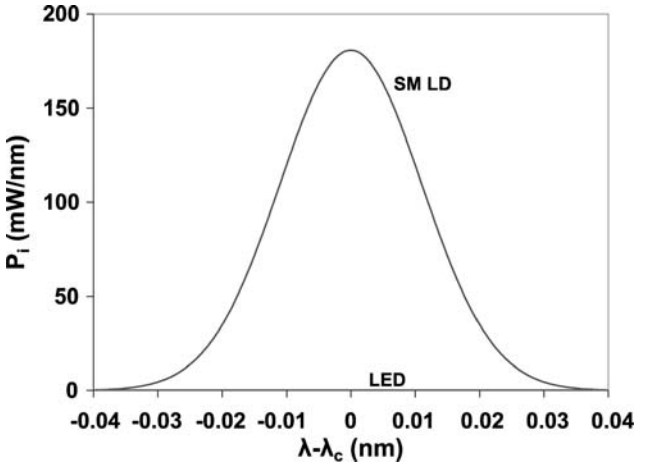
2.4.1. Description of a PMMA-Based PFBG and Light Sources

In order to characterize the thermo-optical behavior of a PFBG, a PMMA PFBG, described in the literatures [6,39,40] was chosen. The selected PFBG was fabricated by the phase-mask technique. The detailed fabrication procedure can be found in [6,39,40]. The Bragg wavelength (λ_B), the grating period (Λ), the effective index (n_{eff}), and the maximum reflectivity ($\rho_{\max}$) are 1576.5 nm, 530.7 nm, 1.4853, and 0.8, respectively. Table 2.1 shows all the inherent and derived parameters of the chosen PFBG, including the material properties and the structural, as well as optical parameters.

Typical power spectra for LED and SM LD light sources were depicted in Figure 2.14, which shows the specific power spectra of the LED and SM LD chosen for this analysis. The FWHMs for the LED and the SM LD are 50 nm and 0.026 nm, respectively, and total power is 5 mW for each of the light sources. Consequently, the peak spectral



(a)



(b)

FIGURE 2.14. Power spectra of Light Emitting Diode and Single Mode Laser Diode—5 mW illumination (a) full scale (b) expanded scale.

power density for the SM LD, approximately 180 mW/nm, is nearly 2000 times greater than the 0.1 mW/nm peak spectral power density of the LED.

2.4.2. Power Variation Along the PFBG

Equations (2.32), (2.33), (2.36), (2.37), (2.39) and (2.40) were utilized to calculate the optical power distribution within the PFBG for various detuning values. The PMMA optical parameters, appearing in Table 2.1 and including an absorption coefficient of 11.513 m^{-1}, were used for the calculation. The resulting non-dimensional power variations along the axis of the BG are plotted in Figure 2.15. At the Bragg condition, for which detuning is zero, a significant power gradient exists in both forward and backward moving waves due to the tight energy coupling between the two waves. It is to be noted that, due to the intrinsic absorption of light energy in this non-ideal fiber grating, the reflectivity, ρ (at $z = 0$), at the Bragg condition is about 7% less than the designed reflectivity of 0.8. As attention is shifted away from the Bragg wavelength, detuning increases, resulting in reduced coupling and a progressive approach to simple light wave propagation along the optical fiber. Thus, for large detuning values the incident light experiences no significant reflection and stronger propagation, controlled only by the inherent light absorption in the polymeric medium.

The limiting detuning value, or grating bandwidth (zones I and II in Figure 2.9), beyond which the BG produces no significant reflection, was numerically determined by comparing the optical powers of the forward moving waves at the outlet of the grating obtained by the modified coupled-mode solution [Equations (2.36)] with those evaluated by the simple Beer's law [Equation (2.39)]. Setting the difference in power at the outlet of the grating obtained with these two light propagation models at less than or equal to 1%, yields an absolute value of detuning, $|\delta|$, equal to or greater than 2.25×10^{-6} nm^{-1}. This results in an effective grating bandwidth of ± 0.6 nm, i.e., wavelengths beyond this bandwidth propagate freely through the grating.

In this detuned region, only the forward moving wave propagates through the PFBG and any decrease in the power is due exclusively to the intrinsic absorption of light. Con-

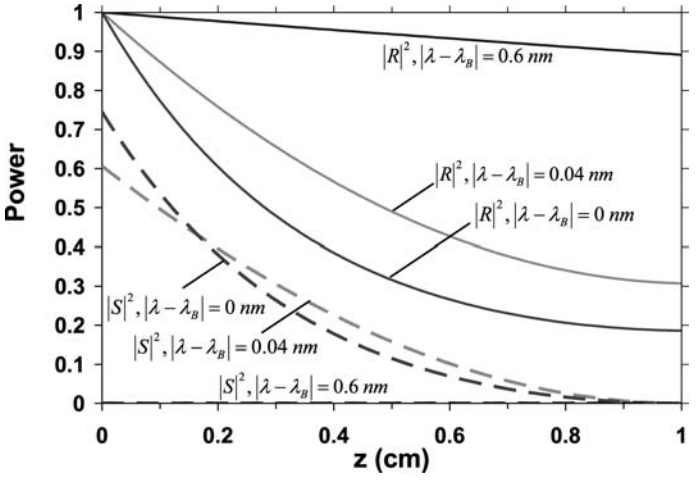


FIGURE 2.15. Incident and reflected power along PMMA fiber Bragg grating.

sequently, for wavelengths that exceed the Bragg wavelengths by $|\lambda - \lambda_B| \geq 0.6$ nm, the power variation can be found by Equation (2.39).

2.4.3. Thermo-Optical Behavior of the PFBG–LED Illumination

2.4.3.1. Thermal Analysis The spectral heat generation density, associated with the illumination of the PFBG by an LED, is shown in Figure 2.16, displaying the assumed symmetric Gaussian distribution with respect to the central wavelength and an FWHM of 50 nm. For incident optical powers of 0.5 to 5 mW, maximum spectral heat generation densities attain values of 0.7 to 7 W cm^{-3} nm^{-1}. Due to the broad-band illumination provided by the LED, extending substantially beyond the bandwidth of the PFBG, heat generation in the fiber is essentially independent of the reflected wave and can be determined almost exclusively from the inherent absorption of the propagating beam, using Equation (2.44). The resulting exponentially decaying axial profile of heat generation along the PFBG is shown in Figure 2.17 for four different incident optical powers of 0.5, 1, 3, and 5 mW. A more precise calculation of the internal heat generation, using Equation (2.43), revealed that including the reflected, backward moving light wave in the narrow spectral window of the grating bandwidth, $|\lambda - \lambda_B| \leq 0.6$ nm produced a negligible amount of additional heat in the fiber (about 0.01%).

The temperature distribution in a passively-cooled PFBG, resulting from LED illumination, was analytically determined using Equations (2.49) and (2.50) and verified by comparison to results generated by the axisymmetric FEA simulation. The effective heat transfer coefficient, h, on the external surface of the volume representing the fiber was obtained by converting the combined convection and radiation heat transfer coefficient (approximately 10 W m^{-2} K^{-1}) on the surface of the fiber, through radial conduction in the cladding and core [41]. The calculated effective heat transfer coefficient, h, for the condition described was 71 W m^{-2} K^{-1}, or 335 W/m K when applied to the length of the fiber in the axi-symmetric FEA model.

Figure 2.18 shows the PFBG axial profiles of the excess temperature (determined relative to a 25°C ambient) for the four incident optical powers. The fiber excess tempera-

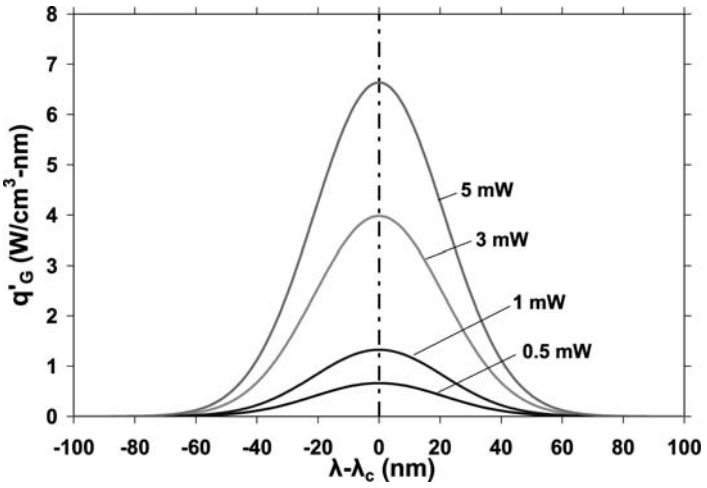


FIGURE 2.16. Spectral heat generation density in PMMA fiber Bragg grating illuminated by Light Emitting Diode.

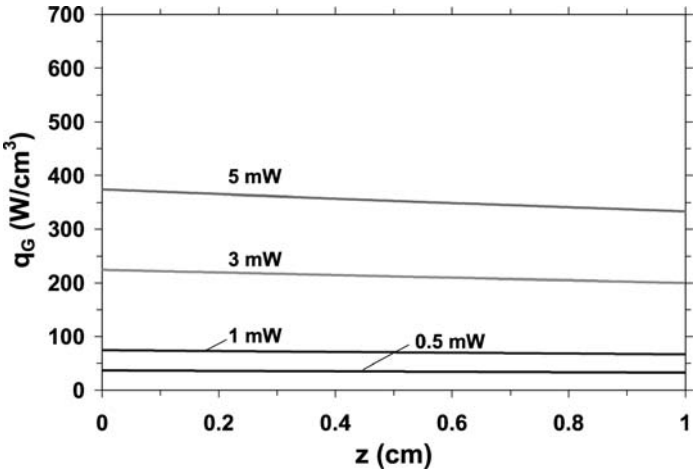


FIGURE 2.17. Heat generation density along PMMA fiber Bragg grating illuminated by Light Emitting Diode.

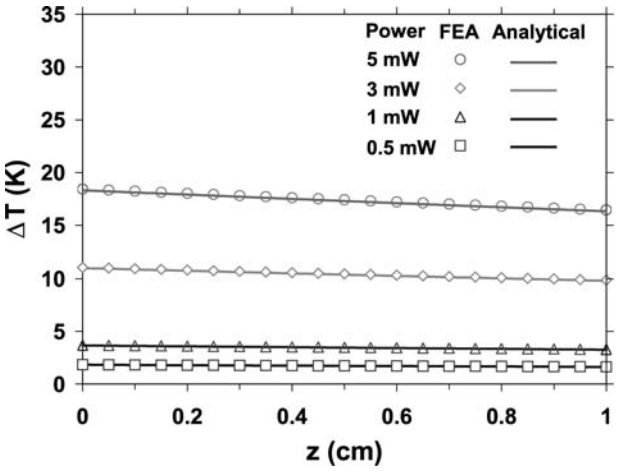


FIGURE 2.18. Analytical and numerical excess temperatures along PMMA fiber Bragg grating illuminated by Light Emitting Diode.

tures vary from 18 K at the inlet of the fiber illuminated with the highest incident power of 5 mW to just 2 K for the 0.5 mW. The analytical and FEA results display the anticipated, though barely discernable, exponential decay and appear to match each other extremely well, typically within 0.7%.

The finite-element model was further utilized to calculate the radial temperature variations in the PFBG. The results are shown in Figure 2.19, where the axial variation of the temperature difference $(T_c - T_{co})$ between the center of the PFBG core (T_c) and the core surface (T_{co}) is plotted. The peak radial temperature differences are seen to range from 0.06 K at 5 mW to 0.01 K at 0.5 mW of incident LED power and thus justify the radially-uniform temperature assumption used in the analytical temperature relations. Figure 2.19 also reveals a slight decrease in the radial temperature difference as the PFBG is traversed from the inlet to the outlet end. It should, nevertheless, be noted that, due to the 7 micron

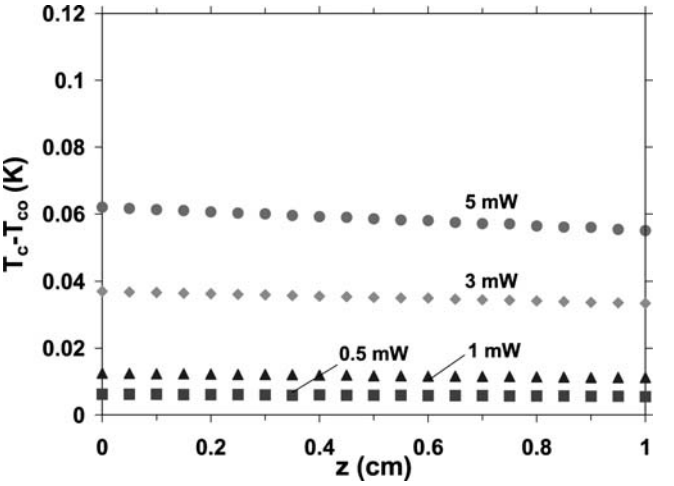


FIGURE 2.19. Radial temperature differences along PMMA fiber Bragg grating illuminated by Light Emitting Diode.

diameter of the PFBG, these radial temperature variations do yield significant temperature gradients.

2.4.3.2. Thermo-Optical Analysis

2.4.3.2.1. Reflectivity Spectrum. In order to obtain the reflectivity spectrum of the specified PFBG associated with the LED, the temperature distribution obtained from Equations (2.49) and (2.50) was integrated with the modified TMM relations [Equations (2.61) and (2.62)]. In the TMM implementation, a total of 200 segments ($\Delta z = 50$ μm) was used with wavelength bands of 0.5 pm.

Figure 2.20 displays the thermally-induced shift in the spectral reflectivity of the specified PFBG illuminated with 5 mW of LED power and operating in an ambient temperature of 25°C. The individual effects of the index change with temperature, (dn/dT), and the grating period change with temperature, $(d\Lambda/dT)$, were determined.

The results clearly indicate that the thermally-driven index change (thermo-optic) produces a negative shift in the reflectivity spectrum, relative to the incident 1576.5 nm-centered LED light, with the dominant "Bragg" wavelength moving lower by -2.03 nm. The change of the grating period due to thermal expansion results in a positive shift in the reflectivity spectrum, driving the dominant wavelength to higher values by 2.0 nm. The jagged character of the reflectivity spectrum is associated with the spectral dispersion induced by the axially non-uniform PFBG temperature and will be explored in a later section. The combined reflectivity spectrum for the 5 mW illuminated fiber shows a very small total shift in Bragg wavelength (-0.03 nm), with modest spectral dispersion. The previously noted, spectral compensation, which results from the comparable magnitude, though opposite sign of the two individual effects, is in clear evidence in this reflectivity spectrum.

The total reflectivity spectra, including both the thermal expansion and thermo-optic change with temperature, are displayed with an expanded wavelength scale in Figure 2.21, for 0.5 and 5 mW of LED illumination. The total reflectivity spectra show a consistent reflectivity shape with the constant FWHM of 0.12 nm, a maximum reflectivity of 0.74, and a small Bragg wavelength shift in the negative direction for each incident optical power.

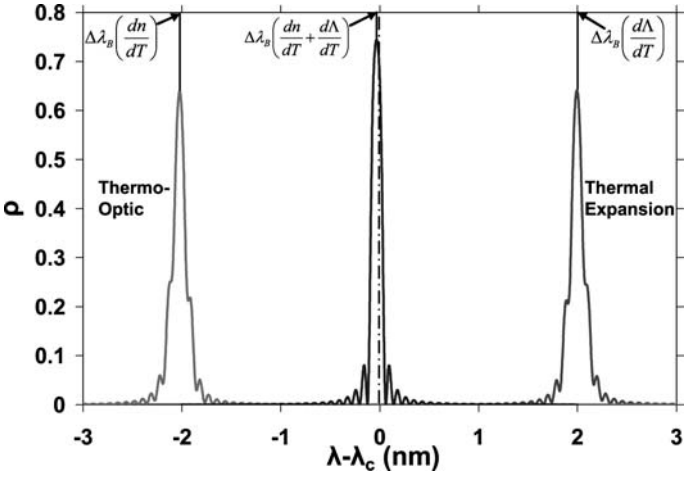


FIGURE 2.20. Thermally-induced Bragg wavelength shifts in PMMA fiber Bragg grating illuminated with a 5 mW Light Emitting Diode—thermo-optic (dn/dT), thermal expansion ($d\Lambda/dT$), and combined effects ($dn/dT + d\Lambda/dT$).

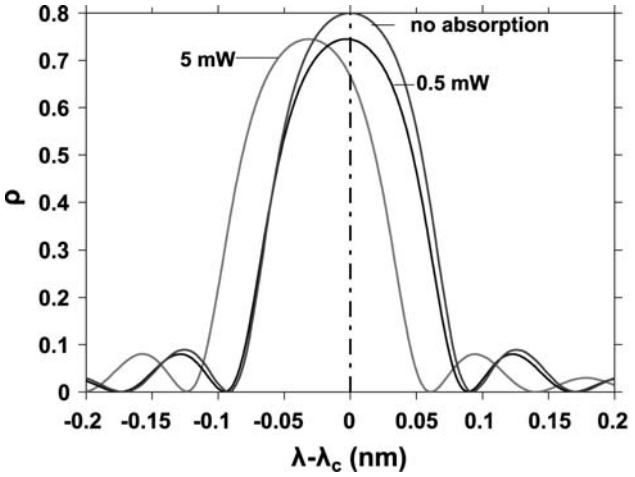


FIGURE 2.21. Reflectivity spectra for Light Emitting Diode illumination of PMMA fiber Bragg grating.

2.4.3.2.2. Reflected Power Spectrum. Although the rigorously evaluated reflectivity spectra provide useful information on the spectral dependence of the reflectivity and the Bragg wavelength shifts, due to the Gaussian spectral distribution of the incident LED illumination, they may not represent the character of the reflected power. To obtain reflected power spectra and the power-based Bragg wavelength shift, it is, thus, necessary to perform a convolution of these two distributions (i.e., Gaussian illumination and coupled-mode reflectivity).

Figure 2.22 shows the reflected power spectra for the four incident optical powers, obtained via a convolution of Equations (2.67) and (2.68), when the ambient is 25°C. Due to the wide-band LED source, the reflected power spectra present similar profiles to those of the reflectivities alone, with nearly identical negative shift in the dominant wavelength.

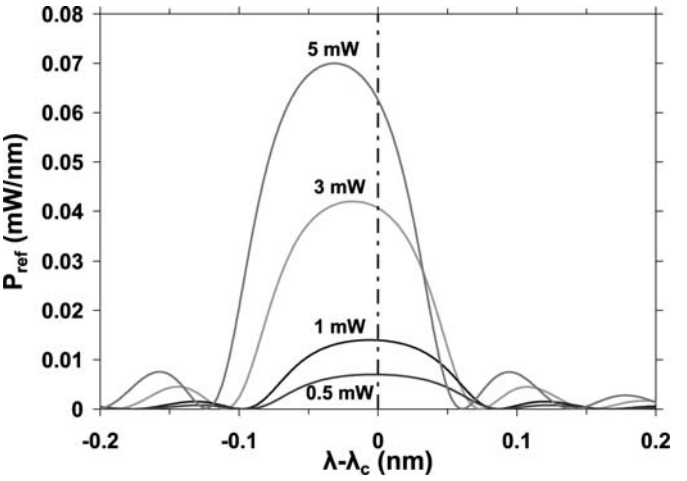


FIGURE 2.22. Reflected power spectra for Light Emitting Diode illumination of PMMA fiber Bragg grating.

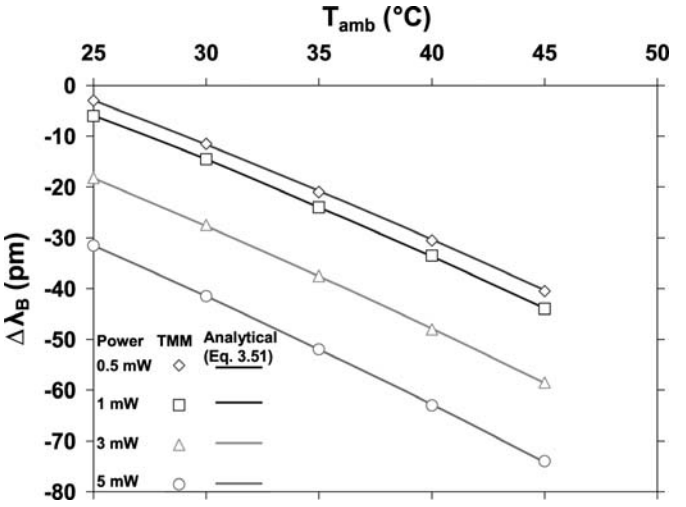


FIGURE 2.23. Bragg wavelength shifts in PMMA fiber Bragg grating illuminated with Light Emitting Diode.

2.4.3.2.3. Parametric Effects on Bragg Wavelength Shifts. The Bragg wavelength shift is, perhaps, the single most important parameter used to characterize the thermo-optical behavior of a PFBG. This Bragg wavelength shift is driven by the temperature change in the PFBG induced by both absorption and the ambient temperature change. Figure 2.23 shows the shifted Bragg wavelengths in an ambient varying from 25°C to 45°C and for four incident optical powers. The results were obtained using both the modified-TMM (the methodology used to produce the spectral reflectivity variations displayed in Figure 2.21) and Equation (2.60), with an axially-averaged PFBG temperature. The analytically determined Bragg wavelength shifts, including the quadratic term in Equation (2.60) and based on the axially-averaged temperatures, are seen to agree very well (typically to within better than 1%), with the values extracted from the more detailed modified TMM methodology.

TABLE 2.2.
Bragg wavelength shifts in PMMA-based polymer fiber Bragg grating with Light Emitting Diode illumination.

Ambient temperature, T_{amb}	Incident optical power, P_{inc} (mW)	Average excess temperature, $\overline{\Delta T}$ (K)	Transfer Matrix Method (pm)	Quadratic Eq. (2.60) (pm)	Linear Eq. (2.55) (pm)
25	0.5	1.73	−3	−2.9	−2.9
25	1	3.46	−6	−5.9	−5.8
25	3	10.38	−18.2	−18.3	−17.3
25	5	17.31	−31.5	−31.5	−28.9
30	0.5	6.73	−11.5	−11.6	−11.2
30	1	8.46	−14.5	−14.5	−14.1
30	3	15.38	−27.5	−27.7	−25.7
30	5	22.31	−41.5	−41.5	−37.3
35	0.5	11.73	−21	−20.8	−19.6
35	1	13.46	−24	−24	−22.5
35	3	20.38	−37.5	−37.6	−34.0
35	5	27.31	−52	−52	−45.6
40	0.5	16.73	−30.5	−30.3	−27.9
40	1	18.46	−33.5	−33.7	−30.8
40	3	25.38	−48	−47.9	−42.4
40	5	32.31	−63	−62.8	−53.9
45	0.5	21.73	−40.5	−40.3	−36.3
45	1	23.46	−44	−43.9	−39.2
45	3	30.38	−58.5	−58.6	−50.7
45	5	37.31	−74	−74.2	−62.3

The less accurate linear superposition formulation, given by Equation (2.55), yields a 2% to 18% discrepancy with those determined by the TMM.

The Bragg wavelength shifts calculated by all 3 methods are presented in Table 2.2. The significant effect of ambient temperatures on the shifts in Bragg wavelength is evident. For example, the effect of 15 K of ambient temperature rise is almost equivalent to that of 10 fold power increase from 0.5 mW to 5 mW.

Figure 2.24 presents the estimated shifts of Bragg wavelengths by TMM with non-uniform temperature, TMM with uniform temperature, quadratic analytical method, and linear analytical method. For these condition, the Bragg wavelength shifts are seem to be governed mainly by the average temperatures and the temperature gradients have a negligible effect.

2.4.4. Thermo-Optical Behavior of the PFBG–SM LD Illumination

2.4.4.1. Thermal Analysis Unlike the modeling approach used with the broad-band LED light source, heat generation associated with narrow-band SM LD illumination (FWHM of just 0.026 nm) strongly depends on the coupling between the forward and backward moving waves within the bandwidth of the PFBG, previously shown to extend well beyond the spectral bandwidth of the incident SM LD light. Since the Bragg wavelength shift can be expected to be of the same magnitude as the FWHM, the thermal and thermo-optical model (the modified TMM) associated with the illumination of the SM LD must be solved in a coupled way.

The spectral bandwidth of the chosen SM LD light source is 0.026 nm, which is expected to be smaller than the boundary of zone I. Consequently, the closed-form solution

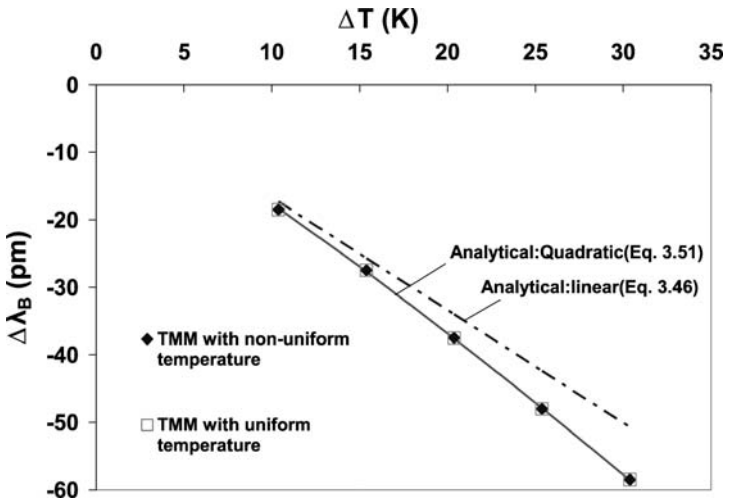


FIGURE 2.24. Transfer matrix method Bragg wavelength shifts—axially uniform and non-uniform temperature.

of the coupled-mode equation [Equations (2.32) and (2.33)] was used to obtain analytically the power variation in the grating. As in the LED analysis, a total of 200 segments ($\Delta z = 50$ μm) was used with a wavelength band of 0.5 pm for the discretization required to apply the TMM methodology.

In order to consider the simultaneous variance of the Bragg wavelength and heat generation, the fully coupled model must satisfy an energy balance $P_{inc} - P_{out} = q_G$, where P_{inc}, P_{out}, and q_G are the incident total optical power, the total optical power at the outlet, and the generated total heat, respectively. However, the optical power and the heat generation depend on the Bragg wavelength, which in turn, depends on the temperature. To deal with this additional interdependence, an iterative procedure, outlined in Figure 2.25, was established to determine the Bragg wavelength at which this energy balance could be attained.

Figure 2.26 shows the spectral heat generation density, associated with the SM LD illumination of the PFBG, for the four optical powers. Due to the narrow-band SM LD source, a much larger spectral heat generation density, than obtained with the LED, is evident at the same total illumination rates. It may also be noticed that the spectral distribution of the heat generation is slightly asymmetric about the central wavelength of the incident light, thus, differing from the perfectly symmetric distribution with the LED light source. This small, though observable shift in the spectral distribution of generated heat, can be related to the combination of the reduced coupling between the incident and reflected wave as well as the strong spectral variation of the incident light.

Figure 2.27 presents the variation in heat generation, obtained from the solution of Equation (2.45) along the axis of the PFBG, associated with the illumination of the narrow-band SM LD, for the four optical powers. It should be noted that the heat generation density of the SM LD at the inlet, peaking at 640 W/cm^3 for 5 mW, is approximately 70% greater than that of the LED. As a result, the axial heat generation gradient of the SM LD becomes much larger.

Since all of the SM LD power, with a FWHM of 0.026 nm, is contained well within the bandwidth of the PFBG, nearly the entire spectrum of the incident wave experiences Bragg reflections from the grating, drastically reducing the amount of light propagating to

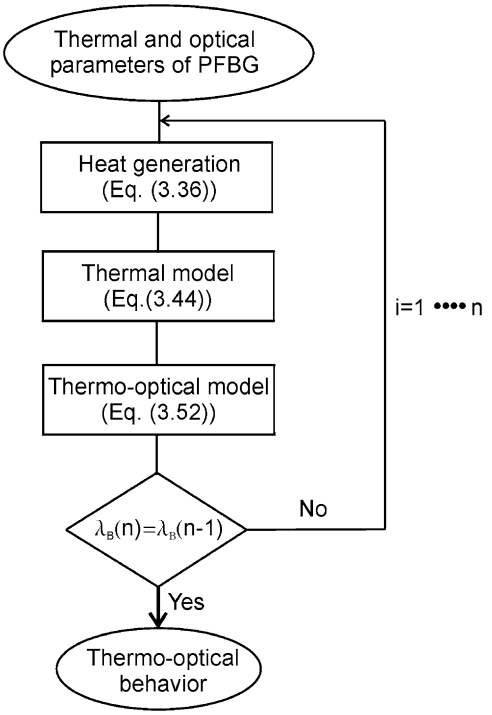


FIGURE 2.25. Iterative solution procedure—fully coupled thermo-optical analysis.

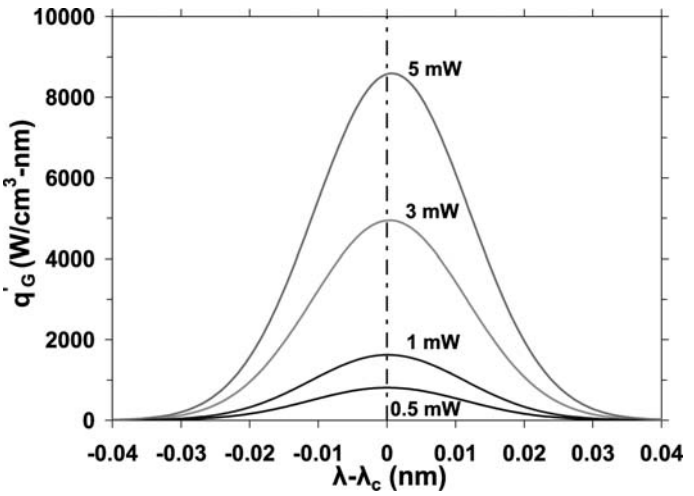


FIGURE 2.26. Spectral heat generation density in PMMA fiber Bragg grating illuminated by Single Mode Laser Diodes.

the outlet end of the grating. Using these internal heat generation profiles, it is now possible to determine the temperature variations within the PMMA FBG using Equations (2.53) and (2.54) as well as the previously described finite-element model. As mentioned earlier, the closed-form solution of the coupled-mode equation is valid in the entire spectral window

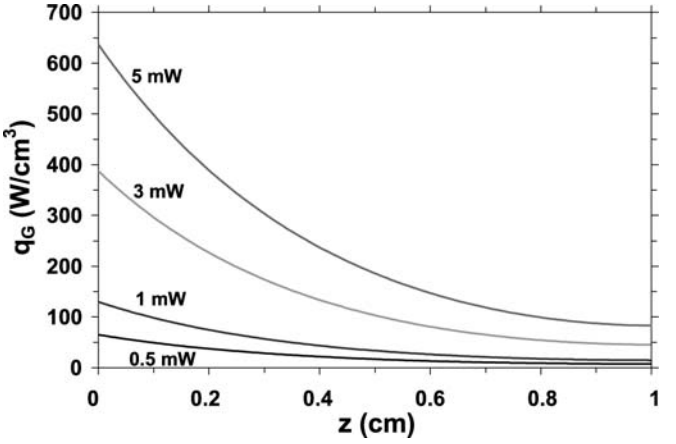


FIGURE 2.27. Heat generation density along PMMA fiber Bragg grating illuminated by Single Mode Laser Diode.

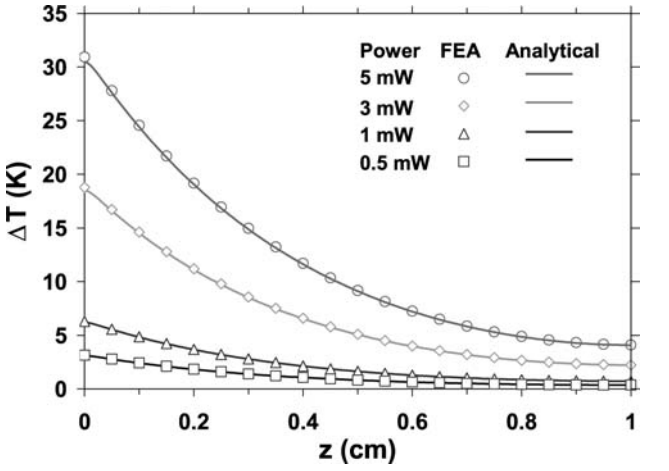


FIGURE 2.28. Analytical and numerical excess temperatures along PMMA fiber Bragg grating illuminated by Single Mode Laser Diode.

for the chosen SM LD. Hence, the integral term associated only with $\lambda_2 \leq \lambda \leq \lambda_3$ (zone I) was evaluated in the temperature solutions of Equations (2.53) and (2.54).

Figure 2.28 shows the PFBG axial profiles of the excess temperature (determined relative to a 25°C ambient) for the four incident optical powers. The results reveal that the inlet of the SM LD illuminated PFBG experiences significant heating, reaching an excess temperature of 31 K at 5 mW and 3 K even at just 0.5 mW of incident power, significantly higher than the peak temperatures induced by the previously described LED illumination. As expected from the highly non-uniform heat generation rate for the SM LD, displayed in Figure 2.27, a large axial temperature gradient, varying from nearly 27 K/cm at 5 mW to just 3 K/cm at 0.5 mW of incident power, is generated in the grating.

The results obtained with the finite-element simulation of the SM LD illuminated PFBG are also shown in Figure 2.28. They agree very well with the analytical results,

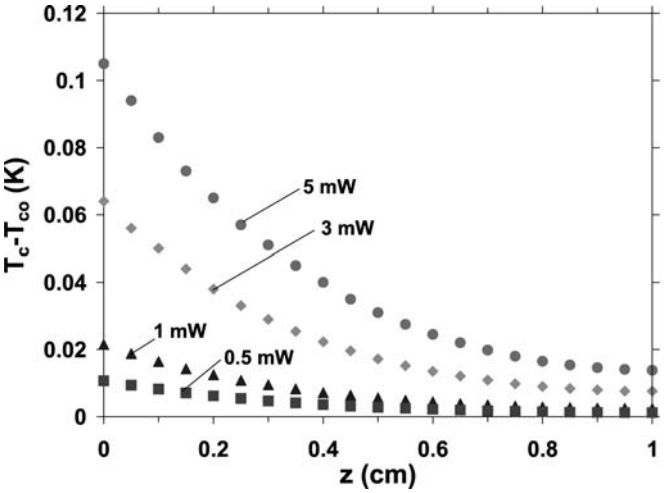


FIGURE 2.29. Radial temperature differences along PMMA fiber Bragg grating illuminated by Single Mode Laser Diode.

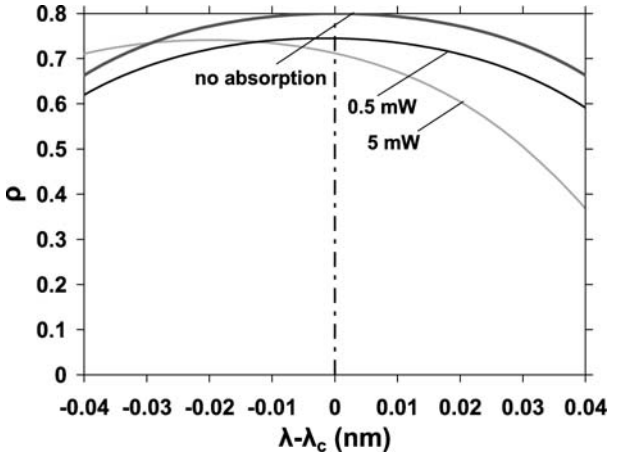
typically to within 1%. Further reinforcing the importance of the strong optical coupling between the incident and reflected waves, which occurs with the narrow-band light source, the average temperature rise of the specified PMMA PFBG induced by SM LD illumination at 5 mW and 0.5 mW is approximately 30% lower than the comparable values for the LED illumination.

Figure 2.29 shows the radial temperature difference between the center of the PFBG core (T_c) and the core surface temperature (T_{co}) along the PFBG. The maximum radial temperature difference lies below 0.11 K at the inlet of the grating and decreases axially to below 0.02 K for the highest light intensity of 5 mW, with similar profiles and even lower radial temperature differences for the lower illumination powers. Thus, the FEA results validate the one-dimensional heat conduction assumption used in the analytical thermal model for the narrow-band light source.

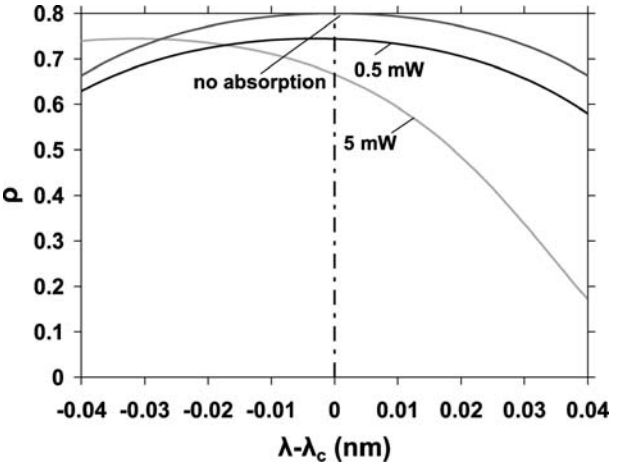
2.4.4.2. Thermo-Optical Analysis

2.4.4.2.1. Reflectivity Spectrum. Figure 2.30(a) shows the total reflectivity spectra for the maximum and minimum incident SM LD powers considered, as well as the no absorption case, all calculated with Equation (2.61), which include both the effects of the thermally-driven index shift and the grating period change due to thermal expansion. It is worth noting that the reflectivity spectrum for the SM LD does not display the multiple side lobes, which were present in the reflectivity spectrum of the LED. However, when compared to the "central" region around the Bragg wavelength for both the LED and SM LD illumination, shown in a magnified view in Figure 2.30(b), it becomes apparent that this difference can be related directly to the narrow-band of the SMLD light source, which contains all the incident light well within the grating bandwidth of the PMMA based PFBG. From an examination of Figure 2.30 it is also possible to deduce the Bragg wavelength shift based on the wavelength of the peak reflectivity of the incident light for the two limiting SM LD illuminations.

Table 2.3 provides additional Bragg wavelength shift values, extracted from reflectivity spectra (TMM), and compares these to the values predicted from Equations (2.60) and (2.55). The Bragg wavelength shifts estimated by Equation (2.60), using the quadratic



(a)



(b)

FIGURE 2.30. Narrow-band reflectivity spectra for PMMA fiber Bragg grating—(a) Single Mode Laser Diode illumination, (b) Light Emitting Diode illumination.

relation, shows generally good agreement with those evaluated by the TMM method, falling typically within 3% of the TMM values. On the contrary, estimates using Equation (2.55), based on a simple linear superposition of the thermo-optic and strain terms, generally show a 3 to 10% discrepancy with the TMM results.

It is to be noted that at the same incident power levels, the SM LD results in nearly 30% smaller Bragg wavelength shifts compared to those produced by the LED light source. This reduction in $\Delta\lambda_B$ can be explained by the approximately 30% lower average temperature rise for the PMMA grating illuminated by the SM LD.

2.4.4.2.2. Reflected Power Spectrum. The reflectivity of a BG is defined as the ratio of the reflected optical power to the incident optical power at the inlet of the PFBG (at $z = 0$). To obtain this reflectivity, it is again necessary to perform a convolution of the incident spectrum with the calculated spectral reflectivity, which yields the reflected power spectra

TABLE 2.3.
Bragg wavelength shifts in PMMA-based polymer fiber Bragg grating with Single Mode Laser Diode illumination.

Ambient temperature, T_{amb}	Incident optical power, P_{inc} (mW)	Average excess temperature, $\overline{\Delta T}$ (K)	Transfer Matrix Method (pm)	Quadratic Eq. (2.60) (pm)	Linear Eq. (2.55) (pm)
25	0.5	1.12	−1.5	−1.9	−1.9
25	1	2.24	−3.5	−3.8	−3.7
25	3	6.83	−11	−11.8	−11.4
25	5	11.85	−20	−21	−19.8
30	0.5	6.14	−10.5	−10.6	−10.3
30	1	7.29	−12.5	−12.6	−12.2
30	3	12.15	−21	−21.6	−20.3
30	5	17.98	−32	−32.8	−30.0
35	0.5	11.18	−19.5	−19.7	−18.7
35	1	12.39	−21.5	−22	−20.7
35	3	17.79	−32	−32.4	−29.7
40	0.5	16.27	−29.5	−29.4	−27.2
40	1	17.6	−32	−32	−29.4
45	0.5	21.4	−39.5	−39.6	−35.7

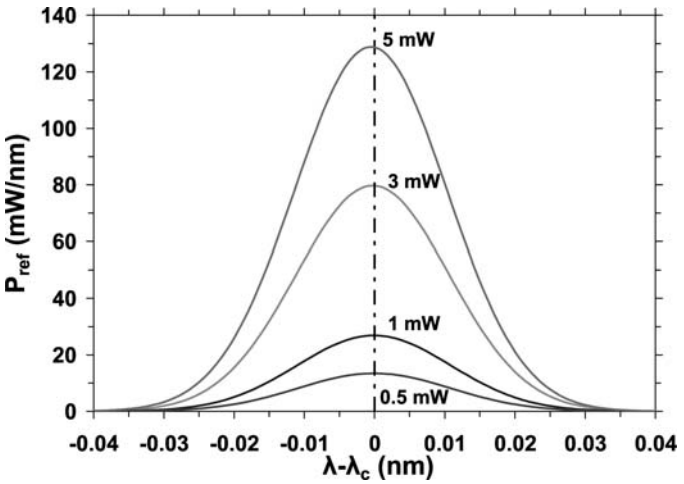


FIGURE 2.31. Reflected power spectra for Single Mode Laser Diode illumination of PMMA fiber Bragg grating.

and then integrate this over the bandwidth of the light source to obtain the total reflected power and grating reflectivity.

This procedure was followed using Equations (2.67) and (2.68). The results are displayed in Figure 2.31, showing the reflected power spectra associated with the four optical powers. The shifts of wavelengths at maximum reflected powers (*apparent Bragg wavelength shift*) with the SM LD are much smaller than those based on the peak reflectivity values (*true Bragg wavelength shift*) and summarized in Table 2.4. Furthermore, comparison of the "power" $\Delta\lambda_B$ values for the SM LD and the LED, reveals the shifts with the narrow-band light source to be far smaller than with the broad-band LED light source.

TABLE 2.4.
True and apparent Bragg wavelength shifts in PMMA-based polymer fiber Bragg grating illuminated with Single Mode Laser Diode.

Ambient temperature, T_{amb}	Incident optical power, P_{inc} (mW)	Average excess temperature, $\overline{\Delta T}$ (K)	True Bragg wavelength shift by Transfer Matrix Method (pm)	Apparent Bragg wavelength shift by Transfer Matrix Method (pm)
25	0.5	1.12	−1.5	0
25	1	2.24	−3.5	0
25	3	6.83	−11	−0.5
25	5	11.85	−20	−0.5
30	0.5	6.14	−10.5	0
30	1	7.29	−12.5	−0.5
30	3	12.15	−21	−0.5
30	5	17.98	−32	−0.5
35	0.5	11.18	−19.5	−0.5
35	1	12.39	−21.5	−0.5
35	3	17.79	−32	−1
40	0.5	16.27	−29.5	−1
40	1	17.6	−32	−1
45	0.5	21.4	−39.5	−1.5

For example, the wavelength shifts at the maximum reflected power with 5 mW of LED and SM LD illumination are 31.5 pm and 0.5 pm, respectively. This much reduced Bragg wavelength shift with the SM LD is mainly due to the strong spectral dependence of the incident power. For wavelength shifts outside the FWHM band of the incident light, the incident power sharply decreases so that the maximum reflected power can only occur near the center of the grating bandwidth.

2.4.4.2.3. Parametric Effects on $\Delta\lambda_B$. Figure 2.32 demonstrates the difference between the true Bragg wavelength, λ_B (the wavelength at the maximum reflectivity), and the apparent Bragg wavelength, λ'_B (the wavelength at the maximum reflected power measured by an optical spectrum analyzer), for an incident power of 3 mW and an ambient temperature of 35°C. As shown in Figure 2.32, the maximum reflected power occurs at $\lambda - \lambda_c = -2$ pm while the maximum reflectivity is found at $\lambda - \lambda_c = -33$ pm, resulting in a considerable difference between the true and apparent Bragg wavelength shift for these specific conditions. Table 2.4 provides further support for these conclusions, showing this variance between the apparent and true Bragg wavelength for a variety of SM LD operating conditions.

Figure 2.33 shows the Bragg wavelengths shifts with ambient temperature for the four optical powers, revealing the significant effect of ambient temperature rise on the Bragg wavelength shift. The effect is even more significant than in the case of the LED light source. An ambient temperature rise of approximately 10 K is seen to produce a Bragg wavelength shift almost equivalent to that resulting from a factor of 10 increase in the incident optical power. A 15 K ambient temperature rise would be needed to produce the same effect in a PFBG subjected to LED illumination.

Despite the relative simplicity of the analytical equation, Equation (2.60) predicts the wavelength shift faithfully; the maximum discrepancy with the shift predicted by the modified TMM is less than 1 pm. The variance between the apparent and true Bragg wavelengths induced by a change in ambient temperature is presented in Figure 2.34. It clearly

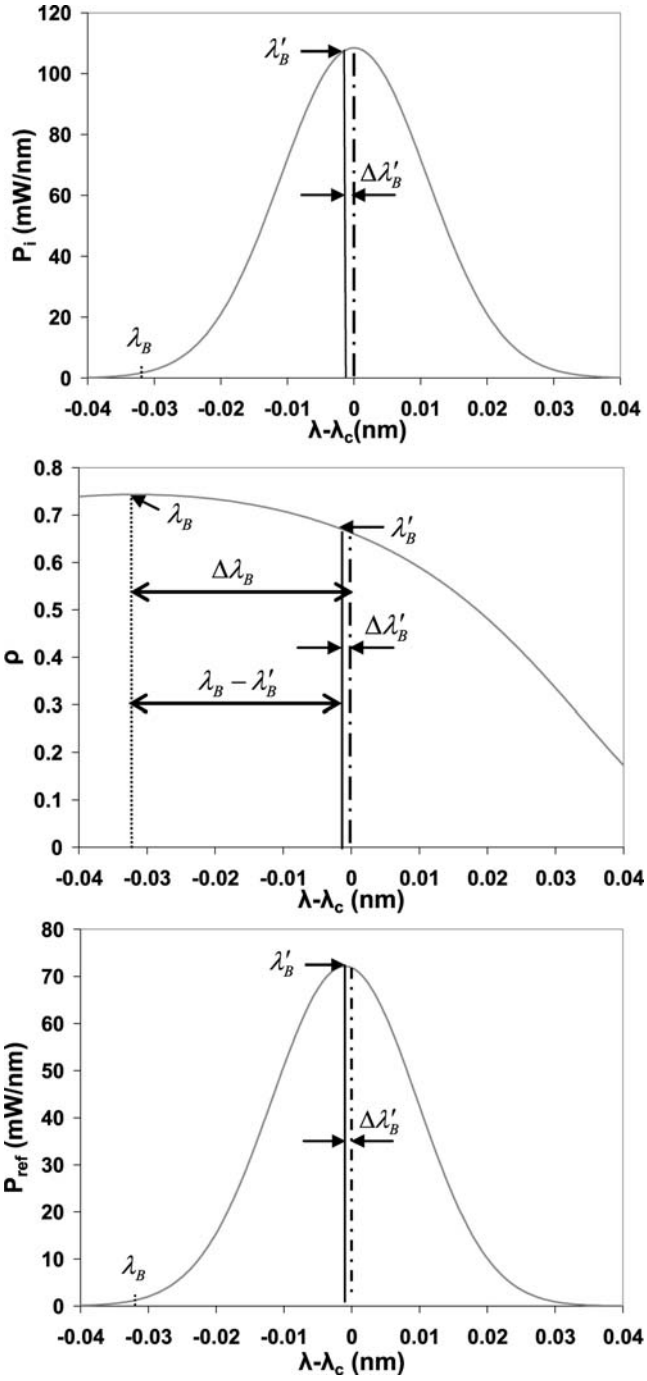


FIGURE 2.32. Apparent and true Bragg wavelength shifts in PMMA fiber Bragg grating illuminated with 3 mW Single Mode Laser Diode.

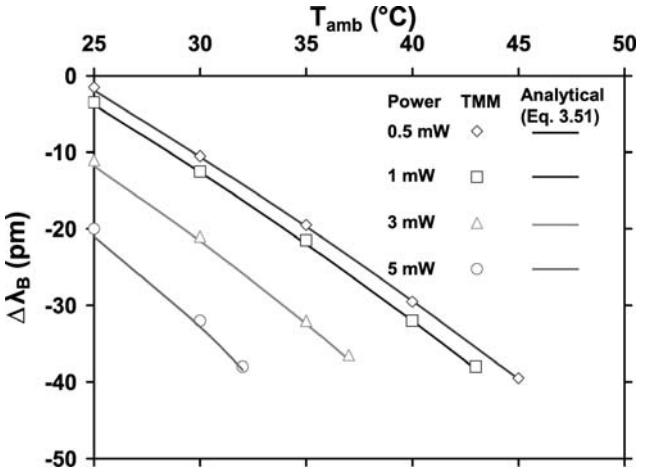


FIGURE 2.33. Bragg wavelength shifts in PMMA fiber Bragg grating illuminated with Single Mode Laser Diode.

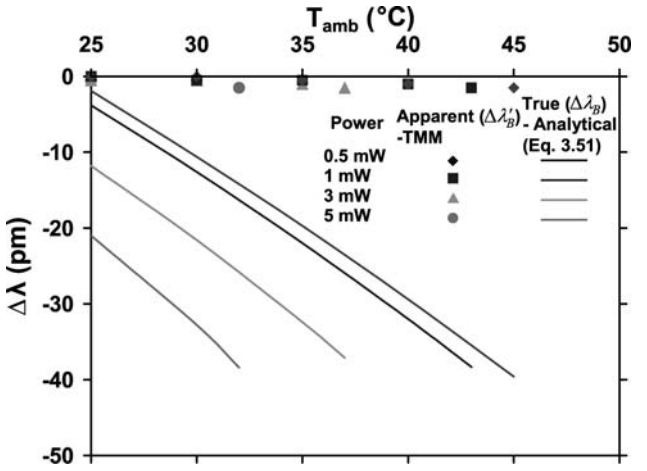


FIGURE 2.34. True and apparent Bragg wavelength shifts induced by ambient temperature and Single Mode Laser Diode illumination.

demonstrates that the considerable discrepancy between the apparent and true Bragg wavelength exists when the SM LD is used. This was caused by the sharp spectral dependence of the power density, as previously discussed.

2.4.5. Thermo-Optical Behavior of the PFBG Associated with Other Light Sources

In the preceding sections, the thermo-optical behavior of the PFBG, induced by illumination from typical SM LD's and LED's, respectively, was characterized. In solving the coupled-mode equations analytically, the entire spectrum of the chosen SM LD was found to lie well within the approximately determined bandwidth of the grating, using a range of engineering criteria for the effective bandwidth of the BG. Alternatively, heating of the PFBG by the broad spectrum of the LED, which considerably exceeded the similarly-

defined grating bandwidth, was found to follow the relatively simple Beer's Law condition. However, it must be recognized that these two illumination sources represent the "narrow-band" and "broad-band" limits, respectively, and that not all light sources result in closed form solutions of the thermo-optic relations.

Furthermore, more severe limitations are imposed by the particular mathematical solution of the simplified first order coupled-differential equations, which assumed a slowly varying amplitude envelope of the incident and reflected waves and the occurrence of energy coupling, between the two counter-propagating waves, in near the Bragg resonance [28]. It is to be noted that relaxation of these assumptions in the coupled-mode theory produces, two-dimensional, second order, coupled-partial differential equations, whose solution would require more advanced numerical methods than used in the foregoing.

2.5. CONCLUDING REMARKS

The foregoing described the development of a thermo-optic modeling methodology for characterizing the behavior of a PFBG, in which intrinsic heating—caused by light absorption in the fiber—is significant. Detailed numerical simulations of both the optical and temperature fields, including the impact of axial temperature gradients, were presented to yield the axial and radial spectral power variation in the forward and backward moving waves, from which the reflectivity and the Bragg wavelength shift can be determined. Analytical approximations of the temperature field and the resulting Bragg wavelength shift, produced by light absorption, as well as ambient temperature variations, were derived and compared to the results of the detailed numerical simulations.

This methodology was applied to a PMMA FBG illuminated alternately with a broad-band Light Emitting Diode (LED) and narrow-band Single Mode Laser Diode (SMLD) light source in the 0.5 mW to 5 mW range. The resulting high volumetric heating rate was seen to produce significant temperature rise and strong temperature gradients, especially with the SM LD illumination, in the grating. The induced temperature field resulted in a increase in the length but to a nearly comparable compensating decrease in the effective index of refraction of the PMMA grating, greatly reducing the Bragg wavelength shift and yielding nearly athermal grating behavior.

Detailed evaluation of the numerical results, obtained with the TMM, revealed that in the PMMA grating, the "chirp," produced by the absorption induced temperature gradients, led to some additional features in the reflectivity spectrum, but did not materially affect the Bragg wavelength shift ($\Delta\lambda_B$) of the PFBG. Consequently, an analytical relation—based on the use of both linear and quadratic excess average temperature terms—was seen to yield $\Delta\lambda_B$ values that are nearly indistinguishable from the more rigorous numerical solution. The TMM simulation has also identified a Bragg shift anomaly for narrow-band illumination, with the reflectivity based shift far larger than power-based shift in the PMMA grating.

REFERENCES

1. L. Eldada and L. Shacklette, Advances in polymer integrated optics, IEEE Journal of Selected Topics in Quantum Electronics, 6, pp. 54–68 (2000).
2. K.O. Hill and G. Meltz, Fiber Bragg grating technology fundamentals and overview, Journal of Lightwave Technology, 15, pp. 1263–1276 (1997).

3. K.O. Hill, B. Malo, et al., Bragg Gratings fabricated in monomode photosensitive optical fiber by UV exposure through a phase mask, Applied Physics Letters, 62, pp. 1035–1037 (1993).
4. K.O. Hill, Y. Fujii, D.C. Johnson, and B.S. Kawasaki, Photosensitivity in optical fiber waveguides, Applied Physics Letters, 32, pp. 647–649 (1978).
5. A. Othonos, Fiber Bragg Gratings-fundamentals and Applications in Telecommunications and Sensing, Artech House, Boston (1999).
6. Z. Xiong, G.D. Peng, B. Wu, and P.L. Chu, Highly tunable Bragg gratings in single-mode polymer optical fibers, IEEE Photonics Technology Letters, 11, pp. 352–354 (1999).
7. Y.-J. Rao, D.J. Webb, D.A. Jackson, L. Zhang, and I. Bennion, In-Fiber Bragg-grating temperature sensor system for medical applications, Journal of Lightwave Technology, 15, pp. 779–785 (1997).
8. E.D.J. Smith, B.A. Patterson, R.J. Webster, P.A. Krug, S.K. Jones, and D.D. Sampson, Engineering a portable quasi-distributed fibre-Bragg-grating temperature sensing system for clinical hyperthermia, Optical Fiber Sensors Conference Technical Digest, 2002, OFS 2002, 15th, 6–10 May 2002, pp. 269–272.
9. B.A. Patterson, D.D. Sampson, et al., In vivo quasi-distributed temperature sensing with fibre Bragg gratings, Lasers and Electro-Optics, 2001, CLEO '01, Technical Digest. Summaries of papers presented at the conference on, 6–11 May 2001, pp. 402–403.
10. A.V. Koulaxouzidis, M.J. Holmes, C.V. Roberts, and V.A. Handerek, A shear and vertical stress sensor for physiological measurements using fibre Bragg gratings, Engineering in Medicine and Biology Society, 2000, Proceedings of the 22nd Annual International Conference of the IEEE, Volume 1, 23–28 July 2000, pp. 55–58.
11. N.E. Fisher, J. Surowiec, et al., In-fibre Bragg gratings for ultrasonic medical applications, Measurement Science & Technology, 8, pp. 1050–1054 (1997).
12. T. Katchalski and E. Teitelbaum, Towards ultra-narrow bandwidth polymer-based resonant grating waveguide structures, Applied Physics Letters, 84, pp. 472–474 (2004).
13. W.C. Wang, M. Fisher, et al., Phase-shifted Bragg grating filters in polymer waveguides, IEEE Photonics Technology Letters, 15, pp. 548–550 (2003).
14. J.-W. Kang, M.-J. Kim, et al., Polymeric wavelength filters fabricated using holographic surface relief gratings on azobenzene-containing polymer films, Applied Physics Letters, 82, pp. 3823–3825 (2003).
15. H. Zou, K.W. Beeson, and L.W. Shacklette, Tunable planar polymer Bragg gratings having exceptionally low polarization sensitivity, Journal of Lightwave Technology, 21, pp. 1083–1088 (2003).
16. T. Augustsson, Proposal of a Bragg grating assisted MMIMI-coupler for tunable add-drop multiplexing, IEEE Photonics Technology Letters, 13, pp. 1011–1013 (2001).
17. A. Sato, S. Atsushi, et al., Holographic edge-illuminated polymer Bragg gratings for dense wavelength division optical filters at 1550 nm, Applied Optics, 42, pp. 778–784 (2003).
18. J.-F. Viens, C. Callender, et al., Compact wide-band polymer wavelength-division multiplexers, IEEE Photonics Technology Letters, 12, pp. 1010–1012 (2000).
19. L. Eldada, Y. Shing, et al., Integrated multichannel OADM's using polymer Bragg grating MZI's, IEEE Photonics Technology Letters, 10, pp. 1416–1418 (1998).
20. S. Tang, Y. Tang, et al., Fast electrooptic Bragg grating couplers for on-chip reconfigurable optical waveguide interconnects, IEEE Photonics Technology Letters, 16, pp. 1385–1387 (2004).
21. T. Erdogan, Fiber grating spectra, Journal of Lightwave Technology, 15, pp. 1277–1294 (1997).
22. C.-L. Chen, Elements of Optoelectronics and Fiber Optics, Irwin, Chicago, 1996.
23. T. Tamir, Guided-wave Optoelectronics, Springer-Verlag, New York, 1990.
24. D. Lee, Electromagnetic Principles of Integrated Optics, John Willey and Sons, New York, 1986.
25. P.K. Cheo, Fiber Optics and Optoelectronics, Prentice-Hall, Englewood Cliffs, NJ, 2nd ed., 1990.
26. Product specification sheets of the Mitsubishi DFB laser diodes (ML9XX12).
27. Product catalog of Osram Opto-Semiconductors LED.
28. H. Kogelnik, Coupled wave theory for thick hologram gratings, The Bell System Technical Journal, 48, pp. 2909–2947 (1969).
29. H. Kogelnik and C.V. Shank, Coupled-wave theory of distributed feedback lasers, Journal of Applied Physics, 43, pp. 2327–2335 (1972).
30. A. Yariv, Coupled-mode theory for guided-wave optics, IEEE Journal of Quantum Electronics, 9, pp. 919–933 (1973).
31. D.G. Zill, Advanced Engineering Mathematics, PWS-KENT Publishing Co., Boston, 1992.
32. M. Ozisik, Heat Conduction, John Willey and Sons, New York, 1980.
33. S. Moaveni, Finite Element Analysis-theory and Application with ANSYS, Prentice Hall, NJ, 2002.
34. R. Steenkiste and G. Springer, Strain and Temperature Measurement with Fiber Optic Sensors, Technomic Publishing Company, Lancaster, PA, 1997.

35. M. Weber, CRC Handbook of Laser Science and Technology Supplement 2: Optical Materials, CRC Press, Boca Raton, FL, 1995.
36. M. Yamada and K. Sakuda, Analysis of almost-periodic distributed slab waveguides via a fundamental matrix approach, Applied Optics, 26, pp. 3474–3478 (1987).
37. J.E. Sipe, L. Poladian, and C.M.D. Sterke, Propagation through non-uniform grating structures, Journal of the Optical Society of America A, 11, pp. 1307–1320 (1994).
38. L.A. Weller-Brophy and D.G. Hall, Analysis of waveguide gratings: application of Rouard's method, Journal of the Optical Society of America A, 2, pp. 863–871 (1985).
39. H.Y. Liu, G.D. Peng, and P.L. Chu, Thermal stability of gratings in PMMA and CYTOP polymer fibers, Optics Communications, 24, pp. 151–156 (2002).
40. G.D. Peng and P.L. Chu, Polymer optical fiber photosensitivities and highly tunable fiber gratings, Fiber and Integrated Optics, 19, pp. 277–293 (2000).
41. A. Kraus and A. Bar-Cohen, Thermal Analysis and Control of Electronic Equipment, Hemisphere Publishing Corp., New York, 1983.

APPENDIX 2.A. SOLUTION PROCEDURE TO OBTAIN THE OPTICAL POWER ALONG THE PFBG

Using $R(0) = 1$ and $S(L) = 0$, the following equation can be obtained from Equation (2.24)

$$r_1 + r_2 = 1, \tag{2A.1}$$

$$s_1 e^{mL} + s_2 e^{-mL} = 0. \tag{2A.2}$$

Substitution of the general solutions into the coupled-mode equation yields

$$\frac{dR}{dz} = \frac{d}{dz}\left(r_1 e^{mz} + r_2 e^{-mz}\right) = -\left(\frac{\hat{a}}{2} + i\delta\right)\left(r_1 e^{mz} + r_2 e^{-mz}\right) - i\kappa\left(s_1 e^{mz} + s_2 e^{-mz}\right). \tag{2A.3}$$

The above equation can be rearranged as

$$e^{mz}\left[r_1\left(m + \frac{\hat{a}}{2}\right) + i(\delta r_1 + \kappa s_1)\right] + e^{-mz}\left[r_2\left(-m + \frac{\hat{a}}{2}\right) + i(\delta r_2 + \kappa s_2)\right] = 0, \tag{2A.4}$$

which produces two additional conditions required to determine the four constants in the general solution as

$$r_1\left(m + \frac{\hat{a}}{2}\right) = -i(\delta r_1 + \kappa s_1), \tag{2A.5}$$

$$r_2\left(-m + \frac{\hat{a}}{2}\right) = -i(\delta r_2 + \kappa s_2). \tag{2A.6}$$

Solving four simultaneous equations [(2A.1), (2A.2), (2A.5), and (2A.6)], the coefficients of the general solution can be obtained as

$$r_1 = \frac{m - \left(\frac{\hat{a}}{2} + i\delta\right)}{m - \left(\frac{\hat{a}}{2} + i\delta\right) + \left[m + \left(\frac{\hat{a}}{2} + i\delta\right)\right]e^{2mL}}, \tag{2A.7}$$

$$r_2 = \frac{\left[m + \left(\frac{\hat{a}}{2} + i\delta\right)\right]e^{2mL}}{m - \left(\frac{\hat{a}}{2} + i\delta\right) + \left[m + \left(\frac{\hat{a}}{2} + i\delta\right)\right]e^{2mL}}, \tag{2A.8}$$

$$s_1 = \frac{m^2 - \left(\frac{\hat{a}}{2} + i\delta\right)^2}{-i\kappa\left\{m - \left(\frac{\hat{a}}{2} + i\delta\right) + \left[m + \left(\frac{\hat{a}}{2} + i\delta\right)\right]e^{2mL}\right\}}, \tag{2A.9}$$

$$s_2 = \frac{\left[m^2 - \left(\frac{\hat{a}}{2} + i\delta\right)^2\right]e^{2mL}}{i\kappa\left\{m - \left(\frac{\hat{a}}{2} + i\delta\right) + \left[m + \left(\frac{\hat{a}}{2} + i\delta\right)\right]e^{2mL}\right\}}. \tag{2A.10}$$

Coefficients for the solution of axial powers are

$$c_1 = \left(m_1 - \frac{\hat{a}}{2}\right)^2 + (m_2 - \delta)^2, \tag{2A.11}$$

$$c_2 = \left(m_1 + \frac{\hat{a}}{2}\right)^2 + (m_2 + \delta)^2, \tag{2A.12}$$

$$c_3 = 2e^{2m_1 L}\left[{m_1}^2 + {m_2}^2 - \left(\frac{\hat{a}}{2}\right)^2 - \delta^2\right], \tag{2A.13}$$

$$c_4 = -4e^{2m_1 L}\left(-m_1\delta + m_2\frac{\hat{a}}{2}\right), \tag{2A.14}$$

$$c_5 = 2e^{2m_1 L}\left\{\left[m_1^2 + m_2^2 - \left(\frac{\hat{a}}{2}\right)^2 - \delta^2\right]\cos(2m_2 L) + 2\left(-m_1\delta + m_2\frac{\hat{a}}{2}\right)\sin(2m_2 L)\right\}, \tag{2A.15}$$

$$c_6 = \left[{m_1}^2 - {m_2}^2 - \left(\frac{\hat{a}}{2}\right)^2 + \delta^2\right]^2 + 4\left(m_1 m_2 - \frac{\hat{a}}{2}\delta\right)^2, \tag{2A.16}$$

$$c_7 = 2\kappa^2 e^{2m_1 L}\left\{\left[{m_1}^2 + {m_2}^2 - \left(\frac{\hat{a}}{2}\right)^2 - \delta^2\right]\cos(2m_2 L) + 2\cdot\left(-m_1\delta + m_2\frac{\hat{a}}{2}\right)\sin(2m_2 L)\right\}. \tag{2A.17}$$

APPENDIX 2.B. SOLUTION PROCEDURE TO DETERMINE THE TEMPERATURE PROFILE ALONG THE PFBG

2.B.1. Solution Procedure of the Temperature Profile Along the PFBG with the LED

Substituting the proposed particular solution into the conduction equation [Equation (2.47)], the coefficient of the particular solution can be obtained as

$$D = \frac{P_{inc}\hat{a}}{(p^2 - \hat{a}^2)(\pi r_o^2 k)}. \tag{2B.1}$$

Using boundary conditions, both ends of Bragg grating are adiabatic,

$$\left.\frac{d\theta}{dz}\right|_{z=0} = pg_1 - pg_2 - \hat{a}D = 0, \tag{2B.2}$$

$$\left.\frac{d\theta}{dz}\right|_{z=L} = pg_1 e^{pL} - pg_2 e^{-pL} - \hat{a}De^{-\hat{a}L} = 0. \tag{2B.3}$$

Two coefficients of the general solution can be determined as

$$g_1 = \frac{\hat{a}D(e^{-\hat{a}L} - e^{-pL})}{p(e^{pL} - e^{-pL})}, \tag{2B.4}$$

$$g_2 = g_1 - \frac{\hat{a}}{p}D. \tag{2B.5}$$

2.B.2. Solution Procedure of the Temperature Profile Along the PFBG with the SM LD

Rearranging the non-homogeneous term in the conduction equation associated with the SM LD [Equation (2.52)] yields

$$\begin{aligned}\left[-\frac{q_G(z)}{k}\right] &= \frac{-P_{inc}}{k\pi r_o^2}\cdot\hat{a}\int_{\lambda_1}^{\lambda_4}\overline{P(\lambda)}\left[(|R(\lambda,z)|)^2 + (|S(\lambda,z)|)^2\right]d\lambda \\ &= \frac{-P_{inc}}{k\pi r_o^2}\cdot\hat{a}\int_{\lambda_1}^{\lambda_4} Be^{4\ln 0.5(\frac{\lambda-\lambda_C}{FWHM})^2}\cdot\left[a_1(\lambda)e^{2m_1(\lambda)z} + a_2(\lambda)e^{2m_1(\lambda)(2L-z)}\right]d\lambda \\ &\quad - \frac{P_{inc}}{k\pi r_o^2}\cdot\hat{a}\int_{\lambda_2}^{\lambda_3} Be^{4\ln 0.5(\frac{\lambda-\lambda_C}{FWHM})^2}\cdot(a_3(\lambda)\cos(2m_2(\lambda)(z-L)) \\ &\qquad + a_4(\lambda)\sin(2m_2(\lambda)(z-L)))d\lambda\end{aligned}$$

$$
\begin{aligned}
&- \frac{P_{inc}}{k\pi r_o^2}\cdot\hat{a}\int_{\lambda_1}^{\lambda_2} Be^{4\ln 0.5(\frac{\lambda-\lambda_c}{FWHM})^2} \\
&\qquad\times\left\{a_3(\lambda)\left[\left(\frac{1-\cos(2m_2(\lambda)L)}{L}\right)z+\cos[2m_2(\lambda)L]\right]\right. \\
&\qquad\left.+a_4(\lambda)\left(\frac{\sin(2m_2(\lambda)L)}{L}z-\sin[2m_2(\lambda)L]\right)\right\}d\lambda \\
&- \frac{P_{inc}}{k\pi r_o^2}\cdot\hat{a}\int_{\lambda_3}^{\lambda_4} Be^{4\ln 0.5(\frac{\lambda-\lambda_c}{FWHM})^2} \\
&\qquad\times\left\{a_3(\lambda)\left(\left[\frac{1-\cos(2m_2(\lambda)L)}{L}\right]z+\cos[2m_2(\lambda)L]\right)\right. \\
&\qquad\left.+a_4(\lambda)\left(\frac{\sin[2m_2(\lambda)L]}{L}z-\sin[2m_2(\lambda)L]\right)\right\}d\lambda, \quad \text{(2B.6)}
\end{aligned}
$$

$$
\begin{aligned}
a_1(\lambda) &= \frac{c_1}{c_1+c_5+c_2e^{4m_1L}}+\frac{c_6}{\kappa^2c_2e^{4m_1L}+2\kappa^2c_5+\kappa^2c_1}, \\
a_2(\lambda) &= \frac{c_2}{c_1+c_5+c_2e^{4m_1L}}+\frac{c_6}{\kappa^2c_2e^{4m_1L}+2\kappa^2c_5+\kappa^2c_1}, \\
a_3(\lambda) &= \frac{c_3}{c_1+c_5+c_2e^{4m_1L}}+\frac{c_6}{\kappa^2c_2e^{4m_1L}+2\kappa^2c_5+\kappa^2c_1}\left(-2e^{2m_1L}\right), \\
a_4(\lambda) &= \frac{c_4}{c_1+c_5+c_2e^{4m_1L}}.
\end{aligned}
$$

The particular solution can take a form as

$$
\begin{aligned}
\theta_p = &\int_{\lambda_1}^{\lambda_4} F(\lambda)e^{2m_1(\lambda)z}d\lambda+\int_{\lambda_1}^{\lambda_4} G(\lambda)e^{-2m_1(\lambda)z}d\lambda+\int_{\lambda_2}^{\lambda_3} H(\lambda)\cos[M(\lambda)(z-L)]d\lambda \\
&+\int_{\lambda_2}^{\lambda_3} N(\lambda)\sin[U(\lambda)(z-L)]d\lambda+\int_{\lambda_1}^{\lambda_2} V(\lambda)z+W(\lambda)d\lambda \\
&+\int_{\lambda_3}^{\lambda_4} V(\lambda)z+W(\lambda)d\lambda+\int_{\lambda_1}^{\lambda_2} X(\lambda)z+Y(\lambda)d\lambda \\
&+\int_{\lambda_3}^{\lambda_4} X(\lambda)z+Y(\lambda)d\lambda. \qquad \text{(2B.7)}
\end{aligned}
$$

By substituting the above particular solution [Equation (2B.7)] into the governing equation (Equation (2.52)), one can obtain the following equation.

$$
\begin{aligned}
&\int_{\lambda_1}^{\lambda_4}\left\{4[m_1(\lambda)]^2-p^2\right\}F(\lambda)e^{2m_1(\lambda)z}d\lambda+\int_{\lambda_1}^{\lambda_4}\left\{4[m_1(\lambda)]^2-p^2\right\}G(\lambda)e^{-2m_1(\lambda)z}d\lambda \\
&\quad-\int_{\lambda_2}^{\lambda_3}\left\{[M(\lambda)]^2+p^2\right\}H(\lambda)\cos\left[M(\lambda)(z-L)\right]d\lambda
\end{aligned}
$$

$$
\begin{aligned}
&- \int_{\lambda_2}^{\lambda_3} \left\{[U(\lambda)]^2 + p^2\right\} N(\lambda) \sin[U(\lambda)(z-L)] d\lambda \\
&- p^2 \int_{\lambda_1}^{\lambda_2} V(\lambda) z + W(\lambda) + X(\lambda) z + Y(\lambda) d\lambda \\
&- p^2 \int_{\lambda_3}^{\lambda_4} V(\lambda) z + W(\lambda) + X(\lambda) z + Y(\lambda) d\lambda \\
&= \frac{-P_{inc}}{k\pi r_o^2} \cdot \hat{a} \int_{\lambda_1}^{\lambda_4} B e^{4\ln 0.5(\frac{\lambda-\lambda_c}{FWHM})^2} \cdot \left[a_1(\lambda) e^{2m_1(\lambda)z} + a_2(\lambda) e^{2m_1(\lambda)(2L-z)}\right] d\lambda \\
&\quad - \frac{P_{inc}}{k\pi r_o^2} \cdot \hat{a} \int_{\lambda_2}^{\lambda_3} B e^{4\ln 0.5(\frac{\lambda-\lambda_c}{FWHM})^2} \cdot \{a_3(\lambda) \cos[2m_2(\lambda)(z-L)] \\
&\qquad\qquad + a_4(\lambda) \sin[2m_2(\lambda)(z-L)]\} d\lambda \\
&\quad - \frac{P_{inc}}{k\pi r_o^2} \cdot \hat{a} \int_{\lambda_1}^{\lambda_2} B e^{4\ln 0.5(\frac{\lambda-\lambda_c}{FWHM})^2} \\
&\qquad\qquad \times \left\{a_3(\lambda)\left(\left(\frac{1-\cos[2m_2(\lambda)L]}{L}\right) z + \cos[2m_2(\lambda)L]\right)\right. \\
&\qquad\qquad \left. + a_4(\lambda)\left(\frac{\sin[2m_2(\lambda)L]}{L} z - \sin[2m_2(\lambda)L]\right)\right\} d\lambda \\
&\quad - \frac{P_{inc}}{k\pi r_o^2} \cdot \hat{a} \int_{\lambda_3}^{\lambda_4} B e^{4\ln 0.5(\frac{\lambda-\lambda_c}{FWHM})^2} \\
&\qquad\qquad \times \left\{a_3(\lambda)\left(\left(\frac{1-\cos[2m_2(\lambda)L]}{L}\right) z + \cos[2m_2(\lambda)L]\right)\right. \\
&\qquad\qquad \left. + a_4(\lambda)\left(\frac{\sin[2m_2(\lambda)L]}{L} z - \sin[2m_2(\lambda)L]\right)\right\} d\lambda. \qquad \text{(2B.8)}
\end{aligned}
$$

The spectrally dependent coefficients of the particular solution can be found by comparing coefficients in each term of Equation (2B.8) as

$$
F(\lambda) = \frac{\dfrac{-P_{inc}}{k\pi r_o^2} \cdot \hat{a} \cdot B e^{4\ln 0.5(\frac{\lambda-\lambda_c}{FWHM})^2} \cdot a_1(\lambda)}{4[m_1(\lambda)]^2 - p^2}, \qquad \text{(2B.9)}
$$

$$
G(\lambda) = \frac{\dfrac{-P_{inc}}{k\pi r_o^2} \cdot \hat{a} \cdot B e^{4\ln 0.5(\frac{\lambda-\lambda_c}{FWHM})^2} \cdot a_2(\lambda) \cdot e^{4L \cdot m_1(\lambda)}}{4[m_1(\lambda)]^2 - p^2}, \qquad \text{(2B.10)}
$$

$$
M(\lambda) = U(\lambda) = 2m_2(\lambda), \qquad \text{(2B.11)}
$$

$$H(\lambda) = \frac{\dfrac{P_{inc}}{k\pi r_o^2} \cdot \hat{a} \cdot Be^{4\ln 0.5(\frac{\lambda-\lambda_c}{FWHM})^2} \cdot a_3(\lambda)}{4[m_2(\lambda)]^2 + p^2}, \tag{2B.12}$$

$$N(\lambda) = \frac{\dfrac{P_{inc}}{k\pi r_o^2} \cdot \hat{a} \cdot Be^{4\ln 0.5(\frac{\lambda-\lambda_c}{FWHM})^2} \cdot a_4(\lambda)}{4[m_2(\lambda)]^2 + p^2}, \tag{2B.13}$$

$$V(\lambda) = \frac{\dfrac{P_{inc}}{k\pi r_o^2} \cdot \hat{a} \cdot Be^{4\ln 0.5(\frac{\lambda-\lambda_c}{FWHM})^2} \cdot a_3(\lambda)}{p^2} \frac{[1-\cos(2m_2L)]}{L}, \tag{2B.14}$$

$$W(\lambda) = \frac{\dfrac{P_{inc}}{k\pi r_o^2} \cdot \hat{a} \cdot Be^{4\ln 0.5(\frac{\lambda-\lambda_c}{FWHM})^2} \cdot a_3(\lambda)}{p^2} \cos(2m_2L), \tag{2B.15}$$

$$X(\lambda) = \frac{\dfrac{P_{inc}}{k\pi r_o^2} \cdot \hat{a} \cdot Be^{4\ln 0.5(\frac{\lambda-\lambda_c}{FWHM})^2} \cdot a_4(\lambda)}{p^2} \frac{\sin(2m_2L)}{L}, \tag{2B.16}$$

$$Y(\lambda) = \frac{-\dfrac{P_{inc}}{k\pi r_o^2} \cdot \hat{a} \cdot Be^{4\ln 0.5(\frac{\lambda-\lambda_c}{FWHM})^2} \cdot a_4(\lambda)}{p^2} \sin(2m_2L). \tag{2B.17}$$

Assuming that all the heat loss occurs from the surface of the fiber and neglecting the axial conduction at the fiber ends, i.e., assuming that both ends of the BGs are adiabatic, the following two boundary conditions can be obtained

$$\left.\frac{d\theta}{dz}\right|_{z=0} = 0 \quad \text{and} \quad \left.\frac{d\theta}{dz}\right|_{z=L} = 0.$$

Applying the above BCs to the general solution for the excess temperature [Equation (2.53)],

$$\begin{aligned}\theta'(0) = {} & pd_1 - pd_2 + \int_{\lambda_1}^{\lambda_4} 2m_1(\lambda)[F(\lambda) - G(\lambda)]d\lambda \\ & + \int_{\lambda_2}^{\lambda_3} H(\lambda)M(\lambda)\sin[L \cdot M(\lambda)]d\lambda + \int_{\lambda_2}^{\lambda_3} N(\lambda)M(\lambda)\cos[L \cdot M(\lambda)]d\lambda \\ & + \int_{\lambda_1}^{\lambda_2} V(\lambda) + X(\lambda)d\lambda + \int_{\lambda_3}^{\lambda_4} V(\lambda) + X(\lambda)d\lambda = 0,\end{aligned} \tag{2B.18}$$

$$\begin{aligned}\theta'(L) = {} & pd_1e^{pL} - pd_2e^{-pL} + \int_{\lambda_1}^{\lambda_4} 2m_1(\lambda)F(\lambda)e^{2m_1(\lambda)L}d\lambda \\ & - \int_{\lambda_1}^{\lambda_4} 2m_1(\lambda)G(\lambda)e^{-2m_1(\lambda)L}d\lambda + \int_{\lambda_2}^{\lambda_3} N(\lambda)M(\lambda)d\lambda \\ & + \int_{\lambda_1}^{\lambda_2} V(\lambda) + X(\lambda)d\lambda + \int_{\lambda_3}^{\lambda_4} V(\lambda) + X(\lambda)d\lambda = 0.\end{aligned} \tag{2B.19}$$

Solving (2B.14) and (2B.15) simultaneously, the two coefficients of the general solution can be determined as

$$\begin{aligned} d_1 = \frac{1}{p(1-e^{2pL})}\Bigg\{ & \int_{\lambda_1}^{\lambda_4} -2m_1(\lambda)\big[F(\lambda)\big(1-e^{L[2m_1(\lambda)+p]}\big) \\ & \qquad - G(\lambda)\big(1-e^{-L[2m_1(\lambda)-p]}\big)\big]d\lambda \\ & - \int_{\lambda_2}^{\lambda_3} H(\lambda)M(\lambda)\sin[L\cdot M(\lambda)] + N(\lambda)M(\lambda)\big(\cos[L\cdot M(\lambda)] - e^{pL}\big)d\lambda \\ & - \int_{\lambda_1}^{\lambda_2} [V(\lambda)+X(\lambda)]\big(1-e^{pL}\big)d\lambda - \int_{\lambda_3}^{\lambda_4} [V(\lambda)+X(\lambda)]\big(1-e^{pL}\big)d\lambda\Bigg\}, \end{aligned} \tag{2B.20}$$

$$\begin{aligned} d_2 = d_1 + \frac{1}{p}\Bigg\{ & \int_{\lambda_1}^{\lambda_4} 2m_1(\lambda)[F(\lambda)-G(\lambda)]d\lambda \\ & + \int_{\lambda_2}^{\lambda_3} H(\lambda)M(\lambda)\sin[L\cdot M(\lambda)] + N(\lambda)M(\lambda)\cos[L\cdot M(\lambda)]d\lambda \\ & + \int_{\lambda_1}^{\lambda_2} V(\lambda)+X(\lambda)d\lambda + \int_{\lambda_3}^{\lambda_4} V(\lambda)+X(\lambda)d\lambda\Bigg\}. \end{aligned} \tag{2B.21}$$

3

Photorefractive Materials and Devices for Passive Components in WDM Systems

Claire Gu[a], Yisi Liu[a], Yuan Xu[b], J.J. Pan[b], Fengqing Zhou[b], Liang Dong[b], and Henry He[b]

[a]*Department of Electrical Engineering, University of California, Santa Cruz, CA 95064, USA*
[b]*Lightwaves 2020 Inc., 1323 Great mall Drive, Milpitas, CA 95035, USA*

Abstract

The photorefractive effect is a phenomenon in which the local index of refraction is changed by the spatial variation of the light intensity. Although the phrase "photorefractive effect" has been traditionally used for such effects in electro-optic materials, new materials, including photopolymers and photosensitive glasses, have been developed in recent years and are playing increasingly important roles in optical fiber communication systems. Photopolymers in combination with liquid crystals are ideal materials for wavelength selective tunable devices. The improved optical quality and large dynamic range of photopolymers make them promising materials for holographic recording. Holographic gratings recorded in photopolymers can be employed as distributed Bragg reflectors (DBR). The large birefringence of liquid crystals can be used to tune the index of refraction to cover a large wavelength range (e.g., 40 nm). In addition, the combination of photopolymer and liquid crystal also leads to a new material known as holographic polymer dispersed liquid crystal (H-PDLC) which provides a medium for switchable holograms. Photonic devices made of these materials can be easily incorporated into an optical fiber system because of the low index of refraction of polymers and liquid crystals. Besides photopolymers, photosensitive glasses are also promising for applications in fiber optic systems. Fiber Bragg gratings (FBG) have been used as bandpass filters and dispersion compensators. In this chapter, we describe the applications of photopolymers, H-PDLCs, and FBGs in fiber optic devices. Specifically, some of the recent works on photonic devices such as filters, switches, and high performance dispersion compensators for wavelength division multiplexing (WDM) systems will be described.

3.1. INTRODUCTION

As computers and the Internet become faster and faster, more and more information is transmitted, received, processed, and stored everyday. The demand for high speed and large capacity information systems is pushing scientists and engineers to explore all

possible approaches including electrical and optical means. Optical data storage and fiber communications have already demonstrated their potential in the competition against other technologies. CD and DVD are showing their advantages in the computer and entertainment market. Optical fiber networks are offering unprecedented bandwidth in communications. What motivated the use of optical waves to transmit, process, store and access information is that light or an optical wave has an enormous capacity (or bandwidth) to carry information because of its short wavelength and parallel nature. Photorefractive materials, including traditional electro-optic photorefractive crystals as well as photopolymers and photosensitive glasses, have demonstrated their potential in information systems.

The conventional photorefractive effect [1] is an optical phenomenon in some electro-optic crystals where the local index of refraction is changed by the spatial variation of the light intensity. Such an effect was first discovered in $LiNbO_3$ crystals in the 1960s. The spatial index variation leads to the distortion of the wave front, and such an effect was referred to as "optical damage." The photorefractive effect has since been observed in many other electro-optic crystals, including $BaTiO_3$, $KNbO_3$, SBN, BSO, BGO, GaAs, CdTe, InP, etc. The photorefractive effect is generally believed to arise from optically generated charge carriers which migrate when the crystal is exposed to a spatially varying pattern of illumination with photons having sufficient energy. Migration of the charge carriers due to drift, diffusion and the photovoltaic effect produces a space-charge separation, which then gives rise to a strong space-charge field. Such a field induces a refractive index change via the electro-optic (Pockels) effect. This simple picture of the photorefractive effect can be employed to explain several interesting optical phenomena in these media.

Photorefractive materials are, by far, the most efficient media for the recording of volume dynamic holograms. In these media, information can be stored, retrieved and erased by the illumination of light, in real time. The holographic recording can be employed for 3D optical data storage with an ultra-high density, such a scheme of volume holographic storage offers the unique property of parallel readout with an extremely short access time. In addition to the efficient holographic response, beam coupling known as two-wave mixing (TWM) occurs naturally in photorefractive media. When two beams of coherent radiation intersect inside a photorefractive medium, a stationary index grating is formed. This index grating is spatially shifted by $\pi/2$ relative to the intensity pattern. Such a spatial phase shift leads to nonreciprocal energy transfer when these two beams propagate through the medium. The unique property of nonreciprocal energy transfer can be employed for many applications, including laser beam clean-up, photorefractive resonators, nonreciprocal transmission window, biased elements for laser gyros, self-pumped phase conjugators, mutually pumped phase conjugators, optical interconnection, neural networks, phase conjugate interferometry, etc. It is important to note that the direction of energy flow in TWM is determined by the orientation of the crystal. In addition to holographic storage and TWM, photorefractive crystals are also efficient media for four-wave mixing (FWM) which is a generic process for the generation of phase conjugate waves. Optical FWM with various boundary conditions can be employed to construct several different types of phase conjugators including, externally-pumped phase conjugators, ring conjugators, self-pumped phase conjugators (SPPC), mutually-pumped phase conjugators (MPPC), etc.

Unconventional photorefractive materials including photopolymers and photosensitive glasses are attracting more and more attention for better materials needed in fabrication of fiber optic devices. Traditional photorefractive materials, such as $LiNbO_3$, have been widely used to record holographic gratings in the applications of optical data storage, information processing, and fiber optic devices. However, new materials with larger

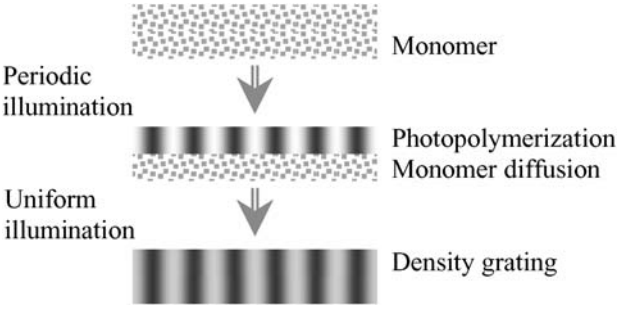


FIGURE 3.1. Grating formation in a photopolymer.

dynamic range, higher photo-sensitivity, lower refractive index, and relatively easier fabrication and integration processes are desirable for these applications, particularly in fiber optic devices.

Photopolymers in combination with liquid crystals are ideal materials for wavelength selective tunable devices. The improved optical quality and large dynamic range of photopolymers make them promising materials for holographic recording. Holographic gratings recorded in photopolymers can be employed as distributed Bragg reflectors (DBR). The large birefringence of liquid crystals can be used to tune the index of refraction to cover a large wavelength range. In addition, the combination of photopolymers and liquid crystals also leads to a new material known as holographic polymer dispersed liquid crystal (H-PDLC), which provides a medium for switchable holograms. Besides photopolymers, photosensitive glasses are also promising for applications in fiber optic systems. Fiber Bragg gratings (FBG) have been used as bandpass filters and dispersion compensators.

New and improved photopolymers have been developed as a result of the search for better holographic materials for optical data storage [2,3]. Figure 3.1 illustrates the process of grating formation in a photopolymer. A photopolymer, before exposed to light, consists of photopolymerizable monomers dispersed in a matrix. Upon illumination by a spatially varying light pattern (sinusoidal intensity pattern generated by two interfering plane waves, for example) monomers in the bright areas become polymers. At the same time, the remaining monomers will diffuse to form a uniform distribution throughout the bulk of the medium. The sum of polymers and monomers form a density gratings, which can be fixed by a uniform illumination after the diffusion of monomers reaches a steady state. As the index of refraction of the polymer depends on its density, the density grating results in an index grating. Since the photopolymerization process is irreversible, gratings recorded in photopolymers are permanent and not optically erasable. The advantages of photopolymers are their relatively large dynamic range and nonvolatile nature. One problem with photopolymers is that the material shrinks during polymerization leading to a Bragg mismatch at read-out. In recent years, there has been significant effort [4–6] to improve the properties of photopolymers, such as higher optical quality, increased thickness, lower shrinkage, as well as larger dynamic range and higher photosensitivity.

A holographic polymer dispersed liquid crystal (H-PDLC) is a photopolymer mixed with liquid crystal (LC). During the grating formation inside the photopolymer–LC mixture, photopolymerization occurs in the bright regions faster than in the dark regions. While the monomer diffuses to the bright regions, the LC molecules diffuse to dark regions [7]. After the final uniform curing, the H-PDLC composite system consists of alternating layers of polymer planes and LC rich droplet planes. If the refractive index of the polymer is

matched with one of the principal refractive indices (n_o or n_e) of the LC but not the other, the grating recorded inside the H-PDLC can be switched on or off by an electrical field that changes the orientation of the LC molecules.

Photosensitivity in glass was first discovered in optical fibers by Ken Hill et al. in 1978 [8], followed by a breakthrough by Gerry Meltz et al. in 1989 reporting on holographic writing of gratings using single-photon absorption at 244 nm [9]. After that, fiber Bragg gratings (FBG) have found numerous applications in fiber optic devices, although the mechanisms of grating formation in photosensitive glass are still under investigation [10,11]. Two of the mechanisms are believed to be involved in the formation of index gratings in germanosilicate fibers: the formation of color centers that changes the index of refraction via Kramers-Kronig relationship and the densification that occurs inside glass fibers upon illumination by UV light.

In this chapter, we describe some recent works on photonic devices such as filters, switches, and dispersion compensators for WDM systems. Photopolymers, H-PDLCs, or photosensitive glass fibers are shown to be the material of choice in the fabrication of these devices.

3.2. TUNABLE FLAT-TOPPED FILTER

As an example, one of the devices that employs photorefractive materials a flat-topped tunable filter [12] for wavelength division multiplexing (WDM) optical networks. A photopolymer can be employed in the implementation of such a filter. As we know, WDM is one of the most promising technologies for increasing the information capacity of optical fiber communication. With WDM, multiple channels at closely spaced wavelengths are sent simultaneously over the same fiber. One of the essential components for WDM is a wavelength selective filter. Previously, several WDM filters have been proposed and discussed. However, these filters do not simultaneously satisfy the two important requirements of a WDM filter: wavelength tunability and flat-topped pass band. In a recent design, a Fabry-Perot etalon with multiple reflection gratings as the distributed Bragg reflectors (DBRs) is used. The DBRs lead to the flat-topped line-shape and an electro-optic material inside the Fabry-Perot etalon gives the tunability of the filter. The filter has a flat-topped pass band with about 1 nm linewidth and its wavelength can be tuned over the 40 nm range provided by Er-doped fiber amplifiers (EDFA).

3.2.1. Principle of Operation

The idea is as follows. For a regular Fabry-Perot etalon, the bandwidth of each transmission peak can be very narrow and only one particular wavelength is transmitted with maximum transmission. At this wavelength the roundtrip phase shift is a multiple integer times 2π, i.e.,

$$\phi_{prop} = 2(2\pi/\lambda_m)nL_C = 2m\pi, \tag{3.1}$$

where ϕ_{prop} indicates the phase shift due to propagation, and L_C is the cavity length. The resonant wavelength λ_m (m is an integer) is determined by the optical thickness nL_C of the cavity, and therefore by n, the refractive index of the medium in the cavity.

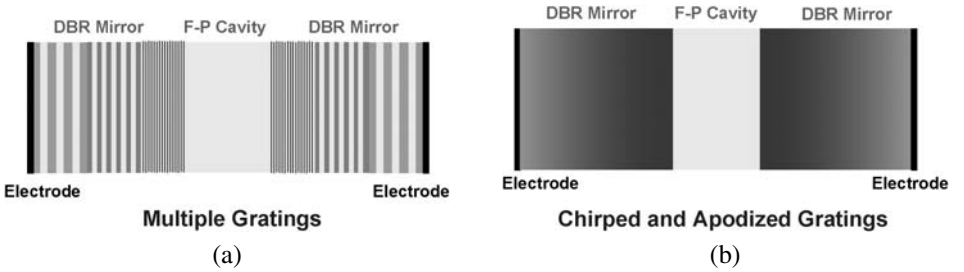


FIGURE 3.2. Fabry-Perot etalon with DBRs. (a) Each DBR consists of multiple spatially separated Bragg gratings. (b) Each DBR consists of a chirped and apodized grating.

Gratings are introduced as DBRs to vary the spectral line shape of the Fabry-Perot etalon. A grating has two functions: (1) a chromatic mirror that reflects light at a wavelength around

$$\lambda_B = 2n\Lambda, \tag{3.2}$$

where Λ is the grating period, n is the refractive index, and λ_B is known as the Bragg wavelength; and (2) a phase shifter that provides additional phase shift to that given in Equation (3.1). The resonant condition for peak transmission with a Fabry-Perot etalon with two symmetric DBRs is now written as

$$\phi_{prop} + 2\phi_{grating} = 2(2\pi/\lambda_m)nL_C + 2\phi_{grating} = 2m\pi, \tag{3.3}$$

where $\phi_{grating}$ is the phase shift upon reflection from each DBR. It is important to note that $\phi_{grating}$ depends on wavelength, as well as grating parameters. In order to have a spectral line-shape with sharp walls (also called tight skirts), one needs a high reflectivity for the DBRs. Therefore, the Fabry-Perot etalon should operate at the wavelength near that given by Equation (3.2). With a single grating DBR, the transmission spectrum still has a peaked line-shape (Lorentzian).

To make the line-shape flat-topped, the Fabry-Perot etalon should resonate at more than one wavelength. This is achieved by using multiple reflection gratings (with different grating period Λ_j, $j = 1, 2, \ldots, N$) as the DBRs. Each of the resonant wavelengths (λ_j, $j = 1, 2, \ldots, N$) results from a pair of gratings with certain grating period (Λ_j, $j = 1, 2, \ldots, N$). With multiple gratings, it is possible to have a narrow range of wavelength approximately satisfying the resonant conditions, Equation (3.3), simultaneously. In this case, the $\phi_{grating}$ in Equation (3.3) is the phase shift upon reflection from all gratings in one end. All wavelengths in this resonant range will have high (near 100%) transmission, resulting in a flat-top. These multiple gratings can be spatially separated, as illustrated in Figure 3.2(a), or they can overlap, or each DBR can be a chirped and apodized grating, as shown in Figure 3.2(b), or a combination of multiple chirped and apodized gratings.

To achieve the tunability, the index of refraction inside the cavity and the background index of refraction of the grating regions need to be modulated (electro-optically, for example).

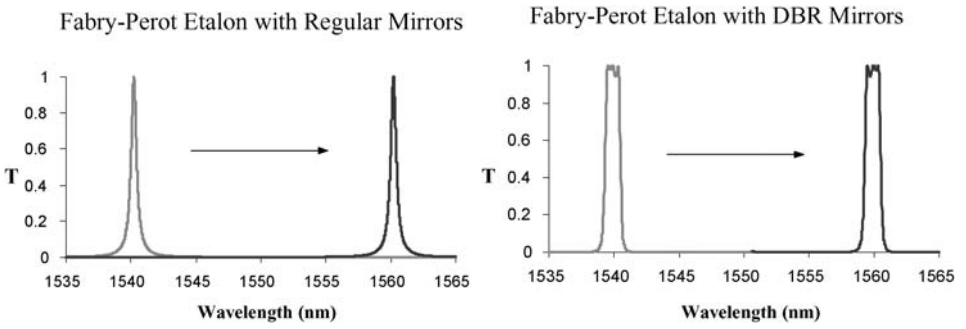


FIGURE 3.3. Comparison between line-shapes of different tunable filters. (a) A regular Fabry-Perot filter. (b) A Fabry-Perot filter with DBRs.

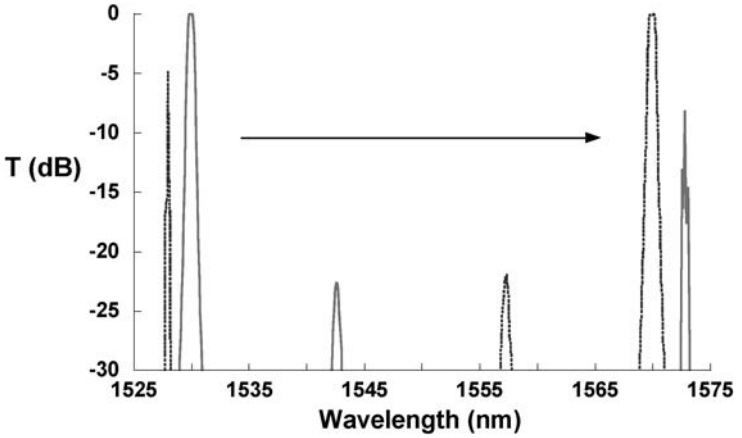


FIGURE 3.4. Transmission spectrum of an optimized flat-topped tunable filter.

3.2.2. Device Simulation

Calculated line-shapes of different tunable filters are compared in Figure 3.3. Figure 3.3(a) shows the transmission spectrum of a Fabry-Perot etalon with regular mirrors. By changing the index of refraction inside the cavity, the transmission peak is tuned from 1540 nm to 1560 nm. Figure 3.3(b) shows the transmission spectrum of a Fabry-Perot etalon with DBRs each consisting of three gratings. The nearly squared line-shape remains throughout the 40 nm tuning range (Figure 3.4) corresponding to the EDFA gain bandwidth. Adjusting grating parameters can minimize the ripples in the high transmission range. In Figure 3.3, we notice that by using DBRs with multiple gratings, the high transmission range is much wider (flat-top) and the slopes of the edges are much steeper (tight skirts). In this design, the pass-band of the filter is well within the stop-band of the DBRs, therefore, the reflectivity of the DBRs is nearly 100%. The tight skirts are achieved by this high reflectivity of the DBRs, therefore high finesse of the Fabry-Perot cavity. A fine tuned DBR with five gratings, two of them chirped, gives us the results shown in Figure 3.4. All side-lobes are either outside the tuning range or below -20 dB. The phase shift upon re-

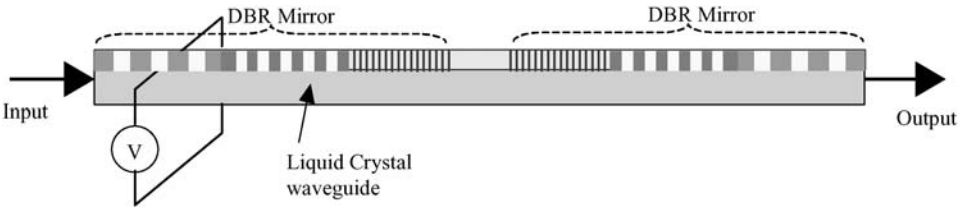


FIGURE 3.5. Schematic of a flat-topped tunable WDM filter based on a liquid crystal waveguide with external gratings, recorded in a layer of photopolymer, as DBRs.

flection from the DBR is calculated using the coupled mode theory. Details of the analysis can be found in Ref. [12].

3.2.3. Design for Implementation

In order to implement the design, one needs to (1) tune the index of refraction both inside the cavity and in the grating region, (2) fabricate the DBRs with multiple gratings. The required large tuning range for the index of refraction ($\Delta n \sim 0.04$) suggests the use of a liquid crystal material. On the other hand, the required interaction length for gratings (100 μm) suggests a holographic medium for the DBRs. The idea to implement a filter that satisfies all the above requirements is shown in Figure 3.5. It consists of a liquid crystal waveguide, whose index of refraction can be tuned by an applied electric field. On top of the liquid crystal waveguide is a layer of holographic material (e.g., photopolymer) which can be used to fabricate the DBRs optically. Multiple exposures in the grating regions can be performed to record multiple holographic gratings. The gratings in the polymer will reflect light waves traveling in the liquid crystal waveguide, therefore serve as DBRs. By applying an electric field across the liquid crystal waveguide, both the index of refraction inside the cavity and the background index of refraction of the grating regions can be tuned simultaneously. The choice of a photopolymer as the holographic material in this case is based on the requirement of low refractive index, permanent gratings, and easy fabrication process.

3.3. WAVELENGTH SELECTIVE 2 × 2 SWITCH

In recent years, the migration of telecom networks to all-optical networks has dramatically increased the demand for all-optical components. DWDM systems capable of increasing the network bandwidth using the currently installed fiber cables have been widely used in the telecommunication systems. Typical DWDM systems require a variety of functional wavelength to be routed throughout the network. Signals need to be optically added and dropped, optically cross-connected and switched. Optical switches with wavelength selectivity are of great importance and application in DWDM systems. One example of such a device is a switchable add/drop module that is capable of switching between all-through state and adding (or dropping) state for the designated wavelength. The building block of this switchable device is an optical add/drop module and a 2 × 2 switch. Most of the existing 2 × 2 switches, such as those involving two prisms operating in total internal reflection mode or total transmission mode, require light being coupled in and out of the fiber resulting in large device size and high insertion loss. Recently, a novel compact-size wavelength-selective switch by recording electrically switchable holographic gratings

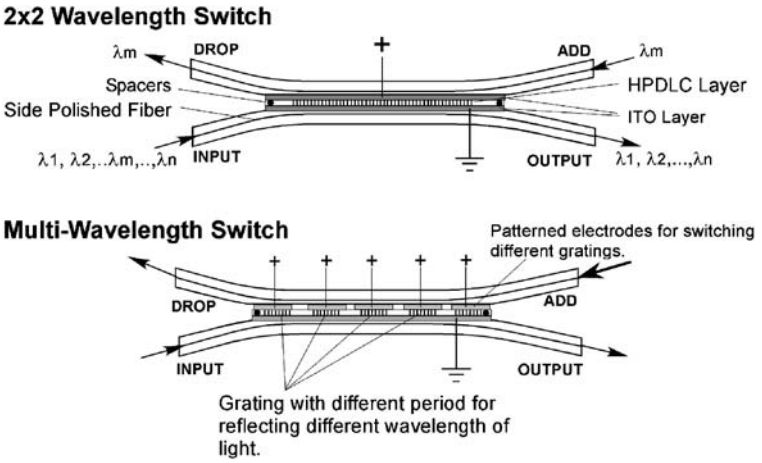


FIGURE 3.6. Structure of the 2 × 2 wavelength switch and multi-wavelength switch.

in a layer of holographic polymer dispersed liquid crystal (H-PDLC) [13–19] sandwiched between two side-polished fibers was proposed, analyzed and demonstrated. This device provides in-line operation capability and is particularly suitable for WDM network reconfiguration.

A holographic polymer dispersed liquid crystal (H-PDLC) is a photopolymer mixed with liquid crystal (LC). During the grating formation inside the photopolymer-LC mixture, photopolymerization occurs in the bright regions faster than in the dark regions. While the monomer diffuses to the bright regions, the LC molecules diffuse to dark regions [20]. After the final uniform curing (which may be optional if most of the monomer is polymerized during grating recording), the H-PDLC composite system consists of alternating layers of polymer planes and LC rich droplet planes. If the refractive index of the polymer is matched with one of the principal refractive indices (n_o or n_e) of the LC but not the other, the grating recorded inside the H-PDLC can be switched on or off by an electrical field that changes the orientation of the LC molecules.

3.3.1. Principle of Operation

A schematic drawing of the basic structure of the 2 × 2 wavelength switch is shown in Figure 3.6. Two fibers are partially cut through their claddings by side polishing. The two polished sides are coated with ITO electrodes. Spacers are placed between the two side-polished fibers to form a cell that is then filled with H-PDLC. A holographic grating can be recorded in the H-PDLC layer by interfering two plane waves from a laser. When exposed to an interference pattern, well defined structures of nano-sized LC droplets channels at lower intensity regions interspersed between polymer-chain channels at higher intensity regions will be formed. Due to the different refractive indices of the polymer and the liquid crystal, an index grating is recorded. When an electric field is applied on the H-PDLC layer, the refractive index of the LC can be changed to be the same as that of the polymer. Therefore, the index grating inside the H-PDLC can be switched on or off by the electric field. If one uses two single-mode fibers with slightly different propagation constants, one can avoid directional coupling between the two fibers when the grating is switched off. At the field-on state, the grating is switched off and light will propagate through each

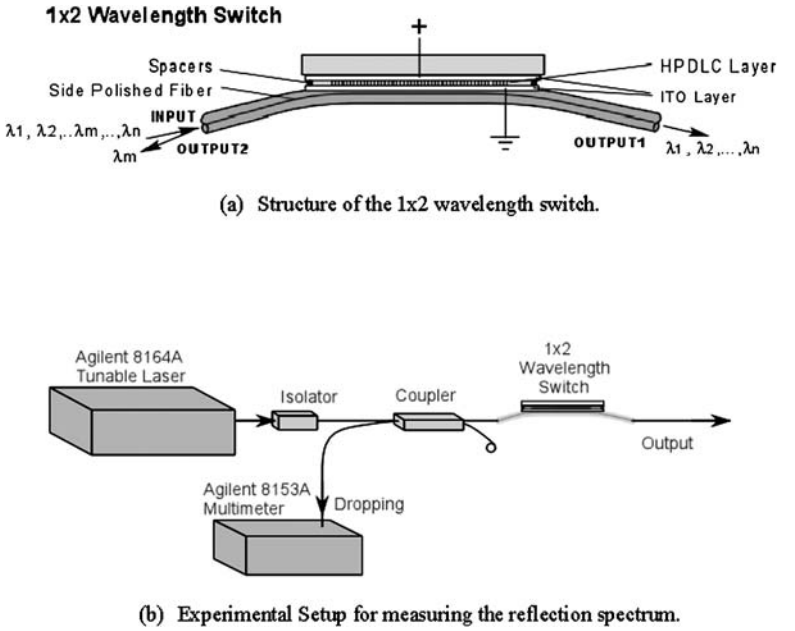


FIGURE 3.7. 1×2 wavelength switch and the experiment setup for demonstrating the device operation.

fiber without any coupling, which provides the all-through state. When there is no electric field, the grating is switched on and the incident light at the INPUT port will be reflected back and come out from the DROP port, which provides the wavelength dropping state. Meanwhile the incident light from the ADD port will be reflected into the OUTPUT port, which provides the wavelength adding state.

3.3.2. Experimental Demonstration

To provide a proof of concept demonstration one side-polished fiber with a H-PDLC cell built on top of it was used. The 1×2 switch structure is shown in Figure 3.7. A Corning SMF-28 fiber was side-polished carefully to a point of 0.5 μm to the core, with a length of 5 mm. ITO coating was deposited onto the polished surface for conduction. Separated by 20 μm spacers an ITO coated glass cover was placed on top of the side polished fiber to form a cell. H-PDLC was introduced into the gap. Then a grating was recorded in the H-PDLC layer by two interfering Ar beams. The angle between the two interfering beams (2θ) was calculated according to $\sin\theta = \bar{n}\lambda_{\mathrm{Ar}}/\lambda$, where λ_{Ar} is the wavelength of the Ar laser (488 nm), λ is the wavelength to be reflected and $\bar{n}$ is the mode index of the corresponding wavelength. To measure the reflected wave, a coupler as shown in Figure 3.7 is placed at the input side.

In the experiment, an Agilent 8164A tunable laser was used as the optical source, an Agilent 8153A multimeter was used to measure the reflected power. A 3 dB coupler was used to split the reflected beam and send half of it to the multimeter; an isolator was used to prevent the other reflected beam from going back into the source. A HP3245A was used to generate a square wave AC signal to drive the switch. The fiber that used in this experiment is a Corning SMF-28 fiber. The parameters are listed in Table 3.1.

TABLE 3.1.
Parameters for 1×2 switch.

Core radius ρ	4.5 μm
Cladding radius	62.5 μm
Core refractive index: n_{core}	1.4505
Cladding refractive index: $n_{cladding}$	1.4447
Distance between the polishing surface to center of the core	5 μm
Grating length	3 mm
Refractive index of H-PDLC	1.502
Thickness of H-DPLC layer	20 μm
Δn of LC in H-PDLC	0.07

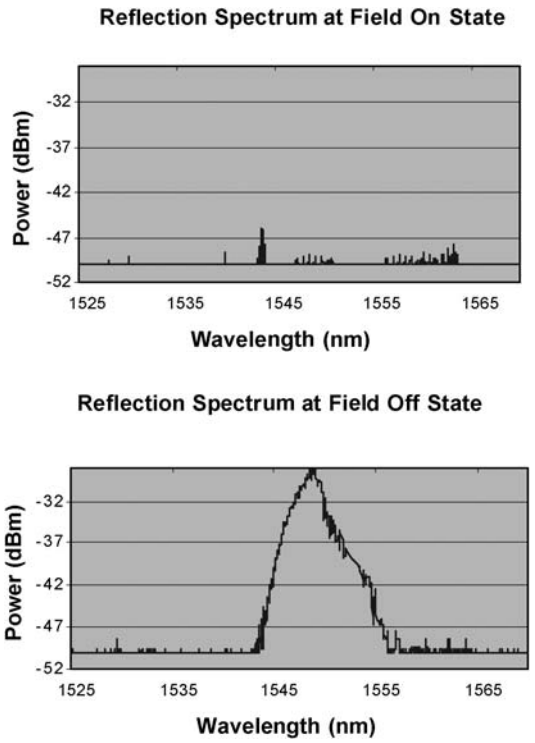


FIGURE 3.8. Reflection spectrum measured at field-on and field-off states of the switch.

The preliminary results of the experiment are shown in Figure 3.8. By adjusting the angle between the two writing beams during the holographic recording a reflection peak at 1548 nm in the field-off state is obtained. The 3 dB bandwidth of the reflection peak is about 3 nm. Adjusting the angle between the two writing beams can control the peak-wavelength of the reflection. When the electric field was applied, the holographic grating was switched off and an extinction ratio of more than 25 dB has been achieved. Further optimization of the experimental condition is expected to improve the results significantly in terms of diffraction efficiency and bandwidth. In this experiment, the H-PDLC has a higher index of refraction than that of the fiber core. In the field-on state, the insertion loss of the device is measured to be 20 dB. This gives us an estimate of the diffraction efficiency of the grating

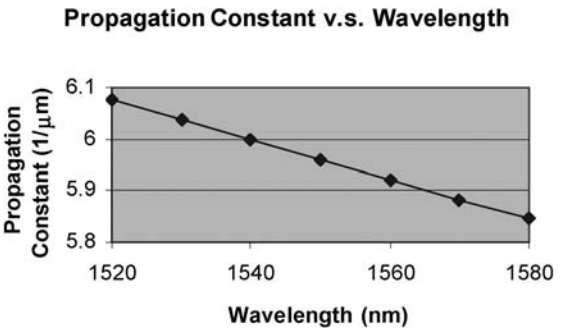


FIGURE 3.9. Propagation Constants calculated for various wavelength in the working range.

to be about 63% (−2 dB). However, the coupling loss due to the high refractive index of the H-PDLC layer can be more than 30 dB. In other words, most of the energy is coupled into the H-PDLC layer and is lost. It is believed that by using an H-PDLC layer with a reduced refractive index, one will be able to minimize the insertion loss. After it is polished to about 5 μm above the core, by using a special polishing technique, the insertion loss is often less than 0.2 dB. By applying low refractive index material on it, it should be able to compensate for the loss at the polished spot. In addition, by adjusting the grating parameters, such as grating length and apodization, etc., one can modify the reflection spectrum to achieve the required bandwidth and line shape. The side-polished fibers can have a polishing length of 10 mm or more.

Simply by replacing the coupler with a circulator and adding another circulator at the output end as the ADD port, one can achieve a Switchable Optical Add/Drop Multiplexer (SOADM). In the field off state, the selected wavelength from the incident light at the input port will be dropped, while the incident light at the ADD port will be added. In the field on state, the grating is invisible due to the matched refractive indices, no light will be added or dropped. By cascading several of these devices, one can also achieve an $N \times N$ Switchable MUX/DEMUX.

Alternatively, all of these devices can also be built based on the 2×2 switch shown in Figure 3.6, where two side-polished fibers are used. In this case, the circulators will no longer be necessary, which provides a more compact and less expensive design.

3.3.3. Theoretical Analysis

To understand the device performance and design improved switches, one needs to analyze the field distribution of the mode in the structure and the coupling between the forward and backward propagating waves. In a recent analysis, they are calculated using the vector modal method [21–26] and the coupled mode theory, respectively.

The modal properties of the D-shaped fiber covered with a dielectric film can be analyzed using the so-called vector modal method, where the field in each region is decomposed into the known eigen-modes of the fiber or slab waveguide, respectively. By matching the boundary conditions the propagation constant and the field distribution can be found. Figure 3.9 shows the propagation constant as a function of the wavelength. As can be noticed, the propagation constant varies almost linearly in the working wavelength range.

Figure 3.10 shows the field distribution inside the complex structure consisting of the D-shaped fiber, the H-PDLC slab waveguide and the outer cladding layer. Since the

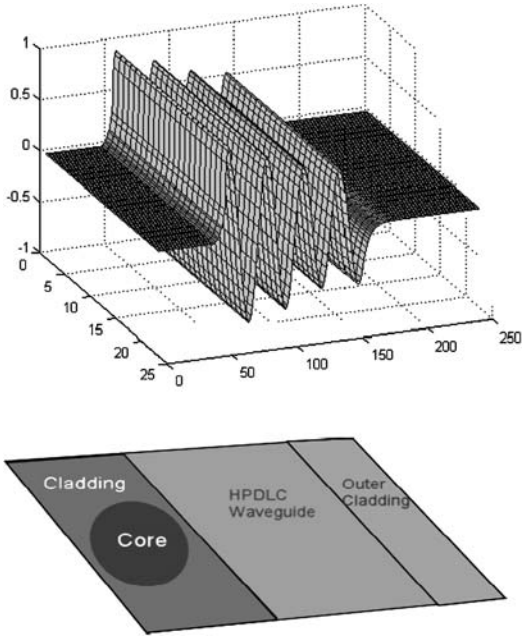


FIGURE 3.10. Field distribution within a side-polished fiber and a (high refractive index) planar waveguide system.

refractive index of the H-PDLC is higher than that of the fiber core, most of the energy is propagating inside the H-PDLC layer. The field remaining inside the fiber core is almost invisible in Figure 3.10. This result explains the high insertion loss that was measured in the experiment.

The coupling between the forward and backward propagating modes is analyzed using the well known coupled mode theory. In the coupled mode equations, the coupling coefficient is given by

$$\kappa = \frac{i\varpi\varepsilon_0 a_l \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \Delta n^2(x)|E(x,y)|^2 dxdy}{4\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}|E(x,y)|^2 dxdy}, \tag{3.4}$$

where $\Delta n(x)$ is the refractive index variation introduced by the grating and $E(x, y)$ is the modal field distribution. Once the coupling coefficient is found, the peak diffraction efficiency of the grating can be easily obtained by $\eta = \tanh^2(\kappa L)$ where L is the length of the grating. The reflection spectrum of the grating can also be obtained using the coupled mode analysis. Using parameters in the experiment, the reflected power is shown in Figure 3.11. The loss (scattering and absorption) of H-PDLC was experimentally measured to be 0.5 dB/cm, thus the absorption and scattering loss in the system was 0.25 dB (waveguide length was 0.5 cm). From the field distribution, the coupling loss would contribute -25 dB to the total power (including the 3 dB coupler in the system).

Furthermore, Equation (3.4) indicates that a stronger coupling (greater coupling coefficient) calls for a field distribution with a greater percentage of power inside the H-PDLC layer where $\Delta n(x)$ is non-zero. However, as discussed above, a large portion of the energy inside the H-PDLC causes a great insertion loss. The trade-off between the coupling effi-

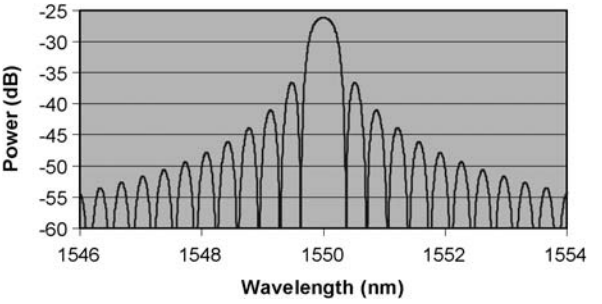


FIGURE 3.11. Theoretical simulation of the spectrum of the 1×2 switch.

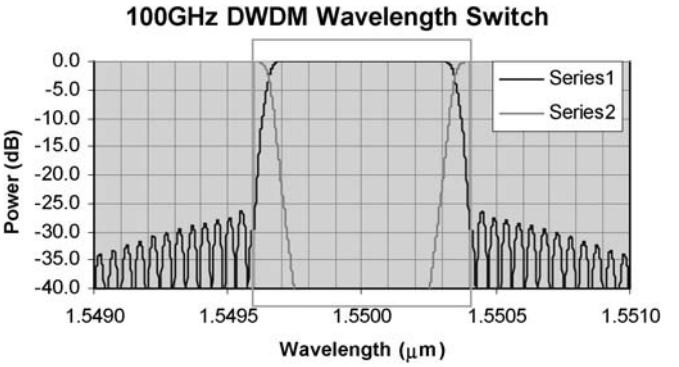


FIGURE 3.12. Spectrum of the 100 GHz DWDM wavelength switch design.

ciency and the insertion loss needs to be taken into consideration when parameters in the design are selected.

3.3.4. Optimized Switch Design

In order for the switch to be practically useful in DWDM systems, it is necessary to improve the performance dramatically. Specifically, it is necessary to decrease the insertion loss, increase the diffraction efficiency to almost 100%, decrease the bandwidth to 100 GHz or 50 GHz, make the line shape flat-topped [27], suppress the side lobes, and increase the switching speed. From the theoretical analysis, the insertion loss can be decreased by reducing the refractive index of the H-PDLC, the diffraction efficiency can be increased by increasing the grating length, the bandwidth and line shape can be modified by the coupling coefficient and the apodization. On the other hand, the trade-off between the coupling coefficient and the insertion loss implies that the refractive index of the H-PDLC cannot be too small. Taking all the restrictions into consideration, the following switches for 100 GHz and 50 GHz DWDM systems, respectively, have been designed.

3.3.4.1. 100 GHz DWDM Wavelength Switch For the 100 GHz DWDM wavelength switch design, the parameters are listed in the following Table 3.2.

Using the transfer matrix method, one can simulate the output spectrum of this device. The result is shown in Figure 3.12. The side-lobes have been successfully suppressed to be less than -26 dB while keeping the bandwidth to be within 0.8 nm. The peak inser-

TABLE 3.2.
Parameters used in 100 GHz DWDM wavelength switch design.

Fiber type	Corning SMF-28
n_{core}	1.4505
$n_{cladding}$	1.4447
r_{core}	9 μm
$r_{cladding}$	125 μm
Distance from the polished surface to core	0.5 μm
Cell thickness	20 μm
$n_{polymer}$	1.4475
Liquid crystal: n_{o}	1.4475
Liquid crystal: n_{e}	1.5125
Δn	0.065
Scattering and absorption loss	0.01 dB/mm
Grating length	16 mm
Apodization envelope function	sinc

TABLE 3.3.
Expected performance of the 100 GHz DWDM wavelength switch.

Switching bandwidth		100 GHz
Channel center wavelength		1550 nm
Reflection band @ 0.2 dB		0.65 nm
Reflection band @ 25 dB		0.8 nm
Channel isolation		>25 dB
Peak insertion loss within reflected band		<0.16 dB
Switching speed:	Rise time:	460 μs
	Fall time:	180 μs
Switching voltage		<200 V

tion loss in the reflection band is only −0.16 dB, which makes this device very desirable for telecommunication.

The switching speed of the H-PDLC cell has been experimentally measured to be a few hundred microseconds. the detailed specification of this design is listed in the following Table 3.3.

3.3.4.2. 50 GHz DWDM Wavelength Switch For the 50 GHz DWDM wavelength switch design, again, a "sinc" apodization envelope function was used to improve the reflection spectrum. The optimized parameters are listed in the following Table 3.4.

The spectrum of this 50 GHz DWDM wavelength switch has been simulated and optimized using the transfer matrix method. The result is shown in Figure 3.13. Notice that the first and the second pair of side-lobes have been suppressed down to less than −32 dB, while the peak insertion loss at the reflection band is only −0.18 dB.

The detailed specification of this 50 GHz DWDM wavelength switch design is listed in Table 3.5.

The switching voltage and switching speed can be further improved. By adjusting the proportion of the ingredients, it is believed that the switching voltage can be lowered to 100 V. By making the curing of H-PDLC more homogeneous, it is also possible to decrease the rise time to less than 200 μs.

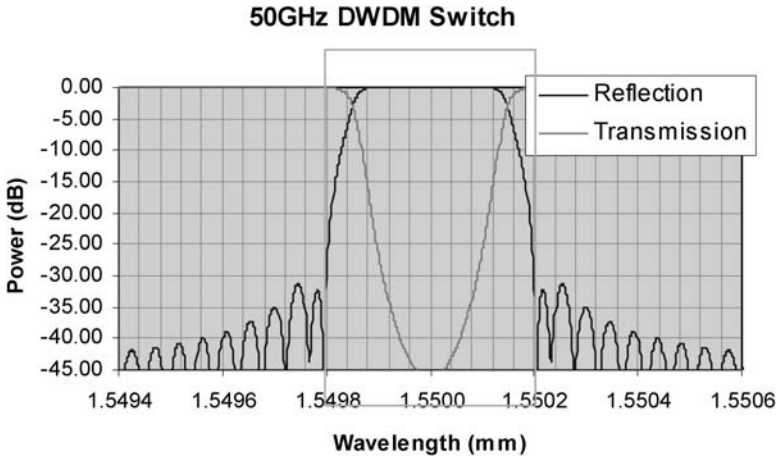


FIGURE 3.13. Spectrum of the 50 GHz DWDM wavelength switch design.

TABLE 3.4.
Optimized parameters used for 50 GHz DWDM wavelength switch design.

Fiber type	Corning SMF-28
n_{core}	1.4505
$n_{cladding}$	1.4447
r_{core}	9 μm
$r_{cladding}$	125 μm
Distance from the polished surface to core	0.5 μm
Cell thickness	20 μm
$n_{polymer}$	1.4475
Liquid crystal: n_o	1.4475
Liquid crystal: n_e	1.4755
Δn	0.028
Scattering and absorption loss	0.01 dB/mm
Grating length	18 mm
Apodization envelope function	sinc

TABLE 3.5.
Expected performance of the designed 50 GHz DWDM wavelength switch.

Switching bandwidth		50 GHz
Channel center wavelength		1550 nm
Reflection band @ 0.2 dB		0.28 nm
Reflection band @ 25 dB		0.4 nm
Channel isolation		>25 dB
Peak insertion loss within reflected band		<0.18 dB
Switching speed:	Rise time	460 μs
	Fall time:	180 μs
Switching voltage		<200 V

3.3.5. Discussion

In this section, a novel in-line 2 × 2 wavelength switch based on H-PDLC gratings sandwiched between two side-polished fibers [28–31] has been discussed. The idea has

been demonstrated experimentally and analyzed theoretically. Based on the experimental and theoretical results, optimized switches for both 100 GHz and 50 GHz DWDM systems have been designed. This device provides the in-line operation capability and is particularly suitable for DWDM network reconfiguration.

In addition, as shown in Figure 3.6, it is possible to build a multi-grating wavelength switch based on the same basic structure. With the polishing technique, the side polished part can be long enough to write several gratings at different spatial locations inside the H-PDLC layer. Using photolithography, patterned electrodes can be fabricated on the ITO layer on one of the side polished fibers. By applying control voltages to each of these electrodes, one will be able to switch the gratings on and off individually, thus choosing different wavelength to be reflected. This structure can be used in wavelength routing.

3.4. HIGH PERFORMANCE DISPERSION COMPENSATORS

Fiber Bragg gratings (FBGs) have been widely used for filters and dispersion compensators. Recently, several high performance dispersion compensators have been designed and implemented. Here, as examples, we present two types of high performance dispersion compensators. One type is a multi-channel dispersion-slope compensator implemented by post writing exposures. The other type is high reflectivity dispersion management filters implemented with a high precision FBG writing set-up.

3.4.1. Multi-Channel Dispersion-Slope Compensator

As transmission speed increases from 2.5 to 10 Gbits/sec and higher, and channel spacing decreases from 200 GHz to 50 GHz, chromatic dispersion and dispersion slope become limiting factors in wavelength division multiplexing (WDM) systems. Several techniques have been proposed for dispersion and dispersion-slope compensation, including: (i) compensation fiber [32,33], which is the most extensively used technology but is expensive and introduces a large amount of loss; (ii) virtual image phased array [34], which has a stability problem; (iii) conventional chirped fiber grating, which is ideal for dispersion compensation and may be used for slope compensation with nonlinear chirp but requires a long grating length to cover enough channels and precise period control to achieve the desired compensation.

Recently, however, sampled fiber Bragg grating (FBG) has attracted attention because of its unique properties, such as multiple reflection/transmission peaks with very precise spacing and relatively easy fabrication procedures compared with long chirped grating writing. These properties make FBG ideal for WDM applications. Multi-channel dispersion compensation has been shown experimentally using sampled chirped FBGs [35], and dispersion slope compensation has been proposed and demonstrated by several sampled grating based techniques. These include the use of a sampled non-linearly-chirped FBG grating [36] and a sampled chirped-grating with a linearly chirped sampling period [37–39]. The sampled nonlinearly chirped grating can also be used for tunable dispersion compensation but it requires precise period control and its channel bandwidth is limited. Theoretically, the chirped sampling period technique [Figure 3.14(a)] holds a lot of promise [37]. For fabrication, however, it requires very precise sampling period control during the grating writing and the grating performance cannot be further modified after writing, which limits repeatability and flexibility.

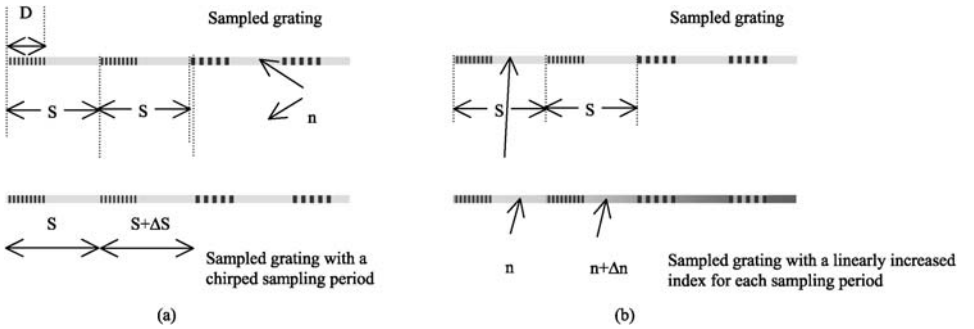


FIGURE 3.14. (a) A sampled grating with a chirped sampling period. (b) A sampled grating with a linearly increased index for each sampling period.

A novel optical fabrication method is used to achieve dispersion-slope compensation based on a sampled linearly chirped FBG with non-uniform refractive index for each sampling period. This grating is relatively easy to be controlled and fabricated and is flexible for different designs. It is demonstrated experimentally in a multi-channel 50 GHz WDM system. The dispersion slope is demonstrated to match that of the Corning LS fiber.

Defining S as sampling period and D as grating length within each sampling period, the optical path length of each non-grating part can be presented by $n(S - D)$, where "n" is the refractive index. Comparing the method of changing the sampling period (Figure 3.14(a), period changing expressed as ΔS) and that of changing the refractive index (represented by Δn) of each sampling period (Figure 3.14(b)) in a sampled grating, they are similar to each other. They both affect the optical path length. Based on this concept, fabricating a sampled grating with a linearly chirped sampling period is equivalent to fabricating a sampled grating with a linearly increased index for each sampling period, as long as $n\Delta S = \Delta n(S - D)$ or $\Delta n/n = \Delta S/(S - D)$. However, adjusting the refractive index is less problematic since Δn can be controlled by applying a UV exposure easily. With a CW laser and step motor control system, a different refractive index in each sampling period can be achieved by post-exposure of UV light with different exposure times. Therefore, a linear or non-linear sampling index profile can be achieved, assuming that the index modulation Δn is proportional to the exposure time Δt. Figure 3.15 shows the procedure for fabricating a sampled chirped grating with a linearly increased index per sampling period. This is equivalent to a sampled chirped grating with a linearly chirped sampling period, which was proposed for dispersion slope compensation [37].

Both simulations and experiments have been done to demonstrate that a sampled chirped grating with a linearly increased refractive index per sampling period is suitable for dispersion slope compensation in a 50 GHz WDM system. A rectangular sampling function and linearly increased index profile were applied, as shown in Figure 3.15.

Figure 3.16 shows the simulation results of a sampled chirped grating's spectrum, group delay (GD) and channel dispersion with a uniform refractive index per sampling period and those with a linearly increased index. It shows that the grating's dispersion has been modified by applying a linearly increased refractive index for each sampling period. The grating's relative dispersion slope (RDS) is optimized to $-0.044\ \text{nm}^{-1}$, which is designed to compensate Corning LS fiber's dispersion slope with RDS equaling to $-0.044\ \text{nm}^{-1}$ [37] (RDS is defined as dispersion slope normalized to dispersion). The grating's sampling period is 2.07 mm, which corresponds to 50 GHz channel spacing, with

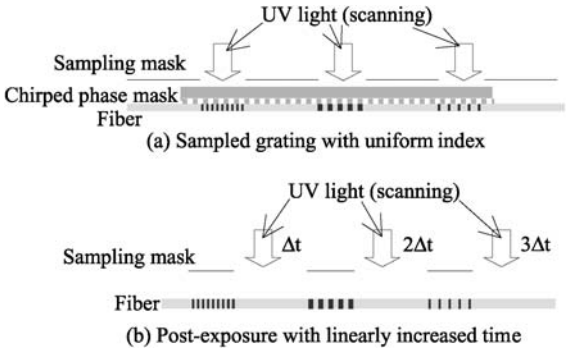


FIGURE 3.15. Schematics of the writing procedure.

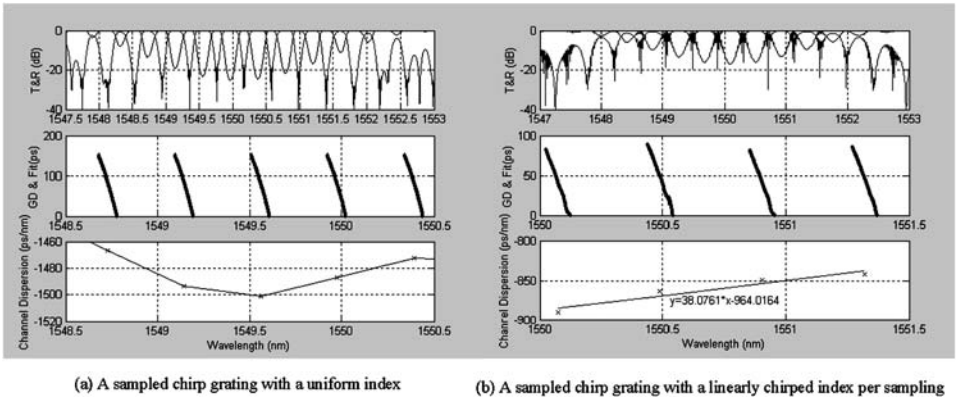


FIGURE 3.16. Simulation results for dispersion slope compensation of a 50 GHz WDM system.

15% duty cycle, defined as D/S. Figure 3.17 shows the experimental results of a sampled chirped grating's spectrum and its channels' dispersion, group delay and group delay ripple (a) before the linear post-exposure and (b) after the linear post-exposure.

In the experiment, the grating was written by a 50 mW 244 nm (frequency doubled Argon) CW laser and a holographic chirped phase mask with a 15% duty cycle sampling mask with 2.07 mm sampling period, which corresponds to 50 GHz channel spacing. The grating's reflection spectrum and group delay were measured by Agilent's "Chromatic Dispersion Test Solution" using a 2 GHz modulation frequency and 10 pm wavelength resolution. The grating length was 60 mm and the fiber was photosensitive and hydrogen loaded. Gaussian apodization was applied to minimize the group delay ripple (GDR).

Before post-exposure, the channel dispersion data points are distributed along a nonlinear curve, while after the post-exposure, these points are lined up along a straight line. With an optimized linearly step-increased index per sampling period, the channels' dispersion slope is adjusted to 23.625 ps/nm^2 and its RDS to -0.042 nm^{-1}, which matches Corning LS fiber's RDS. The relative error of each channel's dispersion compared to the linear fit for the 7 channels is: 0.33%, -0.46%, -0.08%, 0.21%, 1.69%, 5.57%, 1.21%, respectively, which shows that they match very well.

The spectrum's ripple increases with the non-uniform index. This is due to the phase mismatch for each sampling period caused by the linearly increased refractive index. The

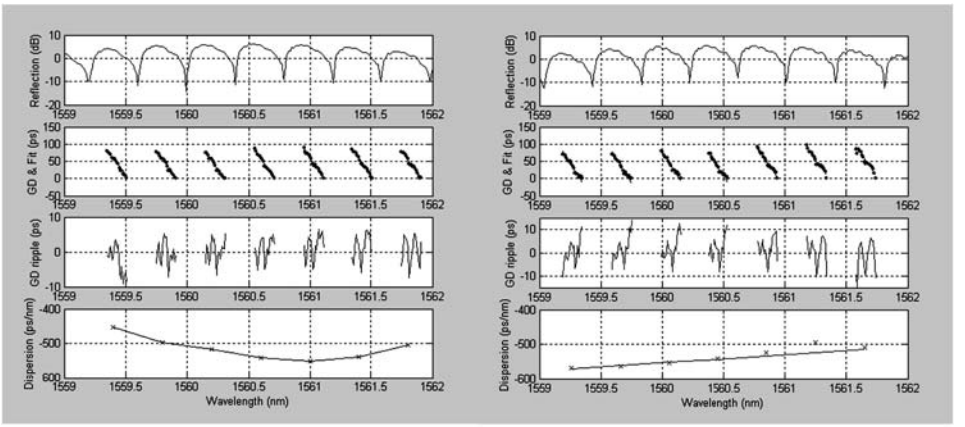


FIGURE 3.17. Experimental results for dispersion slope compensation in a 50 GHz WDM system.

group delay ripple (GDR) also increases, but only slightly. The maximum GDR in (a) is $\sim \pm 8$ ps and is $< \pm 15$ ps in (b) with bandwidth ~ 0.16 nm for each channel. The channel spacing is 50 GHz (0.4 nm) in both cases with $<0.5\%$ error while the center wavelength shifts a little after post-exposure.

Photo-sensitivity of the glass plays an essential role in fabricating this device. The potential of using post recording exposure as a method to compensate for dispersion slope is exceptional due to its ease of fabrication, low cost, flexibility and compatibility with other mature fiber grating techniques.

3.4.2. High Precision FBG Fabrication Method and Dispersion Management Filters

The novel writing system used in the experiment is based on a continuous phase-controlled writing configuration similar to those in previous work [40–44]. It allows apodization and phase of the fiber Bragg grating to be continuously controlled at each grating line. Special attentions have been given to interferometer design, minimum timing jitters, precise translation and accurate phase/apodization control. The fiber is moving constantly on a high precision motor with 0.3 nm resolution. A high-speed shutter system with less than 10 ns response time is utilized to repeatedly write small grating structure (<200 μm) in the fiber as the fiber moves through the interference pattern. The motion is carefully synchronized with the shutter. A compact ultra stable interferometer design is designed and demonstrated. A phase mask is used to split the writing beam equally at desired angle. To minimize the phase perturbation introduced by any perturbations in the optical path, an out-of-plane bi-directional ring structure based interferometer is used. The interferometer looks similar to a folded Mach-Zehnder with the interfering spot folded back to almost the splitting spot. However, a closer look at the beam paths shows that it is different from a simple folded Mach-Zehnder. Each beam is reflected by one mirror in the interferometer to the other mirror, then to the interfering spot where the fiber is located. After reflected by the mirrors, the two beams return to almost the same spot as the incident beam. With this design, the clockwise and count-clockwise beams travel through almost the same optical path. The phase-delay caused by any variation along the optical path will be experienced by both interference beams. Therefore, on the fiber where the grating is written, the

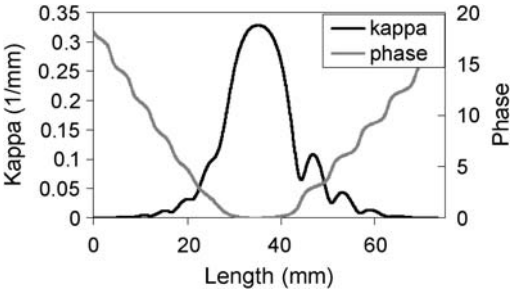


FIGURE 3.18. Grating design for a 25 GHz 99.9% filter with −700 ps/nm of dispersion.

interference pattern's phase remains stable. The fiber grating's phase is varied by the PZT controlled-phase mask, where the phase of interference fringe is accurately adjusted by a closed loop control of the PZT motion.

The first design is a 25 GHz 99.9% grating with a constant −700 ps/nm dispersion, which combines the characters of both dispersion compensator and optical filter. Therefore, the number of components in an optical network is minimized and the network is expected to have a better performance with lower cost. The design is shown in Figure 3.18. The phase structure in gray curve in Figure 3.18 is continuous with smoothed π phase jumps, which is very different from a zero dispersion grating. The coupling coefficient, kappa, does not have the zero-crossings as in a zero dispersion grating associated with the discrete π phase jumps (see the black curve in Figure 3.18). It has some small ripples at the front end and some more pronounced ripples at the back end of the grating. The grating is about 73 mm long. The target reflection and group delay from the design are shown in Figure 3.19 in gray and black solid lines respectively. The measured reflection and group delay are given in gray and black dotted lines respectively. The measured group delay ripple is less than ± 1 ps. It can be seen that a good agreement and performance is achieved for this filter. The target transmission and transmission dispersion are given in black and gray solid lines respectively in Figure 3.20. The measured transmission and transmission dispersion are given in black and gray dotted lines respectively in Figure 3.20. The measured grating has a reflectivity of 99.84%, corresponding to a transmission isolation of 28 dB.

The second design example is a 25 GHz 99.9% grating with a dispersion varying linearly from 1000 ps/nm to −1000 ps/nm. This design is shown in Figure 3.21. Slighted smoothed π phase jumps can be seen in the phase structure along with a slow varying background (see the gray curve in Figure 3.21). The coupling coefficient in black curve in Figure 3.21 bears many similarities to the zero dispersion grating design with many zero-crossings corresponding to the π phase jumps. There is a small asymmetry in the ripples at the two sides of the main peak as well as in the main peak itself. The grating is 72 mm long. This grating has a quadratic delay (see the solid gray line in Figure 3.22). This is very difficult to do by conventional chirp control. But it is, however, easy to do with phase control. The measured delay from a fabricated grating is also shown in Figure 3.22 in gray dotted line. It is in good agreement with the design. The target transmission and transmission dispersion is shown in black and gray solid lines in Figure 3.23, respectively. The measured transmission and transmission dispersion are given in black and gray dotted lines, respectively, in Figure 3.23. The measured grating has a reflectivity of 99.94%, corresponding to a transmission isolation of 32 dB.

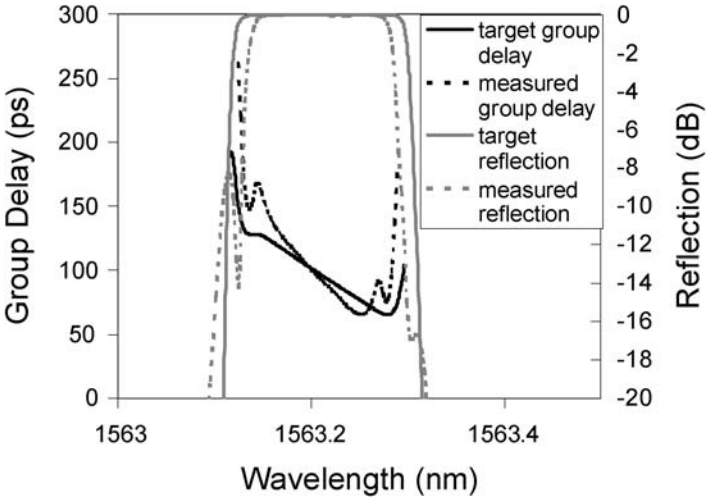


FIGURE 3.19. Target and measured reflection and delay of the grating design shown in Figure 3.18.

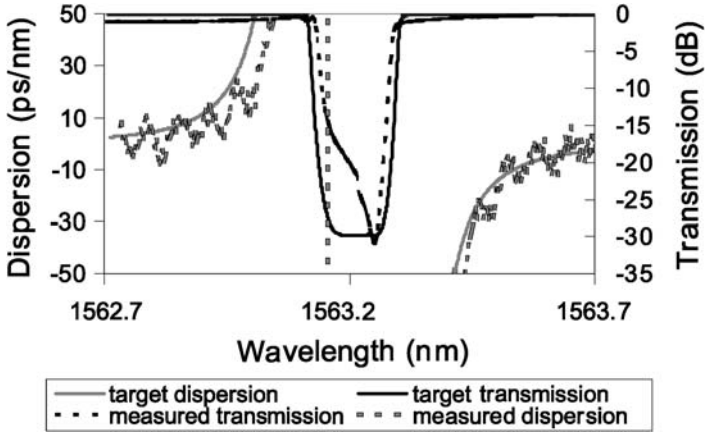


FIGURE 3.20. Target and measured transmission and transmission dispersion of the grating design shown in Figure 3.18.

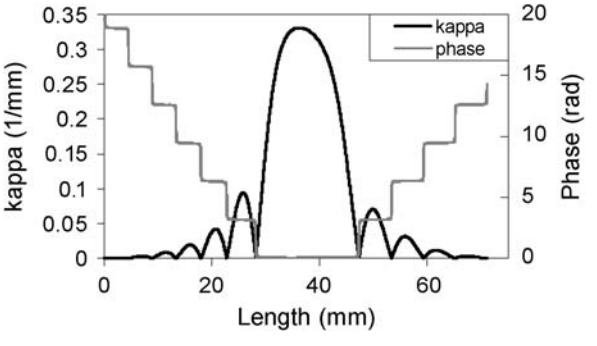


FIGURE 3.21. Grating design for a 25 GHz 99.9% filter with dispersion linearly varying from −1000 ps/nm to 1000 ps/nm.

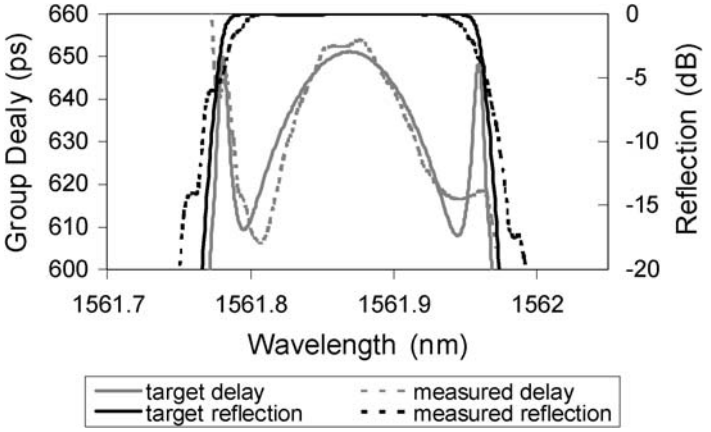


FIGURE 3.22. Target and measured reflection and delay of the grating design shown in Figure 3.21.

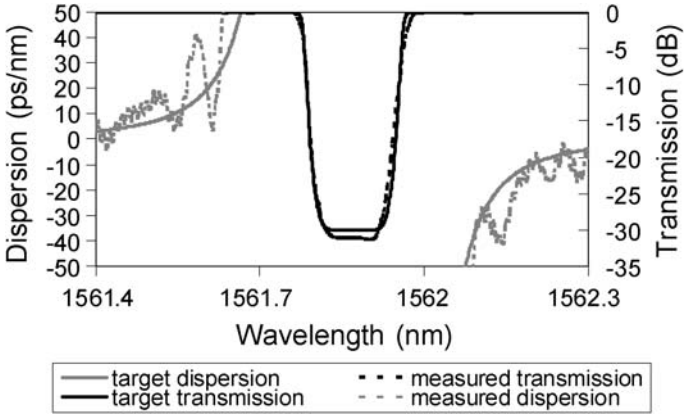


FIGURE 3.23. Target and measured transmission and transmission dispersion with the grating design shown in Figure 3.21.

The two examples here show that optical filters with arbitrarily engineered amplitude and phase and strong enough for sufficient transmission isolation can be accurately achieved by fiber Bragg gratings. This will open up a range of new applications, which have not been possible today.

The ability to continuously control both apodization and phase of a fiber Bragg grating can lead to a large number of new optical filters. The wavelength-division-multiplexing filters with arbitrary amount of chirp can be used at add/drop nodes or terminal nodes to combine the functionalities of filtering and dispersion compensation. This leads to much freedom in designing system dispersion map and location of add/drop nodes. Optical filters with any desired amplitude and phase can be made for various pulses shaping to generate a desired pulse shape and phase from a known in-coming pulse [43]. These filters also enable more freedom in code designs in optical code division multiplexing [44].

3.5. CONCLUSIONS

We have described the applications of photopolymers, H-PDLCs, and FBGs in fiber optic devices. Specifically, as examples, we described a flat-topped tunable WDM filter incorporating DBRs recorded holographically in photopolymers, a wavelength selective switch based on switchable gratings in holographic polymer dispersed liquid crystals, several high performance dispersion compensators using FBGs including a multi-channel dispersion slope compensator using a novel sampled fiber Bragg grating with post recording exposures, and two high reflectivity dispersion management filters. These devices are particularly suitable for WDM networks and have demonstrated the potential for unconventional photorefractive materials to be used in fiber optic devices for WDM systems.

REFERENCES

1. See, for example, P. Yeh and C. Gu, Eds., Landmark Papers on Photorefractive Nonlinear Optics, World Scientific, New Jersey, 1995.
2. L. Dhar, A. Hale, H.E. Katz, M.L. Schilling, M.G. Schnoes, and F.C. Schilling, Recording media that exhibit high dynamic range for digital holographic data storage, Opt. Lett., 24, pp. 487–489 (1999).
3. W.L. Wilson, K. Curtis, M. Tackitt, A. Hill, A. Hale, M. Schilling, C. Boyd, S. Campbell, L. Dhar, and A. Harris, High density, high performance optical data storage via volume holography: Viability at last? Optical and Quantum Electronics, 32, pp. 393–404 (2000).
4. R.T. Ingwall and D. Waldman, Photopolymer systems, in H.J. Coufal, D. Psaltis, and G.T. Sincerbox, Eds., Holographic Data Storage, Springer-Verlag, New York, 2000, pp. 171–197.
5. L. Dhar, M.G. Schnoes, H.E. Katz, A. Hale, M.L. Schilling, and A.L. Harris, Photopolymers for digital holographic data storage, in H.J. Coufal, D. Psaltis, and G.T. Sincerbox, Eds., Holographic Data Storage, Springer-Verlag, New York, 2000, pp. 200–208.
6. T. Bieringer, Photoaddressable polymers, in H.J. Coufal, D. Psaltis, and G.T. Sincerbox, Eds., Holographic Data Storage, Springer-Verlag, New York, 2000, pp. 209–228.
7. C.C. Bowley and G.P. Crawford, Diffusion kinetics of formation of holographic polymer-dispersed liquid crystal display materials, Appl. Phy. Lett., 76, p. 2235 (2000).
8. K.O. Hill, Y. Fujii, D.C. Johnson, and B.S. Kawasaki, Photosensitivity in optical waveguides: Application to reflection filter fabrication, Appl. Phys. Lett., 32, p. 647 (1978).
9. G. Meltz, W.W. Morey, and W.H. Glenn, Formation of Bragg gratings in optical fibers by transverse holographic method, Opt. Lett., 14, p. 823 (1989).
10. A. Othonos and K. Kalli, Fiber Bragg Gratings: Fundamentals and Applications in Telecommunications and Sensing, Artech House, Boston, 1999.
11. R. Kashyap, Fiber Bragg Gratings, Academic Press, San Diego, 1999.
12. M. Yang and C. Gu, Flattopped tunable wavelength-division-multiplexer filter design, Appl. Opt., 38, pp. 1692–1699 (1999).
13. R.L. Sutherland, L.V. Natarajan, V.P. Tondiglia, T.J. Bunning, and W.W. Adams, Switchable volume hologram material and devices, U.S. Patent #5,942,157, 1996.
14. M. Date, Y. Takeuchi, and K. Kato, Droplet size effect on the memory-mode operating temperature of smectic-A holographic polymer dispersed liquid crystal, Journal of Physics, D-Applied Physics., 32(24), pp. 3164–3168 (1999).
15. K. Tanaka, K. Kato, and M. Date, Fabrication of holographic polymer dispersed liquid crystal (H-PDLC) with high reflection efficiency, Japanese Journal of Applied Physics Part 2-Letters, 38(3A), pp. L277–L278 (1999).
16. T.J. Bunning, L.V. Natarajan, V.P. Tondiglia, and R.L. Sutherland, Holographic polymer-dispersed liquid crystals (H-PDLCs), Annual Review of Materials Science, 30, pp. 83–115 (2000).
17. K. Mimura and K. Sumiyoshi, Diffraction efficiency improvement in Holographic Polymer Dispersed Liquid Crystal (H-PDLC) devices, Molecular Crystals & Liquid Crystals, 346, pp. 239–244 (2000).
18. R.L. Sutherland, Polarization and switching properties of holographic polymer-dispersed liquid-crystal gratings. I. Theoretical model, Journal of the Optical Society of America, B-Optical Physics, 19(12) pp. 2995–3003 (2002).

19. J. Qi, H.Q. Xianyu, J.H. Liang, and G.P. Crawford, Active U-turn electrooptic switch formed in patterned holographic polymer-dispersed liquid crystals, IEEE Photonics Technology Letters, 15(5), pp. 685–687 (2003).
20. C.C. Bowley and G.P. Crawford, Diffusion kinetics of formation of holographic polymer-dispersed liquid crystal display materials, Appl. Phy. Lett., 76, p. 2235 (2000).
21. A.W. Snyder and J.D. Love, Optical Waveguide Theory, Chapman and Hall, London, 1983.
22. C. Vassallo, Rigorous theory for modes of optical fibers with cladding limited by a plane, ibid., 1986, 22, p. 944–945.
23. M.S. Dinleyici and D.B. Patterson, Vector modal solution of evanescent coupler, Journal of Lightwave Technology, 15(12), pp. 2316–2324 (1997).
24. M.S. Dinleyici and D.B. Patterson, Calculation of the wavelength filter properties of the fiber-slab waveguide structure using vector mode expansion, Journal of Lightwave Technology, 16(11), pp. 2034–2039 (1998).
25. D. Marcuse, F. Ladouceur, and J.D. Love, Vector modes of D-shaped fibers, Inst. Elec. Eng. Proc. J., 139, (Apr.), pp. 117–126 (1992).
26. D. Marcuse, Investigation of coupling between a fiber and an infinite slab, Journal of Lightwave Technology, 7(1), pp. 122–130 (1989).
27. M. Yang and C. Gu, Flat-topped tunable wavelength-division-multiplexer filter design, Appl. Opt., 38, pp. 1692–1699 (1999).
28. C. Gu, Y. Xu, Y. Liu, J.J. Pan, F. Zhou, and H. He, Applications of new photorefractive materials in fiber optic devices, SPIE Annual Meeting, Seattle, July 7–11, 2002, paper 4803-17.
29. C. Gu, Y. Xu, Y. Liu, J.J. Pan, F. Zhou, and H. He, Applications of Unconventional Photorefractive Materials in Fibre Devices, J. Opt. A: Pure Appl. Opt. 5, pp. S420–S427 (2003).
30. C. Gu, Y. Xu, Y. Liu, J.J. Pan, F. Zhou, and H. He, Applications of photorefractive materials in information storage, Processing and Communication, Optical Materials, 23, pp. 219–227 (2003).
31. C. Gu, Y. Liu, Y. Xu, J.J. Pan, F. Zhou, and H. He, Send a hologram, IEEE Circuits and Devices, 19(6), pp. 17–23 (2003).
32. V. Srikant, Broadband dispersion and dispersion slope compensation in high bit rate and ultra long haul systems, OFC2001, TuH1-1-3, 2001.
33. K. Aikawa, T. Suzuki, K. Himeno, and A. Wada, New dispersion-flattened hybrid optical fiber link composed of medium-dispersion large-effective-area fiber and negative dispersion fiber, OFC 2001, TuH6-1-3, 2001.
34. M. Shirasaki, Compensation of chromatic dispersion and dispersion slope using a virtually imaged phased array, OFC 2001, TuS1-1-3, 2001.
35. M. Ibsen, M.K. Durkin, M.J. Cole, and R.I. Laming, Sinc-sampled fiber Bragg gratings for identical multiple wavelength operation, IEEE Photonics Technology Letters, 10, pp. 842–844 (1998).
36. Y. Xie, S. Lee, Z. Pan, J.X. Cai, A.E. Willner, V. Grubsky, D.S. Starodubov, E. Salik, and J. Feinberg, Tunable compensation of the dispersion slope mismatch in dispersion-managed systems using a sampled nonlinearly chirped FBG, IEEE Photonics Technology Letters, 12, pp. 1417–1419 (2000).
37. W.H. Loh, F.Q. Zhou, and J.J. Pan, Sampled fiber grating based dispersion slope compensator, IEEE Photonics Technology Letters, 11, pp. 1280–1282 (1999).
38. X. Chen, Y. Luo, C. Fan, T. Wu, and S. Xie, Analytical expression of sampled Bragg gratings with chirp in sampling period and its application in dispersion management design in a WDM system, IEEE Photonics Technology Letters, 12, pp. 1013–1015 (2000).
39. X. Chen, X. Li, L. Xia, J. Wang, S. Xie, and Y. Yin, Numerical investigation of a stress-gradient sampled Bragg grating for dispersion compensation applications in wavelength-division multiplexing systems, Optics Communications, 187, pp. 363–367 (2001).
40. M. Ibsen, R. Feced, P. Petropoulos, and M.N. Zervas, 99.9% Reflectivity dispersion-less square-filter fiber Bragg gratings for high speed DWDM networks, OFC, PD21, Baltimore, 2000.
41. M.K. Durkin, R. Feced, C. Ramirez, and M.N. Zervas, Advanced fibre Bragg gratings for high performance dispersion compensation in DWDM systems, OFC, TuH4-1, 2000.
42. M. Ibsen et al., Broadband fibre gratings for pure thord-order dispersion compensation, OFC, Postdeadline paper FA7-1, 2002.
43. P. Petropoulos, M. Ibsen, A.D. Ellis, and D.J. Richardson, Rectangular pulse generation based on pulse reshaping using a superstructured fiber Bragg grating, Journal of Lightwave Technology, 19(5), pp. 746–752 (2001).
44. P.C. Teh, P. Petropoulos, M. Ibsen, and D.J. Richardson, Phase encoding and decoding of short pulses at 10 Gbit/s using superstructured fiber Bragg gratings, Photonics Technology Letters, 13(2), pp. 154–156 (2001).

4

Thin Films for Microelectronics and Photonics: Physics, Mechanics, Characterization, and Reliability*

David T. Read[a] and Alex A. Volinsky[b]

[a]*National Institute of Standards and Technology, Boulder, CO, USA*
[b]*University of South Florida, Tampa, FL, USA*

4.1. TERMINOLOGY AND SCOPE

4.1.1. Thin Films

Thin films of various types are a key component of modern microelectronic and photonic products. Conducting films form the interconnect layers in all chips, and dielectric films provide electrical insulation. With silicon-on-insulator (SOI) and strained silicon, semiconductor films have entered commercial design practice. The term *thin films* as used here refers to material layers deposited by vapor- or electrodeposition, with thicknesses too small to permit characterization by conventional mechanical testing procedures for bulk materials as described in, e.g., ASTM standards. Accordingly, our upper limit of thickness is taken as around 20 μm. Layers in this thickness range formed by other special processes, such as SOI (silicon on insulator) layers, also fall outside the standard mechanical test methods, and require thin-film characterization methods. Copper traces within multi-chip packages may be thicker than the definition given above, but some package designs include films within the present scope. Interconnect layers on die are included in a book on packaging because these layers are often considered to be "Level 0 packaging," since they are part of the packaging chain that connects the active devices to the outside environment electrically, mechanically, and thermally.

The relatively new field of microelectromechanical systems (MEMS) utilizes semiconductor fabrication techniques, especially lithographic patterning, to produce devices with moving parts and mechanical functions. Commercially important examples include

*Contribution of the U.S. National Institute of Standards and Technology. Not subject to copyright in the U.S. Certain trademarks are used in this manuscript for clarity; no endorsement of any commercial organization or product is given or implied.

accelerometers, used to trigger the deployment of automotive air bags, and pressure sensors. More recent examples are the movable mirrors used by Lucent, in optical switches, and by Texas Instruments, in Digital Light Processor (TM) systems. Thin films of a material set different from those in microelectronic devices are used for MEMS; polycrystalline silicon (polySi), designed for mechanical functions, is an important MEMS material.

The dimensions of films commonly used in microelectronic products have progressed well into the nanoscale at present, and MEMS technology is evolving the ability to produce nanoelectromechanical systems (NEMS). The elements of these products that are well below 1 μm in thickness challenge the current leading edge of mechanical characterization of thin films.

4.1.2. Motivation

The competitive pressure to produce "smaller, faster, cheaper" microelectronic devices means that reliability must be achieved using only the minimum amount of material, and also the minimum amount of testing. Actual testing of complete devices is the most definitive means to find out whether a design is reliable, and also the slowest and most expensive. Device designers use modeling and simulation to substitute for actual testing whenever possible. It seems evident that simulations based on an accurate understanding of the physics and mechanics of the materials involved, and carried out using accurate values of material properties, are more valuable than simulations based on *ad hoc* schemes and using guessed or estimated properties. The International SEMATECH roadmap has for years included the need for accurate modeling based on actual material properties. The 2003 version includes the statement [1]: *Cost effective first pass design success requires computer-aided design (CAD) tools that incorporate contextual reliability considerations in the design of new products and technologies. It is essential that advances in failure mechanism understanding and modeling, which result from the use of improved modeling and test methodologies, be used to provide input data for these new CAD tools. With these data and smart reliability CAD tools, the impact on product reliability of design selections can be evaluated.*

4.1.3. Chapter Outline

The current understanding of thin film materials has benefited from the energetic attention of materials science researchers. Several recent books discuss the deposition and behavior of thin films, with the general theme being the relationship between fabrication process, structure, and properties [2–5]. Since many different thin-film materials are critical components of various commercial products, and many different deposition processes are in use, this literature is voluminous. This chapter presents a broad overview of the physics and mechanics of thin films, with the goal of introducing topics and issues relevant to the reliability of films used in microelectronics and photonics. The latter sections, on mechanical properties and their characterization, treat their subjects in some depth. Again, however, complete descriptions must be sought in the original references. Section 4.2 presents a review of the physics and mechanics of thin films, with emphasis on basic physical features and recent findings relevant to the reliability of microelectronics and photonics. Section 4.3 introduces the physics underlying the most common failure modes and mechanisms of thin films, all of which relate closely to the state of stress, and argues that mechanical characterization is basic to design against these challenges; Section 4.4 reviews the techniques

available for mechanical characterization of thin films; Section 4.5 presents general expectations and rules of thumb for the mechanical behavior of thin films; and Section 4.6 gives results for some specific materials of interest for microelectronics and photonics. Section 4.7 identifies some issues where progress in understanding and characterization is urgently needed.

4.2. THIN FILM STRUCTURES AND MATERIALS

Microelectronic and photonic devices contain a variety of different types of thin films, introduced in this section. These devices are fabricated on massive substrates of single-crystal semiconductors. Millions, and more recently billions, of transistors must be placed on each die to create VLSI or ULSI (very or ultra large scale integrated) devices; semiconductor lasers and other photonic elements are also being driven to small form factors. A major branch of thin film technology, denoted as interconnect, has been developed to create the structures that connect these very tiny and unique active elements to each other and to the outside world. Interconnect films have thicknesses similar to metallurgical and anti-reflective coatings, but the deposition technologies and design schemes used in microelectronics and photonics are highly specialized. The reader should keep in mind that this book chapter focuses on films for microelectronics and photonics; information about the materials and deposition techniques particularly relevant to other technologies must be sought elsewhere.

Here we begin with substrates commonly used for electronic and photonic devices; we next touch on the conceptually simple but experimentally challenging case of epitaxial films; we introduce dielectric, metal, and polymer films, and special films for MEMS devices and for adhesion, barrier, buffer, and seed layers.

4.2.1. Substrates

The very idea of a thin film presumes the existence of a substrate; the difference in form and properties between film and substrate are fundamental to the state of stress in both the film and the substrate. Some useful values of mechanical properties of a few common substrate materials are given in Table 4.1. Crystalline substrate materials, especially silicon, are commonly elastically anisotropic, requiring three second-order elastic constants as shown in the table. Note that cubic symmetry is sufficient to ensure that the thermal expansion coefficient and the electrical and thermal conductivities are isotropic [6].

4.2.2. Epitaxial Films

An epitaxial film is a crystalline film on a crystalline substrate, where a lattice spacing of the film matches a periodic spacing of the substrate lattice. The benefit of epitaxy is that the microstructure of the film is controlled through selection of the substrate. Machlin gives a thorough, though "not encyclopedic", review of epitaxy [3]. He points out that many of the cases of epitaxy studied before 1985 were actually cases of graphoepitaxy. Examples would be metal films on sodium chloride or mica. These are cases of heteroepitaxy, because the film and substrate have different compositions. But graphoepitaxy refers specifically to the phenomenon where the bonding between the film and substrate atoms is so weak that the crystallographic alignment is produced by regular steps and terraces on the substrate

TABLE 4.1.
Selected room-temperature properties of some substrate and film materials [7–12]. Handbook properties are listed; microstructural effects such as texture may produce different values in specific cases.

Material	Crystal structure	Lattice parameter, Å	C_{11}, GPa	C_{12}, GPa	C_{44}, GPa	Anisotropy	Average Young's modulus, GPa	Average biaxial modulus, GPa	Biaxial modulus (100), GPa	Thermal expansion coefficient, $10^{-6}/°C$
Silicon	Diamond cubic	5.431	165.8	63.9	79.6	0.64	162.8	209.4	180.4	2.6
Germanium	Diamond cubic	5.658	128.5	48.3	66.8	0.60	131.6	166.0	140.5	5.8
Gallium arsenide	Zincblende (cubic)	5.653	118.8	53.7	59.4	0.55	116.2	153.1	123.9	5.74
Indium arsenide	Zincblende (cubic)	6.058	86.77	48.57	39.56	0.48	76.6	108.4	81.0	5.0[a]
Fused silica[b]	Amorphous	NA	NA	NA	NA	1	73.06	87.2	NA	0.49
Copper	FCC	3.615	168.4	121.4	75.4	0.31	128.2	195.4	114.7	16.8
Aluminum	FCC	4.050	106.8	60.7	28.2	0.82	70.0	107.3	98.4	23.6

[a] Average of values found in [10] and [11].

[b] The crystalline form of SiO_2 at room temperature is α-quartz. Its crystal structure is reported in the literature as both hexagonal and trigonal. It is actually trigonal, but is very close to hexagonal. Quartz crystals are used as a component in film thickness monitors in deposition systems. Ballato [13] notes an informative incident in the history of the measurement of the elastic constants of quartz. Atanasoff and Hart [14] made an extensive series of measurements of resonant frequencies of a set of quartz plates, as a means of measuring the elastic constants. Their worrisome conclusion was that their results were inconsistent with the trigonal symmetry then accepted for α-quartz. Later in the same year, Lawson [15] provided the explanation, that the piezoelectric nature of α-quartz influences its elastic behavior. By a proper mathematical consideration of the "converse piezoelectric effect," he showed that the experimental data of [14] were consistent with the trigonal symmetry, and provided a corrected set of elastic constants. These are listed in Table 4.1a. Table 4.1b lists the (hexagonal!) lattice parameters from [11] and thermal expansion coefficients for α-quartz from Touloukian et al.'s monumental work [12].

TABLE 4.1a.
Elastic constants of α-quartz, according to the trigonal crystal structure, derived by using the "converse piezoelectric effect" to correct the data of [14] at room temperature, from [15].

Elastic constant	Value, in units of GPa (or equivalently, 10^{10} dynes/cm^2)
C_{11}	86.75
C_{12}	6.87
C_{44}	57.86
C_{14}	17.96
C_{13}	11.3
C_{33}	106.8

To add yet another footnote to this cautionary tale, these same elastic constants are listed in [13], but C_{14} is shown with a different sign because [13] uses a different sign convention.

surface. Because graphoepitaxial films are prone to defects, they are not considered to be likely candidate materials for electronic or photonic devices.

Modern apparatus can maintain sufficiently high vacuum to produce true epitaxy, controlled by atomic bonding between the substrate atoms and the film atoms. The inter-

TABLE 4.1b.
Room-temperature lattice parameters [11] and thermal expansion coefficients [12] for α-quartz, reported according to the hexagonal crystal structure.

Direction	Lattice spacing, Å	Thermal expansion coefficient, 10^{-6}/°C
a	4.914	12.4
c	5.405	6.8

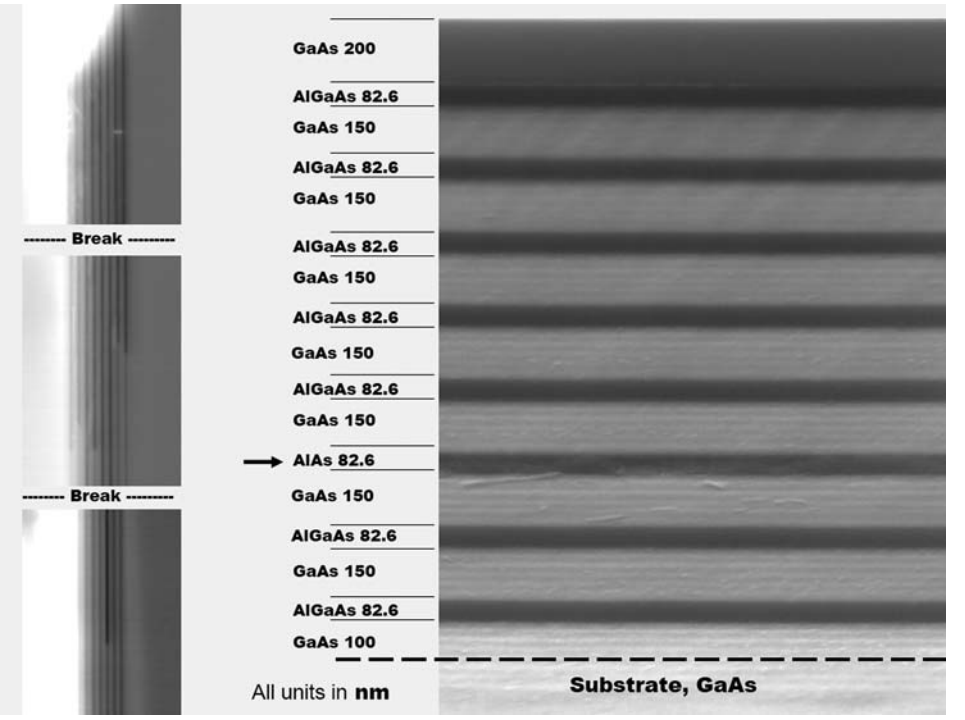


FIGURE 4.1. SEM micrograph showing heterogeneous epitaxial layers in a vertical cavity surface emitting laser (VCSEL) structure. This image is from a study of strains produced by oxidation of the AlAs layer [16,19].

face between epitaxial film and substrate can be more or less coherent, depending on the closeness of the match of their lattice parameters, the thickness of the epitaxial layer (or, epilayer), and other factors. According to Ohring [5], epitaxial deposition of silicon on silicon is the most commercially significant example of homoepitaxy, where the deposited material is of the same species as the substrate [5]. The epitaxial film is purer than the substrate and has fewer defects.

Technologically driven applications of heteroepitaxy are mainly in fabrication of optoelectronic devices made from wide-bandgap semiconductors. An experimental VCSEL (vertical cavity surface emitting laser) structure is shown in Figure 4.1 [16]. Intense commercial activity at present in this field has produced a bewildering variety of devices. Moglestue et al. describe the improvement produced by introducing AlGaAs layers into a GaAs photodetector [17]. The AlGaAs layers act as barriers to prevent the electrons and holes from traveling to low field regions in the GaAs; the result is improved perfor-

mance at high frequencies. Zirngibl et al. systematically describe a more sophisticated design [18]. These authors report a high-speed photodetector made of a strained superlattice of InGaAs/GaAs on a GaAs substrate. Here, the term "superlattice" refers to a series of alternating epitaxial layers of InGaAs and GaAs. In this example 120 pairs of layers were used, each on the order of 6 nm thick. The term "strained" refers to the fact that neither of the epilayers (epitaxial layers) fabricated in this study exists at its equilibrium lattice parameter, even though both the substrate and one of the layer types are composed of GaAs. This paper argues that the use of strained layers opens up the number of available materials systems for use in photonic devices, and that properly designed strained layers may remain stable and perform as well as lattice-matched layers. Epitaxy can be used to exploit even more sophisticated optoelectronic phenomena, such as distributed Bragg reflection and quantum wells and quantum dots.

The issue in all these structures is the manufacturability of a device with the needed electronic-photonic properties. The use of heterostructures may improve the optical performance, but may require that materials with different equilibrium lattice parameters be deposited in adjacent layers. The result is strain between the layers. If this strain becomes too large, the layer-to-layer interface will break down in some fashion, perhaps by the appearance of dislocations or other lattice defects, or perhaps by complete delamination. A thin layer can sometimes remain thermodynamically stable under more strain than a thick layer, because the strain energy accumulates with volume. But the variety of factors to be considered in selection of layer thicknesses includes a host of factors beyond strain, and may extend, for example, to the wavelength of light to be generated or detected.

So far we have considered only "blanket" layers covering the whole surface of a wafer. But designers are beginning to seriously consider quantum dot structures, where the epitaxy must extend in three dimensions, not just one. "Strain engineering" is the name of the effort to use strain effects to enhance the self-assembly and optoelectronic performance of quantum dots. Analytical techniques that treat the atoms in and around quantum dots are needed to fully describe their behavior; continuum mechanics calculations of strain provide only a rough estimate for the atomic displacements considered in the design of quantum dots.

4.2.3. Dielectric Films

It is often noted that the success of early thin film processes for metal-oxide-silicon (MOS) transistors was due as much to the robustness of readily produced silicon dioxide dielectric as to the semiconductor properties of silicon. One might also put the tenacious, electrically insulating natural oxide of aluminum into the list of material characteristics that contributed to the success of the aluminum–silicon dioxide interconnect structure. The traditional dielectric film for VLSI devices was some variant of silicon dioxide, for example, PECVD-TEOS, which stands for plasma enhanced chemical vapor deposition, using the precursor tetraethylorthosilicate. But for a variety of related reasons associated with electrical performance, which can be summarized as RC (resistive-capacitive) delay and power dissipation, insulating films with lower dielectric constant (k) are being incorporated into commercial devices [20]. These materials are known collectively as low-k dielectrics. The radical approach to the dielectric problem would be to use vacuum or gas as the dielectric, because vacuum has the lowest dielectric constant, and most gases do not raise it appreciably [21]. This approach has not yet proven to be popular but it is still being investigated. Silicate and silicate-carbonate films are the most widely studied and produced low-k films.

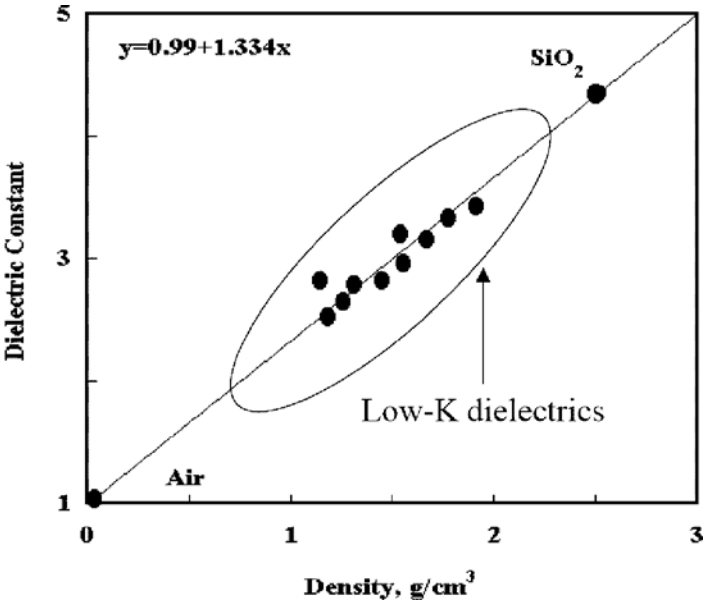


FIGURE 4.2. Dielectric constant plotted against density for a series of low-k dielectrics [25]. Values for air and SiO_2 are shown for reference.

These are amorphous materials and may be made highly porous to further reduce the dielectric constant. Figure 4.2 shows the trend for dielectric constant with density. This trend raises a problem, because the decrease in density implies an increase in porosity. Most low-k materials studied so far lack mechanical strength; the strength tends to go down along with the dielectric constant [22], because the lower-k materials are more porous. Volinsky et al. used nanoindentation, discussed below, to document the mechanical weakness of some low-k films [23–25]. The lack of strength is an issue because the interconnect stack on the back of a ULSI device is built up layer by layer, so that the first layers must be able to survive the manufacturing steps applied to later layers, especially CMP (chemical-mechanical planarization). The move to flip-chip designs, with solder bumps placed on top of the interconnect stack, means that the interconnect stack requires a certain degree of mechanical robustness. An improved low-k material would aid both manufacturing yield and device performance, so there is a great deal of research under way in this field at present.

4.2.4. Metal Films

Metal films are used in interconnect structures for their electrical conductivity. In the last few years, copper has displaced aluminum from ULSI devices largely because of its higher conductivity, even though it lacks the robust natural insulating oxide of aluminum. Highly reflective metal coatings are key elements in optical switch structures. Adhesion of the film to the substrate is important for useful films. Careful polishing, etching, and cleaning techniques are applied to commercially available substrates. The cleanliness of the substrate and the deposition system is critical to adhesion. A typical practical film deposition process for a metal film such as aluminum or copper onto a silicon substrate might include plasma cleaning of the substrate within the deposition chamber and the use of an adhesion layer such as Ti or Ta. A film intended for use as a conductor might have a thick-

ness around 1 μm, although large variations, say from 0.1 to 10 μm, are found depending on what the purpose of the film is and where it is located in the interconnect stack. In electrodeposition, the adhesion issue applies to both the seed layer on the substrate and the electrodeposit on the seed layer. The purity of the electroplating solution is obviously important.

4.2.5. Organic and Polymer Films

Organic polymers are traditional insulators used with metallic conductors and as coatings on optical fibers. Two examples are benzocyclobutene (BCB) and polyimide (PI). These polymers feature good thermal and mechanical stability; some varieties are photosensitive, which adds processing efficiency. These materials have elastic moduli only a few percent of those of metals, so they are highly compliant. Their strength can be comparable to that of metal films, ranging up to 100 MPa.

Organic semiconductors are being considered as active electronic elements for price-sensitive applications, where ULSI integration and gigahertz performance are not required [26]. Early organic semiconductor transistors were primitive [27], but considerable progress has been made since then. Organic semiconductors still face a challenging set of obstacles before widespread commercialization [28]. In particular, their reliability is likely to be an issue; few standard mechanical characterizations of such materials have been reported.

4.2.6. MEMS Structures

Microelectromechanical systems are microscopic mechanical systems constructed by use of the powerful photolithographic techniques developed in the microelectronics industry. Controllable micromirror arrays, accelerometers and orifices for inkjet printing are the leading commercial applications of MEMS at present. The fabrication of MEMS relies on chemical etching to remove one or more sacrificial layers. Typical materials used are a silicon nitride film to passivate the silicon substrate, followed by alternating layers of BSG (borosilicate glass, the sacrificial layer), and polySi. After the sacrificial BSG layers have been removed, the polySi mechanical elements become free to flex, extend, or rotate. MEMS-style structural elements, such as cantilever beams, are sometimes used as mechanical test structures for non-MEMS materials systems such as CMOS, as well as for MEMS materials [29]. The polySi layers are deposited by a chemical vapor deposition (CVD) route. The grain sizes of some very strong and adherent polySi films are quite small, only a few nanometers, and the strength may approach the traditionally predicted strength limit for crystalline solids, around one-thirtieth of the Young's modulus [30].

The packaging of MEMS systems is often very challenging and expensive, especially when the device must be in contact with the external environment. MEMS-style lithography techniques have been used to fabricate microfluidic devices, the "lab on a chip," which may become powerful tools for chemical analysis and bioassays.

4.2.7. Intermediate Layers: Adhesion, Barrier, Buffer, and Seed Layers

The use of one or more intermediate layers is a time-tested strategy in the design of thin film structures. The thickness of such layers is typically thin compared to the layer they are included to promote, and their electrical properties may be poor or irrelevant. A classic

example of buffer layers for chemical stability is the under bump metallurgy (UBM), the series of layers between the top ULSI interconnect layer and the solder. The need for intermediate layers, such as nickel, between aluminum and solder is evident, because to this day aluminum can be soldered only with great difficulty. However UBM is still considered useful for chips with copper interconnect layers [31]. Probably the earliest example of intermediate layers is the use of chromium below gold films on glass or quartz, to dramatically improve the adhesion. While both titanium and chromium [32] promote the adhesion of copper to SiO_2, titanium, tantalum (tantalum nitride), and Ti-W are more likely to be found in ULSI devices.

In today's competitive microelectronics/photonics design environment, the mantra "smaller, faster, cheaper" may also imply hotter, because higher operating currents promote speed. An important deleterious effect of higher temperatures on integrated microelectronic devices may be due to enhanced diffusion of chemical species into, out of, or through thin films. This issue was encountered in the silicon–aluminum–silicon oxide system when silicon diffused into the aluminum, causing a variety of problems. These were mitigated by alloying the aluminum films with a small amount of silicon, to prevent the uptake of additional silicon by the films. Later, when aluminum films containing a small amount of copper came into use because of their better electromigration resistance, titanium and titanium nitride barrier layers were used. This problem has grown more severe with the copper–silicon systems, and is controlled by the use of barrier layers to prevent copper from diffusing into the surrounding dielectric [33,34]. Titanium nitride and tantalum nitride barrier layers are used. A barrier layer that also enhances adhesion is clearly a design efficiency, and both titanium and tantalum nitride layers promote adhesion as well as serving as diffusion barriers.

It was noted above that coherent epitaxy and minimization of strain are incompatible objectives for heteroepitaxy when the epi layer has a lattice parameter different from that of the substrate. Buffer layers are used to "spread out" the strain. In such a case, the intermediate layers would be alloy layers sharing the composition of the substrate and epitaxial layers, with a graded composition to grade the lattice parameter.

Electrodeposition requires a conductive seed layer that is electrochemically compatible with the plated layer. The most straightforward choice is to use a seed layer of the same chemical element as the plated layer, for example, a copper seed layer for copper electrodeposition. The seed layer, of course, must be deposited by some technique other than electrodeposition, especially when the substrate is insulating. While copper seed layers have been used for ULSI designs with copper interconnects, research is in progress to use the barrier layer as the seed layer [35]. A combination seed, barrier, and adhesion layer would be an admirable example of design efficiency. As yet the verdict is still out on the ruthenium-copper system for ULSI.

4.3. MANUFACTURABILITY/RELIABILITY CHALLENGES

Manufacturability and reliability are critical when failure of a single transistor among millions can negate the economic value of a die. Producers of films for electronics and photonics are invariably most concerned with functional diagnostics; for example, does the film meet its intended electrical or optical performance criterion, such as electrical resistivity or optical reflectivity? But functional failures have their roots in the physical and mechanical behavior of the films. The selection of topics for this chapter has been guided

toward those relevant to manufacturability and reliability of microelectronic and photonic devices.

We note that mechanical failure is not the only threat to electronic devices; electromigration is included here as a special case because it clearly falls under the topic of reliability and it is a thin film failure mode. The prevention of corrosion, whether electrochemical or not, is an important necessity in the design and manufacture of electronic devices. But for corrosion, the physics and mechanics of thin films are secondary to chemistry, so it is left to the reader to pursue this topic elsewhere.

Throughout thin film technology, the concept of stress has been utilized as a guide to understanding, predicting and preventing instances of mechanical failure. Stresses created in the film deposition process are measured and minimized; sources of stress during service are identified and controlled; the response of materials to stress, namely, mechanical properties, are characterized and optimized; and even quantities that may influence the response of a thin film to stress, such as grain size, are considered and monitored.

4.3.1. Film Deposition and Stress

4.3.1.1. Thin Film Deposition Processes A set of deposition processes used for crystalline materials such as metals and ceramics applies energy to the film source material to separate individual atoms from the source and to launch them toward the substrate. Outlines of the principles of some methods for film deposition are given here to introduce the basis of some of the manufacturability and reliability issues; a complete treatment can be found in [4]. In *physical vapor deposition (PVD)*, commonly used to produce metal films, thermal energy is supplied to the source material by heating it, either electrically or by a directed electron beam. As the source material approaches its boiling temperature, individual atoms fly off the surface and travel in random directions. The process is carried out in vacuum, so that the atoms can reach the substrate before colliding with atmospheric gases. The usual practice is that only the source itself is hot, so the atoms are deposited on whatever surface they encounter first, which is often, but inefficiently, the walls of the vacuum chamber. The rate of production of free atoms is expressed in plots of vapor pressure *vs* temperature. The residual pressure in the deposition system, the composition of the residual gases, the temperature of the source, which controls the rate of evaporation, the temperature of the substrate, the distance from source to substrate, and the duration of the deposition are important variables. In *sputtering*, the source is bombarded with energetic atoms, usually atoms of an inert gas such as argon. Atoms from the source are knocked or "sputtered" off the surface, and travel around the chamber, sometimes steered by electric or magnetic fields. The pressure and energy of the sputtering gas are important parameters specific to this process. An advantage of sputtering is that source materials with low vapor pressure at practical temperatures can be deposited. The stress in sputtered films is highly variable, depending on the deposition conditions. Epitaxial films were discussed above; in *expitaxial* deposition, the uniformity of the crystal structure of the substrate surface, a carefully chosen match between the crystal structures of the substrate and the deposit, and overall cleanliness are the key issues. In *chemical vapor deposition*, the source is not used in pure form, but rather in the form of a chemical compound that can be transported into the deposition zone as a vapor. Within the deposition zone, a chemical reaction occurs, typically controlled by heating the substrate, which converts the element to be deposited to a form that sticks to the substrate and forms a solid film. Since the source is gaseous, it can be prevented from sticking to the walls of the deposition chamber or to any surface

it encounters by controlling their temperatures; therefore complex and highly curved surfaces can be coated. In *electrodeposition*, the electrically conductive source material and the substrate are immersed in a plating solution. Individual atoms or clusters of atoms are separated from the source by a combination of chemical dissolution and the local electric fields at the surface. The electric field set up between the substrate and the source directs the dissolved atoms to the substrate, where the local electric field promotes their incorporation into the deposit. This process incorporates a host of new variables, including the chemistry and temperature of the solution, the voltage, and the current density. An important consideration is the seed layer, needed to establish the electric field before any plating occurs. Additives or impurities in low concentration in the plating solution can influence the plating process significantly. Polymer films, which are rarely crystalline, are deposited by spreading a thin liquid layer of polymer, often mixed with solvent, on the substrate surface and drying and baking it. Typical examples are photoresist and polyimide.

In all deposition process, atoms collect on the substrate in structures that initially have low density and low order compared to the equilibrium structure of the film material. The mechanisms by which crystallographic order is gradually recreated within the deposited film have been subjects of intense study, as reviewed by Machlin [3] and other investigators cited therein. A key feature of these processes is the creation of stress in the newly deposited film.

4.3.1.2. Thin Film Stress It is safe to say that stress occurs during every thin film deposition process, although specially designed processes may be capable of eliminating stress, for example, by the incorporation of high temperature annealing built into the process. As an atom of the species being deposited strikes the substrate it loses most of its kinetic energy, especially if the substrate is maintained at a temperature well below the melting point of the film material. Stresses arise when the deposited atoms lack sufficient thermal energy to diffuse far enough and long enough to assume their equilibrium conditions of density and crystal structure. Stresses remaining after the substrate and film reach room temperature are termed "intrinsic" stresses; some authors use the term "residual stresses." Doerner and Nix have reviewed several stress-generation mechanisms [36]. Detailed schemes relating the deposition temperature to the stress are given in [3]. Floro et al. discuss the evolution of stress in SiGe and selected metals [37]. Sputtering is capable of producing compressive stresses. Deposition stresses may be great enough to peel a film from the substrate, to create plastic strains through stress relaxation, or to distort the shape of the substrate, creating the infamous "potato chip" effect. Ohring [5] gives a table of stresses encountered in several films, mainly dielectrics, but notes that the stresses are highly dependent on experimental variables and that no quantitative predictive theory is available.

At this point it is worth considering just what is meant by the phrase "thin film stress." Conceptually, as with stress in a small volume element of a continuum, the existence of stress implies that the atoms in the stressed element are constrained in a state of elevated elastic potential energy, which relieves itself instantaneously when the element is freed to deform. Four types of stress are traditionally distinguished: externally applied; lattice misfit; intrinsic, as defined above; and thermal. Lattice misfit stresses were mentioned above in the context of epitaxy. The principle is that the epitaxial layer may have an equilibrium lattice parameter different from that of its substrate; the rigid substrate prevents the epitaxial layer from stretching or shrinking to reach its nominal lattice parameter; thus, the epitaxial layer is in a state of stress. Thermal stresses arise when two materials with different thermal expansion coefficients are joined, and then heated or cooled while held rigidly in place. There are three main operational definitions or "acid tests" for the presence of film stress:

(a) if a substrate deforms evidently during a deposition process, or if a film spontaneously peels from a substrate, as described in the previous paragraph, one deduces that the film was stressed so highly that its tendency to deform could not be resisted by the substrate or by the adhesion of film to substrate; (b) if one of the crystal lattice parameters of the film is found to be different from its value in the bulk, stress is suspected, because films thicker than a few atomic layers have the same equilibrium lattice parameters as the bulk; (c) if a film is deposited on a previously flat rigid substrate, and the substrate is subsequently found to be curved, one deduces that stress in the film is the source of the force needed to elastically deform the substrate. Measurements of film lattice parameters, by X-ray techniques, and of substrate curvature, mainly by optical techniques, are discussed next. It is also noted that Swadener et al. have reported that intrinsic stress can be measured using depth-sensing indentation [38]. As discussed below, nanoindentation can also promote observable delamination of film from substrate; this phenomenon can be used as a diagnostic for film stress.

X-ray methods are common for measurement of lattice parameter and crystallographic texture for metal and ceramic films of all types. They can even provide some information about amorphous films in special cases. Grazing incidence X-ray reflectometry can be used to measure the thickness and density of very thin layers; the upper limit is poorly defined but is of the order of 0.1 μm. Gravimetric methods of measuring film density are impractical. Many characterization techniques applied to films, such as for example electrical conductivity or sound wave speed, depend on both the thickness and the density. The density is typically assumed to be nominal, but this is not necessarily the case. Therefore, X-ray reflectometry methods, even though they require special apparatus, may be advantageous in some cases because of their specificity.

X-ray diffraction to characterize a thin film can be carried out using a diffractometer set up for measurements of "powder patterns," with an appropriate specimen holder. The incident radiation would have a monochromatic spectrum. For a typical metal film deposited under relatively normal conditions, the X-ray diffraction record reveals the same crystal structure as for the bulk material; for example, the face-centered-cubic (FCC) structure of aluminum, copper, and gold is maintained. The diffraction peaks would be broadened, relative to bulk materials, due to the smaller thickness, grains, and subgrain domains in thin films. The intensity is low because the scattering material is only about 1 μm thick, but it is sufficient even for general-purpose instruments. The spectrum might be interrupted if a strongly-scattering set of planes in the substrate comes into alignment with the X-ray source and detector. In films just as in some bulk metals such as drawn wires or rolled plates, the relative amplitudes of the peaks may be influenced by preferred orientation or texture of the film. Measurements over a wide range of solid angles are needed to accurately evaluate the film texture. The lattice parameters deduced from the spectrum give essentially the same values of lattice parameter as listed in the handbooks for bulk material of the same composition. X-ray diffraction can be used to evaluate intrinsic stress if the lattice parameter can be measured with sufficient accuracy, and if, importantly, the value of the lattice parameter of the film material in its unstressed condition is known. However, most of the literature reports on this approach are for hard films rather than electronic interconnect materials.

Curvature measurement techniques for both wafers and cantilevers have been used to follow stress evolution, e.g., [37,39]. Wafer curvature as applied to mechanical properties measurement will appear again below. However, we note that Malhotra et al. [40] reported a difference between stress measured by curvature and by X-ray diffraction. The Stoney formula, which relates substrate curvature to film thickness and stress, contains the biaxial

elastic modulus. Because of the power of this equation in a variety of measurement and application contexts, it will be described below. Example data sets needed for application of this equation were given in Table 4.1. The concepts of tensor elasticity are used to relate single crystal elastic moduli, in particular the stiffnesses C_{ij} or the compliances S_{ij}, to practical moduli such as the biaxial modulus in the substrate plane [2]. Representative elastic moduli of polycrystalline aggregates are relevant for metal films like aluminum and copper. Some useful examples were listed in Table 4.1. de Lima et al. studied the thermal expansion coefficients and biaxial moduli for several combinations of materials relevant to microelectronics and photonics [41]. The scatter among their data and differences between their thin-film data and bulk values indicate that both preparation of representative specimens and carrying out precise and accurate measurements of these properties remain experimental challenges.

The Stoney formula is

$$\sigma_f = \frac{M_s t_s^2}{6Rt_f}. \tag{4.1}$$

Derivations, and the link back to the original source of this formula, are given in [2] and [5]. Here σ_f is the stress in the film; M_s is the biaxial modulus, given by $M_s = E_s/(1-\nu_s)$, where E_s is the Young's modulus of the substrate and ν_s is the Poisson's ratio of the substrate; t_s is the thickness of the substrate; t_f is the thickness of the film, and R is the radius of curvature of the substrate. This equation shows that for no curvature, where R is infinite, the stress in the film is zero. As the curvature increases, R becomes finite and the stress increases. For a given curvature, the stiffer the substrate (high M_s value), the larger the stress. And, for a given curvature, the thicker the film, the smaller the stress. This formula as written assumes that the stiffness of the thin film is negligible compared to that of the thick substrate. The assumption implies that Equation (4.1) can be applied without knowledge of the elastic properties of the film, an advantage in some cases.

The Stoney formula can be extended to a 2D case:

$$\sigma_R = \frac{E_s}{1-\nu_s}\frac{h_s^2}{6h_f R_1}\left[1+\left(\frac{\nu}{1+\nu}\right)\cdot\left(\frac{R_1}{R_2}-1\right)\right], \tag{4.2}$$

where R_1 and R_2 are the radii of curvature in the x–z and y–z planes respectively.

Substrate curvature radii can be measured accurately with laser deflection and optical lever cantilever beam techniques. This allows constructing a stress map over the whole wafer (Figure 4.3). Recently Shen et al. [42] have extended the curvature methods to measure residual stresses in patterned line structures. Treatment of MEMS-style cantilever beams, where the thicknesses of the two layers are comparable, would require a more complete formulation.

4.3.2. Grain Structure and Texture

A second major consequence of film deposition processes is that thin films generally are polycrystalline with grain sizes much smaller than bulk materials. Epitaxial semiconductor films are an exception, of course, and polymer films have no grains, hence no grain size. Relevant examples are all metal films produced by physical vapor deposition (PVD) or electrodeposition. Polycrystalline silicon (polySi) is not known in bulk form, but CVD

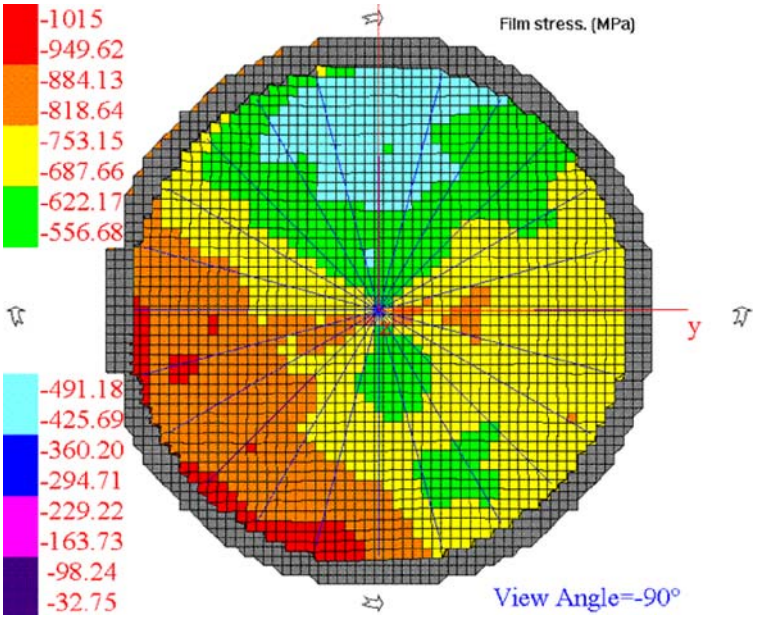


FIGURE 4.3. Residual stress map of a TiW_xN_y film 1 μm thick on a 150 mm diameter GaAs wafer.

polySi films have a very small grain size, well below 1 μm. Grain sizes of 1 μm in PVD metal films are commonplace; the effects on strength will be discussed below. In fine-grained films the volume of material in and near grain boundaries has been considered to be large enough to influence the overall behavior of the film, including density and elastic properties. Machlin [3] gives an extended treatment of the structure zone model, which relates the grain structure to the film growth conditions, particularly the substrate temperature. Figure 4.4, from [43], presents an overview that applies to metal films.

It is important to measure thin film grain size, since it affects the mechanical properties, specifically yield stress. Grain boundary scattering of electrons in very fine metal lines is now considered to have a serious deleterious effect on the electrical conductivity [44]. In the case of a nanocrystalline columnar grain Cu film, the in-plane grain size can be measured by means of atomic force microscopy (AFM), where grains can be resolved on the surface. The out of plane grain size is typically equal to the film thickness for films with thickness of a few hundred nanometers. Measurements from the AFM section analysis provide the average grain size. One of the problems with measurements of grain size of electroplated Cu films is that annealing causes grain coalescence through the film thickness, but not necessarily surface reconstruction that would replicate the new bigger grain size. Focused ion beam imaging (FIB) is a more suitable technique than AFM for thin film grain size measurement, because it directly samples the lattice, not the surface topography. It is similar to scanning electron microscopy (SEM), but instead of the electron beam, a focused ion beam of gallium is used to raster along the sample surface. The image is constructed by collecting secondary ions or electrons, which produce a contrast according to the grain orientation. The FIB can also be used to clean oxide from the surface by sputtering. A FIB image of a 2 μm thick electroplated Cu film is displayed in Figure 4.5. The sample is tilted 45° to the ion beam, so all grains appear elongated in the horizontal direction. Grain size can be directly measured from such an image, where much of the intragranular contrast

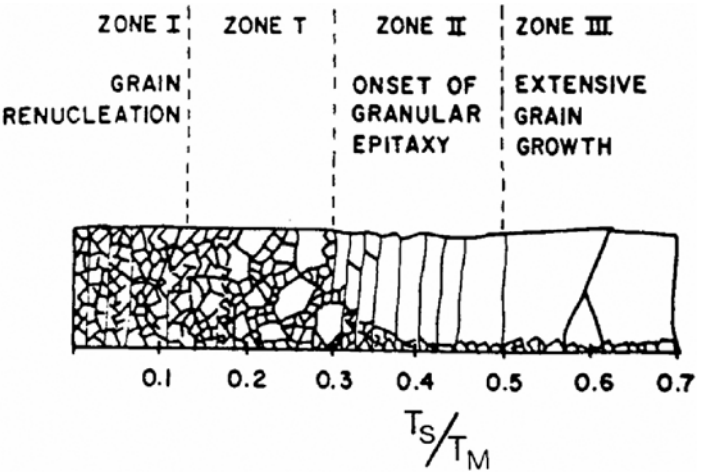


FIGURE 4.4. Effect of substrate temperature on the grain structure of a metal film. T_s/T_m is the ratio of the substrate temperature to the melting temperature of the film material in bulk form [43]. Figure reprinted with permission.

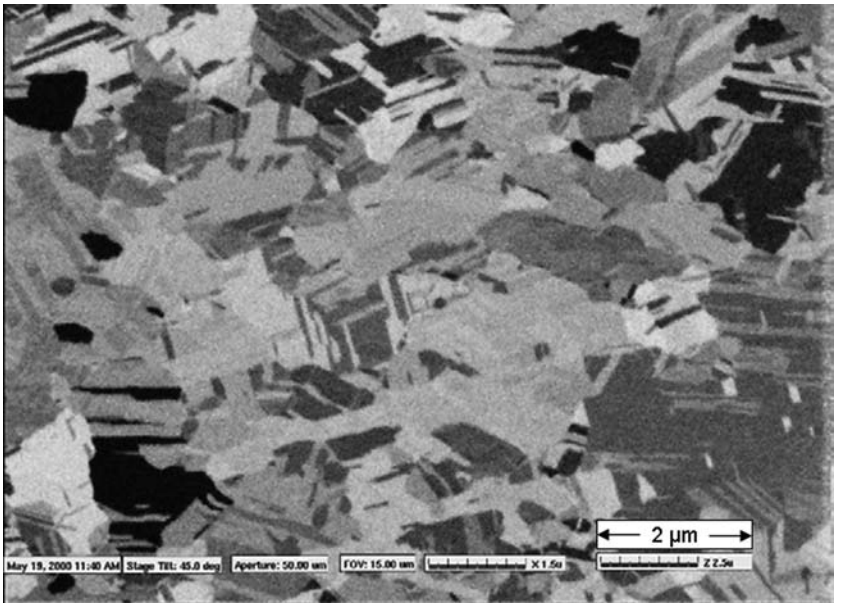


FIGURE 4.5. Focused ion beam image of a 2 μm thick electroplated Cu film (45° tilt).

is due to Cu twinning. While more time consuming, especially because of the specimen preparation required, TEM is the definitive technique for grain size measurements. The new SEM technique of electron back scatter diffraction (EBSD), combined with precise cameras and special software, is a powerful technique for crystallographic measurements in films with clean, flat surfaces.

Crystallographic texture is the third property that arises as a result of the deposition process. X-ray diffraction is the technique typically used to measure thin film texture; the

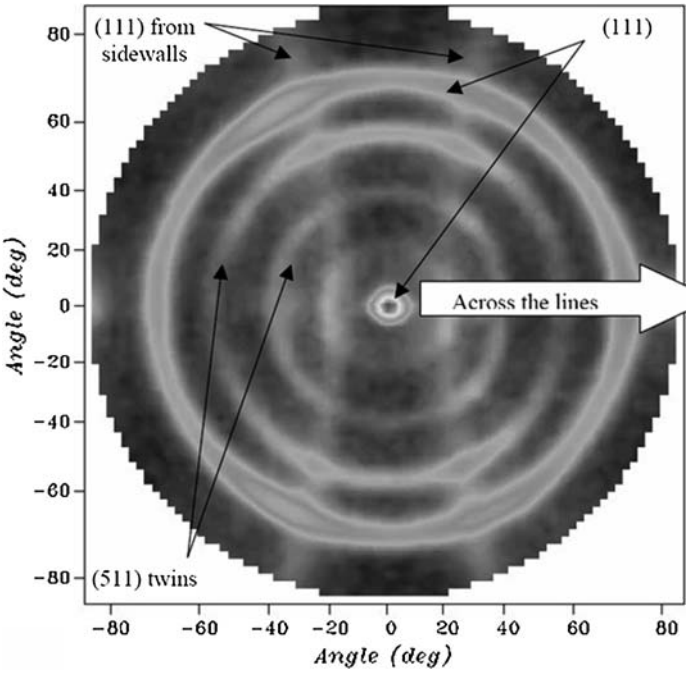


FIGURE 4.6. (111) pole figure of a 500 × 500 μm array of parallel Cu lines on a TaN underlayer. The direction transverse to the lines is indicated. The (111) fiber texture of the film on the wafer surface and the (511) twins appear as the central spot and the rings, indicated by arrows in the image. The sidewalls of the damascene lines produce (111) reflections that form the curved bands from top to bottom.

theory for the measurements is adopted from treatment of bulk materials [45]. In blanket films as-deposited in typical processes, the only special direction is the direction perpendicular to the substrate, especially since substrates are often rotated during deposition to ensure that the film properties are the same along all in-plane directions. The resulting thin-film textures are fiber textures, symmetrical about the direction perpendicular to the substrate. Aluminum films made by PVD commonly have a strong (111) texture. Typically electroplated Cu films have (111) texture, sometimes with a weak (100) component. Some film patterning process, such as lift-off and damascene, deposit metal on a patterned substrate; these patterns can introduce their own characteristic crystallographic texture. Even though the residual stress and the mechanical properties of various Cu films are very similar, the X-ray pole figure analysis technique can show the difference in electroplated Cu films on different underlayers in terms of their microstructure. Figure 4.6 shows a (111) pole figure obtained from a horizontal array of Cu interconnect lines on top of a TaN underlayer. The pole figure shows standard (111) peak intensities at 0 and 70.5°, as well as (511) twins. The fainter vertical features are the (111) intensities from the interconnect sidewalls. Electron back scatter diffraction can be used to evaluate texture, as well as grain size, because the diffraction patterns can conveniently be analyzed to give the complete crystallographic orientation of the sampled volume of material.

The crystallographic texture of electrodeposited copper films can change with time for hours or days after deposition [46]; silver appears to share this behavior. This texture evolution is just one among several manifestations of the change in microstructure of elec-

trodeposited copper films that can occur at room temperature—certain grains can also grow by recrystallization. Annealing accelerates the effects.

Zhang et al. [44] report a careful study of the electrical resistivity of electrodeposited copper. They conclude that the effects of grain boundaries are significant in raising the electrical resistivity of their films. For relatively thick films, they report a combined resistivity effect of 0.08 microohm·cm for impurities and grain boundaries. This is lower than a value reported over 10 years earlier for PVD films, discussed below. This study [44] is relevant to a current "hot topic," the increase of electrical resistivity with decreasing film thickness. Zhang et al. observe the resistivity increase with decreasing thickness, but attribute the greater part of the effect to the smaller grain sizes in the thinner films. The effect of line width on electrical resistivity was discussed by Hanaoka et al. [47] and by Josell et al. [48].

The physics behind these resistivity phenomena is based on the electron mean free path. The mean free path in pure bulk copper at room temperature is limited by electron-phonon scattering. The existence and significance of electron-phonon scattering can be evaluated from measurements of resistivity as a function of temperature. As the temperature is lowered, the phonon density decreases, the mean free path increases, and the resistivity decreases dramatically, especially in very pure materials. The residual resistivity ratio (RRR), which is the resistivity at room temperature divided by the resistivity at 4 K (−269°C), can reach values in the thousands for bulk specimens of very pure copper. Values of the RRR for thin films are much lower.

As linewidths reach the room-temperature electron mean free path in copper, the effective resistivity increases because electrons scatter from the sidewalls, top, and bottom of the line. However, electrons can scatter in different ways. Diffuse scattering, where the final velocity is randomly distributed, has a large effect, whereas specular scattering, where the electron retains its momentum in the plane of the boundary, has no effect. So a need for control of the roughness of the edges of narrow interconnect lines (line edge roughness) has been hypothesized; however experimental achievement of surfaces smooth enough to reduce diffuse scattering of electrons has not been reported. This effect is taken seriously, to the extent that novel alternatives to copper are being actively discussed, such as carbon nanotubes, which may offer "ballistic conduction," where the electron flies unimpeded down the bore of the nanotube.

4.3.3. Impurities

The rain of metal atoms impinging on and sticking to a substrate during film deposition as described above in Section 4.3.1.1 is accompanied by a rain of impurity gases, even in PVD systems with very good vacuum. Some of the gas atoms remain in the films. Similarly, impurities from an electrodeposition solution are deposited along with the intended metal. These components may be deleterious to the electrical conductivity of a film, but they are generally carefully controlled only during epitaxial growth of multilayer semiconductor films.

The best current value for the electrical resistivity of pure bulk copper at room temperature (300 K) was given by Schuster et al. as 1.72 microohm·cm [49]. These authors also give a detailed discussion of the temperature dependence of the resistivity of copper. Johnson [50] studied the electrical resistivity of copper and nickel thin films produced by thermal evaporation under a range of deposition conditions relevant to electronic interconnects. He noted that an impurity concentration of around 1% was typical for his PVD films; this produced a resistivity increase of about 0.2 microohm·cm. Johnson examined other possible sources of electrical resistivity in his thin films and concluded that

the next largest contribution, about 0.1 microohm cm, was from intragranular dislocations (discussed below), and that the contributions of the remaining possible sources were insignificant. Johnson concluded that PVD copper films made under optimized deposition conditions can be expected to have electrical resistivities 18% higher than the bulk copper value. He found the situation for nickel to be different in some aspects, with a much worse achievable electrical resistivity.

4.3.4. Dislocations

A dislocation is a line defect in a crystal lattice [51], familiar from undergraduate materials science. Dislocations are a concern in all the crystalline materials in electronic devices. Silicon wafers are carefully handled to make sure they are practically "dislocation free," because dislocations would offer paths for current leakage and fast diffusion, as well as likely sites for impurities. The effects of dislocations in other crystalline substrates are similarly undesirable. Dislocations may play a slightly more constructive role as strain-absorbing structures at sites of lattice-parameter mismatch, for example, in heteroepitaxy or around quantum dots. The intentional use of dislocations in such structures is in its infancy at present, and its future is uncertain. The typical approach today is to seek to control creation of dislocations through the use of graded buffer layers.

Metal films made by PVD contain a large dislocation density compared to familiar bulk materials. This is an issue because dislocations, along with surfaces and grain boundaries, raise the electrical resistivity and are paths for rapid diffusion. Dislocations exacerbate the failure modes of diffusion, which can create unwanted conduction paths, and electromigration, which can cause open circuits. Figure 4.7 [52] shows dislocations in an

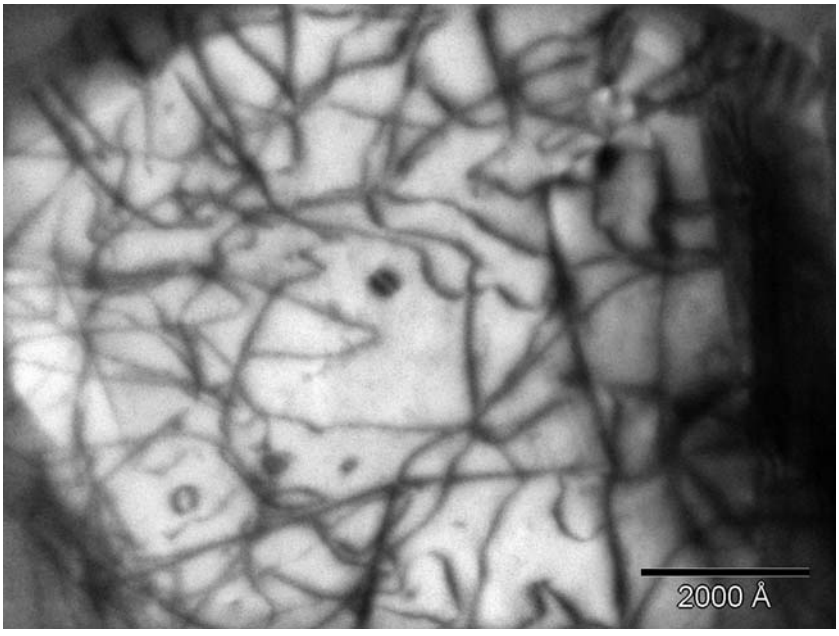


FIGURE 4.7. TEM micrograph of a thin film of aluminum, as grown. The dark lines are dislocations, which produce contrast because the crystal lattice is disturbed at the dislocation core. The dislocation density is about 2×10^9 cm^{-2}.

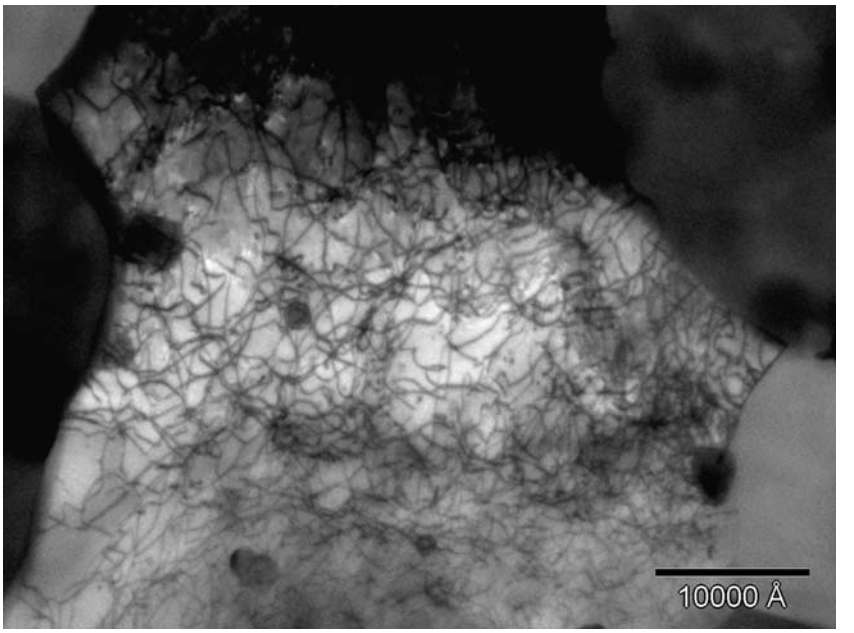


FIGURE 4.8. TEM micrograph of a thin film of microtensile specimen of aluminum, strained to failure in tension. The dark lines are dislocations. The dislocation density is about 10^{11} cm^{-2}.

aluminum film deposited as contact metal in a commercial CMOS process. The dislocation line density is about 2×10^9 cm/cm^3. Dislocations were discovered as the mechanism for plastic strain in metals, and mobile dislocations are the carriers of plastic strain in sufficiently thick metal films, as can be discovered by TEM examination of microtensile and nanoindentation specimens. Figure 4.8 [52] shows dislocations in a microtensile specimen of the same type as shown in the previous figure, that had been strained to failure. The dislocation density has increased by more than an order of magnitude.

Metal films below a minimum thickness in the range of tens to hundreds of nanometers tend to lack mobile dislocations. This conclusion has been reached both in analytical predictions and in experiments [53–55]. Figure 4.9 [16,52] shows a TEM micrograph of a tensile specimen in the thin region adjacent to the failure site. This specimen, again the same material as in the previous two figures, exhibited a chisel-point fracture. The material stretched by thinning, and then parted. Figure 4.9 shows that dislocations are absent from the region near the fracture, and reappear in the thicker region further away. Figure 4.9 also shows another feature, prismatic dislocation loops, which are immobile and therefore are retained in the failed specimen. These loops are the residue of clusters of vacancies; vacancies agglomerate as a disk, the flat surfaces bond together, and a dislocation is created as the boundary. The prismatic loops are the small, round structures that resemble coffee beans.

4.3.5. Electromigration and Voiding

Electromigration is the phenomenon of mass transport of metal atoms being driven out of a conductor by the flux of electrons impinging on each atom. Figure 4.10 [56] shows a void and a hillock created by electromigration in an aluminum line. Electromigration

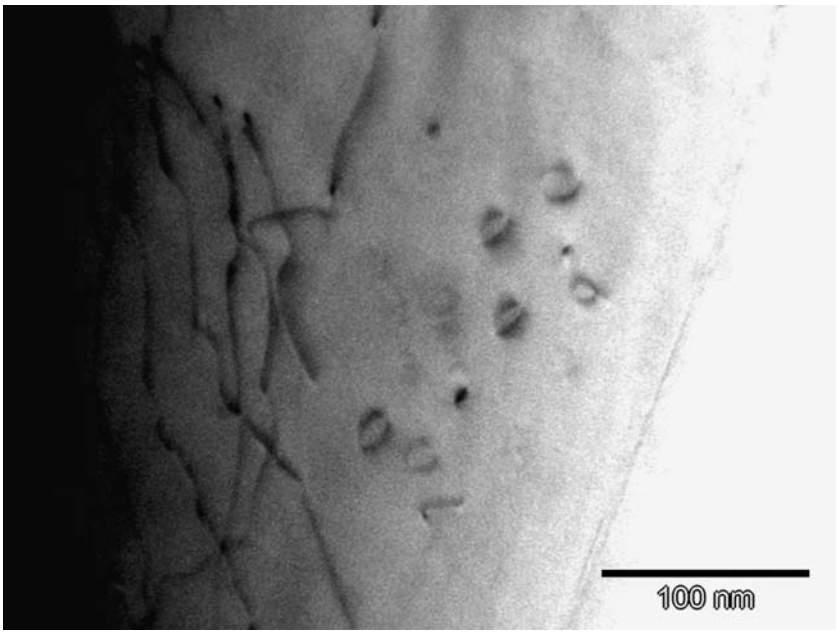


FIGURE 4.9. TEM micrograph of the failed edge of an aluminum thin-film microtensile specimen. The linear features on the left side of the image are dislocations; the thickness decreases from left to right in this image. The round features on the right are prismatic dislocation loops.

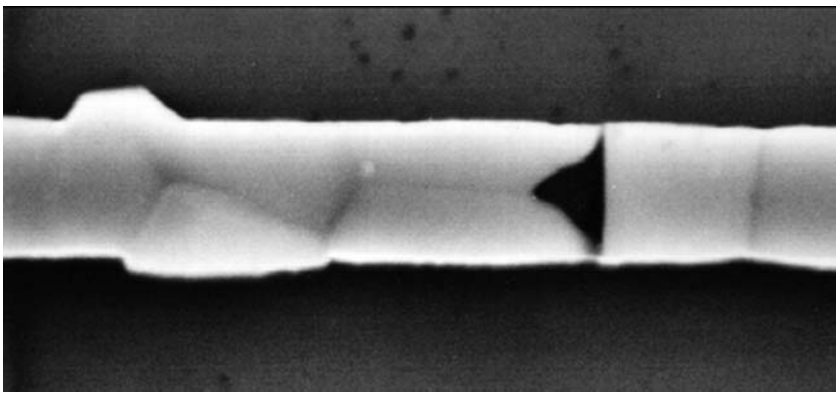


FIGURE 4.10. Hillock and void produced by electromigration in an aluminum line [56]. Reprinted by permission.

has been recognized as a reliability issue in integrated microelectronics devices for many years [57]. Tamura et al. [58] report on the stresses created by electromigration in copper. Stress and temperature affect electromigration. The stress effect operates to create a conductor length effect, as described by the Blech formula [5]. Conductive paths that are shorter than the "Blech length" do not fail by electromigration because the stresses created by electromigration itself, by moving atoms into a "crowded" region, are sufficient to halt the process. Standardized tests for electromigration have been developed; a thin-film con-

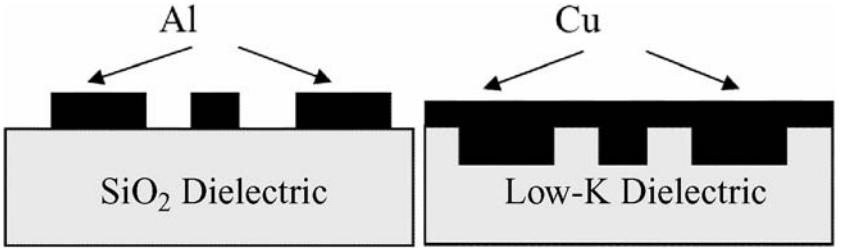


FIGURE 4.11. Schematics of Al/SiO_2 vs. Cu/low-*k* (damascene) interconnect structures.

ductor can withstand currents of 10 MA/cm^2 for short periods of time. Use of design rules that limit current densities to 0.1 MA/cm^2 controls electromigration in practical devices.

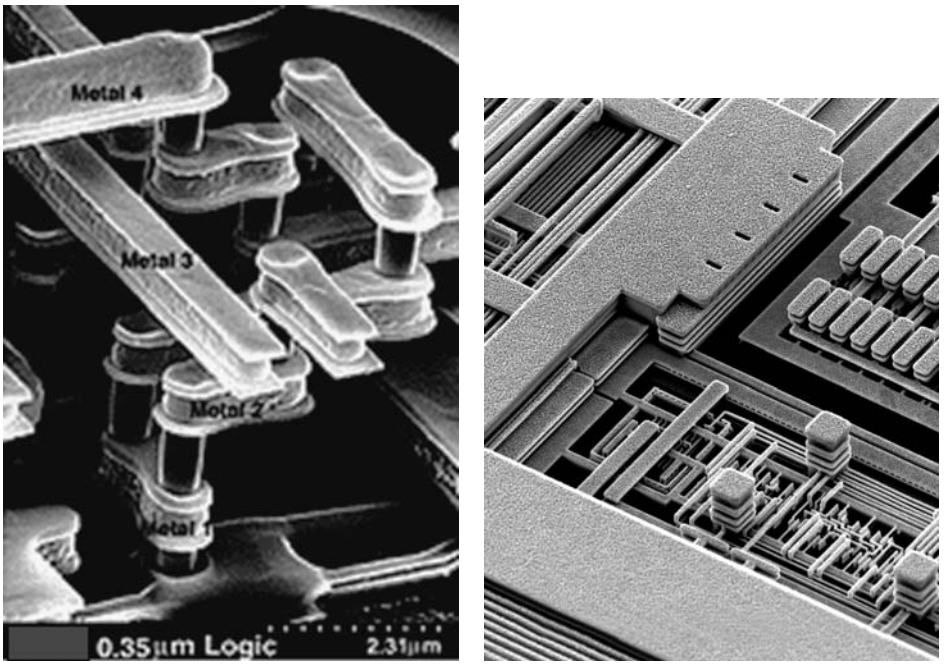
Thermal stress-induced voiding [59] is another diffusion-based reliability issue. It arises from high tensile hydrostatic stresses induced by encapsulation of metal lines in rigid dielectrics. Void growth is accelerated by increasing temperature and increasing stress.

4.3.6. Structural Considerations

The classic single-layer line-on-substrate design could not provide the interconnect density needed for large-scale integrated circuits, so multilevel interconnect structures became necessary decades ago. Interlayer dielectric films were applied over aluminum lines, via holes were patterned and tungsten vias placed, and the next layer of interconnect was deposited. However, for the new copper interconnect structures, a different design has been adopted, the damascene approach, Figure 4.11. The trenches are patterned first, and then Cu is electroplated on top of the sputter deposited barrier and Cu seed layers. Excessive Cu is removed by the chemical mechanical planarization (CMP) process. This is a combination of chemical and mechanical polishing that adds a new set of challenges to the mechanical integrity of the interconnect structure. Chemical mechanical planarization can also be used with Al technology; its widespread use with Cu layers allows for a larger number of interconnect layers, currently around 10, in the device, thus higher complexity and level of integration. Figure 4.12 shows perspective images of interconnect structures [60].

4.3.7. Need for Mechanical Characterization

Mechanical characterization is useful in design for reliability, material acceptance, and quality control. Challenges to the mechanical reliability of thin films in microelectronic devices can arise during manufacture and in service, from a variety of sources; the main source of the stress is usually differential thermal expansion. Additional factors such as diffusion (promoted by high temperature), high voltage, and electrochemical corrosion (promoted by humidity) can significantly accelerate the failure process. Microelectronic devices contain a variety of different materials, with different thermal expansion coefficients and different elastic properties. These materials are rigidly joined with each other, often as alternating layers atop a stiff silicon substrate. The chips are heated from ambient to operating temperature, e.g., around 90°C for a typical desktop computer, at least once for every on-off cycle. Changes in the mode of use, e.g., beginning and ending a numerically intensive calculation, can add more cycles. Each temperature change produces thermal stresses caused by differential thermal expansion among the materials of the chip. High stresses can occur in encapsulated lines and in the enclosing dielectric. Since the



(a) (b)

FIGURE 4.12. Aluminum (a) and copper (b) interconnect structures [60]. (b) Photo courtesy of International Business Machines Corporation. Unauthorized use not permitted.

processing speed increases with the current into and out of the transistors, high current levels are often favored. But, the higher the current, the higher the temperature, so commercial pressures favor devices capable of operation at high temperatures.

4.3.8. Properties of Interest

Here we focus on the set of properties needed in considerations of manufacturability and reliability of thin film structures, and specifically, on their measurement. Other properties, particularly electrical and optical properties related to the intended function of a film, are also of wide interest. Many of these can be measured on blanket films, or with straightforward test structures. Some properties of films, for example, the thermal expansion coefficient, are not measured; where data are needed, bulk properties are used. Mass density is becoming more important with the growing use of porous low-k (low dielectric constant) dielectric layers. However, the mass density of thin film layers is rarely measured directly, as discussed above.

Analytical or numerical simulation of the mechanical behavior of thin film layers relies on a handful of material properties, including Young's modulus or, in special cases, the anisotropic elastic constants, thermal expansion coefficient, yield strength, ultimate tensile strength, elongation to failure, residual stress, fatigue resistance in the form of stress-N or strain-N curves (where N is number of cycles), creep resistance, and adhesion. The differences from the design set for macroscopic structures are the importance of thermal expansion as a source of stress; the ubiquity of adhesion relative to joining; and the shift of fracture toughness issues from bulk materials such as welds to bimaterial interfaces. Thin

films in microelectronics devices have microstructures much different from those of traditional bulk materials. For metals, conventional wrought and annealed materials often have grain sizes in the tens of micrometers, while for films, sub-micrometer grain sizes are the rule. For polymers, the thin layers applied by spin- or dip-coating are likely to be different from extruded or cast bulk material. Polycrystalline silicon is not produced in bulk. Because material properties depend strongly on microstructure, characterization of thin films produced by the same deposition methods used in manufacturing of actual devices, and with specimens of the same general size scale as the films in actual devices, is necessary.

How can the values of these properties be obtained for a given material? Use of conventional mechanical testing instrumentation, such as the tensile test machine, for measurements on thin films is simply not appropriate. The films cannot be removed from their substrates for gripping, and the force sensitivity needed is orders of magnitude finer. The need for experimental techniques adapted specially for testing thin films has been recognized since at least the 1950's. The challenge of accurately measuring mechanical properties has stimulated the development of specimen designs and fabrication routes, and special test apparatus. The next section describes the most successful techniques for measuring mechanical properties of thin films.

4.4. METHODS FOR MECHANICAL CHARACTERIZATION OF THIN FILMS

The main methods in current use for mechanical characterization of thin films include microtensile testing (MT) and instrumented indentation, also referred to as nanoindentation (NI). Other methods in wide use include wafer curvature, the pressurized bulge test, and a variety of tests of the adhesion of a film to its substrate.

4.4.1. Microtensile Testing

Tensile testing is the standard means of obtaining mechanical properties of structural metals. Because the stress field is uniform throughout the gage section, the Young's modulus, yield strength, and ultimate tensile strength can be obtained from an accurate force-displacement record. So it was natural to apply this time-tested method to thin films. Early attempts to pull thin films in conventional testing machines used specimens lifted from the substrate; researchers encountered problems in placing the specimen on the grips without excessive wrinkling, and depended on special separation layers beneath the specimen film, such as water soluble sodium chloride. Early tests of metal films revealed the main phenomena still seen today: high strength, and low elongation to failure [61]. There is at present no standard test method for microtensile testing of thin films; individual investigators adapt the standard methods for bulk metal specimens to fit their specific specimen geometry. Standardization is hindered by the multitude of specimen sizes and designs that are in use, which has resulted from the difficulty of fabricating microtensile specimens.

The problems with the early methods led to improved procedures, and progress continues. It became evident that since films in actual devices are always produced on substrates, the use of the substrate to support the thin film specimen is appropriate. But the substrate is always much more massive than the film, so it must be removed at least from beneath the gage section of the specimen. Ding et al. [62] reported the use of a silicon frame design for testing doped silicon. The first realization of this scheme for metal films was the silicon frame tensile specimen [63]. Bulk micromachining of MEMS devices had

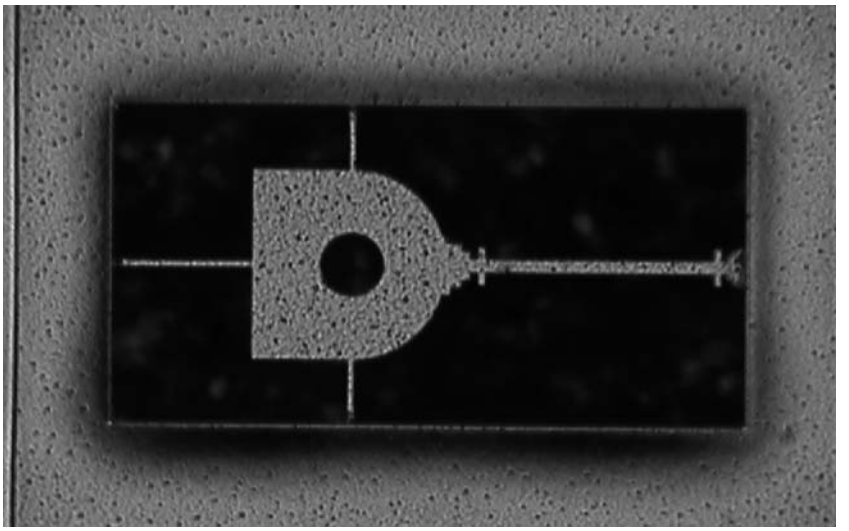


FIGURE 4.13. Microtensile specimen of aluminum, fabricated through the MOSIS process. The loading tab, with its 50 μm hole, is to the right. The gage section, with "ears" for use in digital image correlation for displacement measurement, is to the left. The silicon substrate has been etched away to a depth of 60 μm or more. The three slender aluminum lines connecting the field to the loading tab are tethers which are manually cut just before testing.

been developed by this time, demonstrating the concept of etching away a selected portion of the substrate to form a useful device. To produce the silicon frame tensile specimen, photolithographic patterning is used to form a straight and relatively narrow gage section with larger grip sections on a silicon frame. The substrate beneath the gage section is removed by a suitable etchant. The silicon frame, carrying its tensile specimen of a thin film, is mounted on a suitable test device capable of supplying force and displacement [64]. The silicon frame is cut, while leaving the specimen undamaged. This step has been accomplished manually with a dental drill, using a temporary clamp to hold the specimen in place, and by the use of a cutting wheel mounted on a movable stage [65].

All the MT techniques include measurements of force and displacement. The force is measured by a load cell, either commercial or custom-built. For the specimen introduced above, the force might amount to 0.1 N; commercial load cells with this range are available. Displacement has been measured by interferometric techniques such as electron speckle pattern interferometry (ESPI), for example as in [66], or by diffraction from markers placed on the specimen surface [67].

The specimen fabrication challenge with these techniques is the chemical selectivity required to etch through hundreds of micrometers of silicon without damaging the metal specimen. Aqueous hydrazine has been used, but this material is hazardous. Another disadvantage is the large width of the gage section, 100 μm or more, compared to the linewidths used in interconnect, and also compared to a typical film thicknesses of 1 μm.

A new generation of smaller-scale specimens and complementary test techniques has been developed. In this version, the specimen width is around 10 μm and the gage length is around 200 μm, while the thickness remains near 1 μm, Figure 4.13 [68]. The surface micromachining concept is used; the substrate is removed to a depth of around 100 μm beneath the specimen, using xenon difluoride. This etchant is less hazardous than hydrazine,

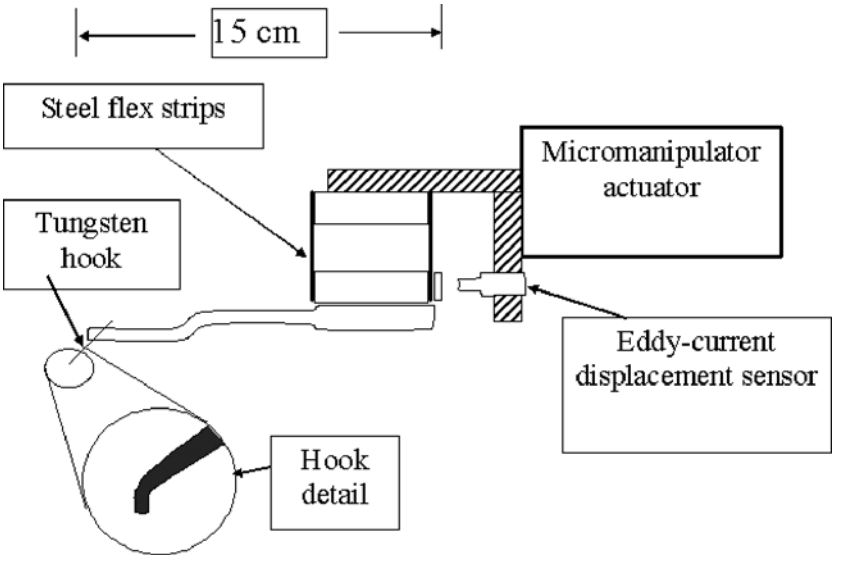


FIGURE 4.14. Tungsten "hook" carried by instrumented micromanipulator to load tensile specimen and measure the force.

and is very selective for silicon masked by SiO_2, aluminum, copper, etc. The control of edges; the elimination of the need to manually lift a film off a substrate, often with the aid of a water bath, and place it on tensile grips; and the use of specimens with aspect ratios (width to thickness) of 10 or less, have significantly mitigated the presence and effects of wrinkles and edge nonuniformity pointed out by Brotzen [69] in his review, discussed further below. Young's modulus can usually be measured in these specimens, but Poisson's ratio has been measured only by special techniques on relatively large specimens [70,71], because the transverse displacements are so small on a specimen only a few micrometers wide. In an early version of this test, the specimen was loaded by engaging a tungsten probe tip, 50 μm in diameter, to a hole in the loading tab, Figure 4.14. A recent variant of this technique is the membrane deflection tensile test, applied to a series of face-centered-cubic (FCC) metals by Espinosa et al. [72], Figure 4.15.

A new advance is the cofabrication of a specimen and a protective frame that includes a force sensor, Figure 4.16 [54]. This specimen is suitable for use inside a transmission electron microscope (TEM).

A recent round robin showed reasonable agreement among several laboratories in the strength of polySi, although most labs required their own unique specimen geometry. The different geometries were produced on the same MEMS chip [30]. The strength values obtained for polySi were impressively high, of the order of 1/30 of the polycrystalline Young's modulus, which is the usual estimate of the theoretical strength of a solid.

4.4.2. Instrumented Indentation

The nanoindentation test similar to the conventional hardness test, but is performed on a much smaller scale by use of specialized equipment—a nanoindenter [73]. The force, P, required to press a sharp diamond indenter into tested material is continuously recorded as a function of the indentation depth, h, as indicated schematically in Figure 4.17. The

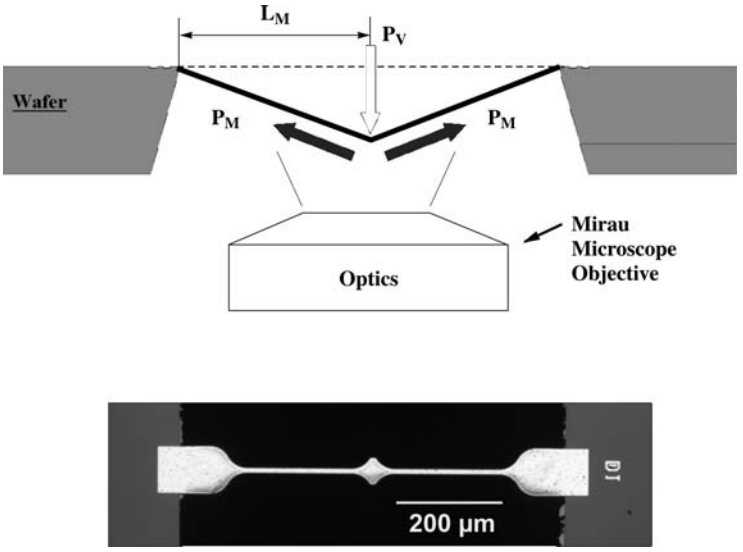


FIGURE 4.15. Setup for the membrane deflection tensile test [72].

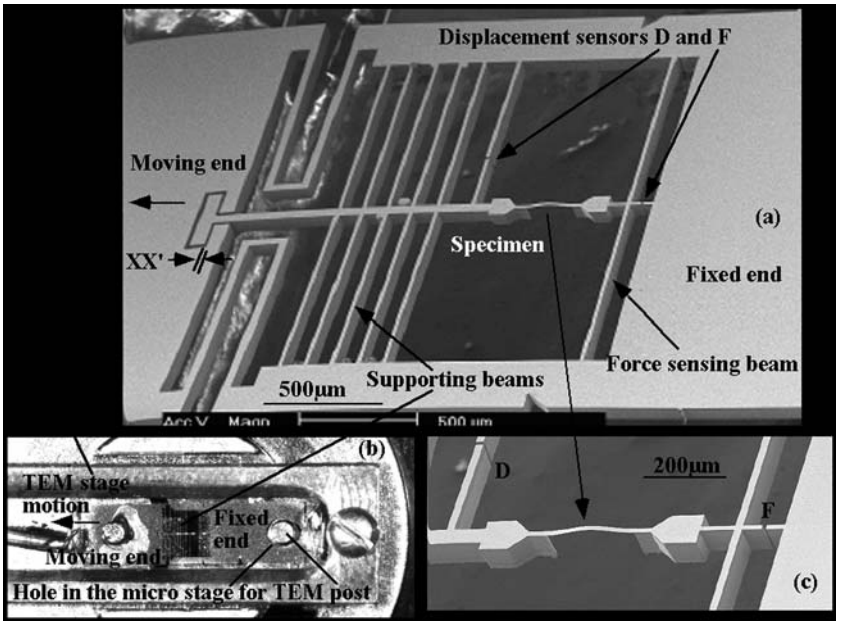


FIGURE 4.16. Tensile specimen assembly including aluminum tensile specimen and MEMS support assembly and force gage for use in the TEM [54].

actuation mechanism can be based either on electromagnetic or electrostatic application of force. Since the depth resolution is on the order of angstroms, it is possible to usefully indent even very thin (∼100 nm) films. The nanoindentation load-displacement curve, similar to one shown in Figure 4.18, provides a "mechanical fingerprint" of the material's response to contact deformation. Elastic modulus and hardness are the two parameters that can be

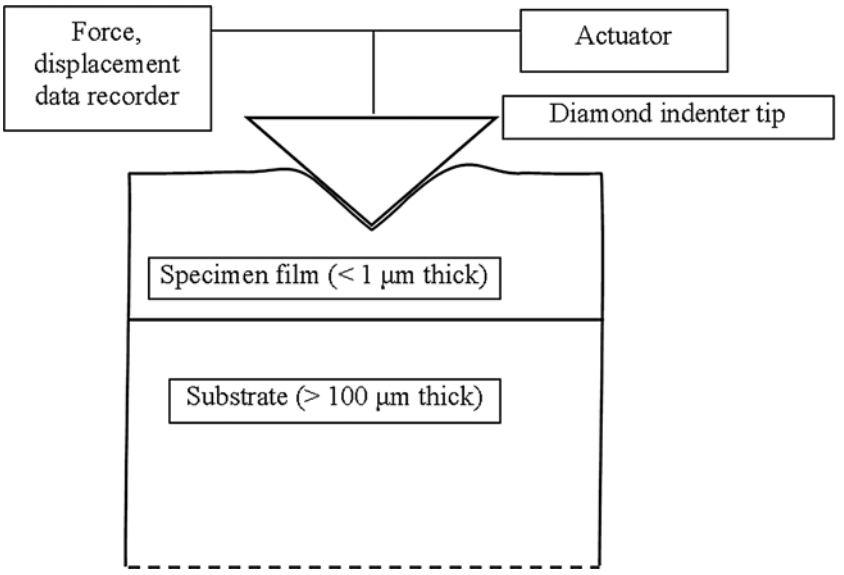


FIGURE 4.17. Schematic of instrumented instrumentation apparatus.

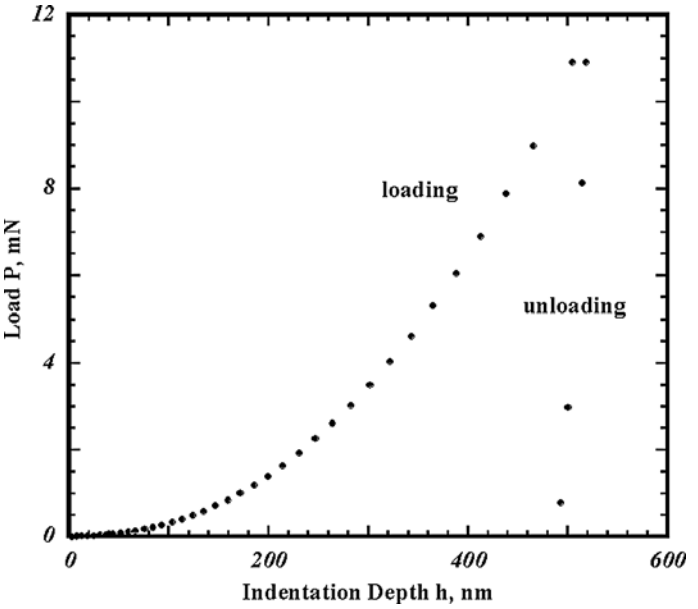


FIGURE 4.18. Load-displacement record from an instrumented indentation test of an electrodeposited Cu film 1 μm thick.

readily extracted from the nanoindentation load-displacement curve. Elastic property measurements by nanoindentation were originally proposed by Loubet et al. [74]. Later, Doerner and Nix [75] suggested that a linear fit to the upper 1/3 of the unloading portion of

the indentation curve could be used to determine film stiffness, $S = dP/dh$, from which the reduced elastic modulus, E_r, could be found as

$$E_r = S\frac{\sqrt{\pi}}{2\sqrt{A}}, \tag{4.3}$$

where A is the contact area and E_r is a combined elastic property of the film and indenter materials. Since the indenter material itself has finite elastic constants, its deformation contributes to the measured displacement. The reduced modulus E_r, is given by:

$$\frac{1}{E_r} = \frac{1-\nu_i^2}{E_i} + \frac{1-\nu_f^2}{E_f}. \tag{4.4}$$

In this equation E is the elastic modulus, ν is the Poisson's ratio, and the subscripts f and i refer to the film and the indenter materials respectively. A more elaborate power law fit to the unloading portion of the load–displacement curve was suggested by Oliver and Pharr [76], and is widely known as the Oliver and Pharr method.

Hardness H, a material's resistance to plastic deformation, is defined as

$$H = \frac{P_{\max}}{A}, \tag{4.5}$$

where A is the projected area of contact (a function of the indentation depth) at the maximum load $P_{\max}$.

The expressions for both elastic modulus and hardness contain the contact area, which is correlated to the indentation depth both theoretically, through the known geometry of the indenter, and experimentally, by indenting a material with known elastic modulus. This tip calibration procedure consists of indenting a standard material (often fused quartz or single crystal Al) to various maximum indentation depths. Since the contact area is determined from tip calibration, various tip geometries can be used, with the most common being the Berkovich three-sided pyramid geometry. From the manufacturing standpoint, a three-sided pyramid always ends as a point, and the tip radius can be as sharp as 10–50 nm. Other geometries are also used, and include Vickers (a standardized square pyramid), cube corner, conical and wedge indenters. The unloading slope, dP/dh, is related to the tip geometry as

$$\frac{dP}{dh} = 2\beta\sqrt{\frac{A}{\pi}}E_r, \tag{4.6}$$

where h is the indentation depth, and β is a constant, near unity for a given tip geometry. King et al. calculated β values for different tip geometries using finite element analysis [77].

One should note that the tip calibration does not account for either plastic pile-up or sink-in of both the standard and the specimen materials, which causes inaccuracies in indentation depth and contact area determination. In addition, the total test compliance, i.e., the inverse of stiffness, is affected by the indentation contact. One should also account for the test frame compliance, C_f, as it offsets the total test compliance:

$$C_{total} = C_f + \frac{\sqrt{\pi}}{2\sqrt{A}E_r}. \tag{4.7}$$

In order to avoid substrate effects on the measured mechanical properties, a film should be indented only up to a certain percentage of its thickness (up to 10–20%). There is also an influence of the residual stress and substrate effects that are hard to account for in the analysis [78,79]. Indentation curve analysis has been extended in the past few years with new FEM-based models being developed [80,81].

A more powerful variant of the indentation technique uses a "continuous stiffness" approach (continuous stiffness method, CSM, in which the contact stiffness is measured by multiple small (2 nm) loading and unloading cycles (0.5 to 300 Hz) superimposed on the load-displacement curve. The original idea was proposed by Pethica and Oliver [82], where they used a force controlled indenter, superimposing a small sinusoidal component on the quasistatic indentation load. This resulted in the indenter displacement phase shift with respect to the excitation force:

$$\tan(\phi) = \frac{\omega D}{\dfrac{SK_f}{S+K_f} + K_s - m\omega^2}, \tag{4.8}$$

where ϕ is the phase angle, ω is the frequency of oscillation, m is the mass of the indenter tip, Figure 4.19, $K_f = 1/C_f$ is the stiffness of the load frame, K_s is the spring constant of the indenter support springs, and D is the damping coefficient. The contact stiffness can be calculated from [76]

$$\frac{\Delta P}{h(\omega)} = \sqrt{\left(\frac{SK_f}{S+K_f} + K_s - m\omega^2\right)^2 + \omega^2 D^2}, \tag{4.9}$$

where ΔP is the magnitude of the force oscillation, and $h(\omega)$ is the magnitude of the displacement oscillation.

Both elastic modulus and hardness can be calculated for the unloading portion of each tip oscillation during indentation. Figure 4.20 shows elastic modulus and hardness obtained with the CSM method for a 1 μm thick electroplated Cu film on a Si substrate.

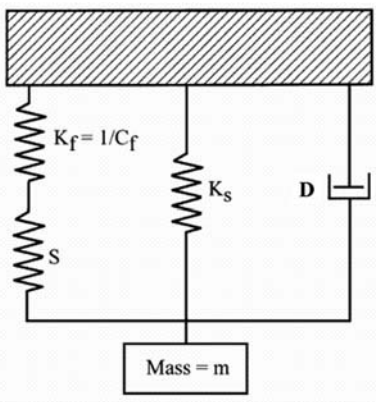


FIGURE 4.19. Schematic of the indentation dynamic model (adapted from [23]). K_f is the stiffness of the instrument frame, S is the stiffness of the tip-specimen contact, K_s is the stiffness of the springs that support the tip, D is the damping coefficient, and m is the mass of the tip assembly.

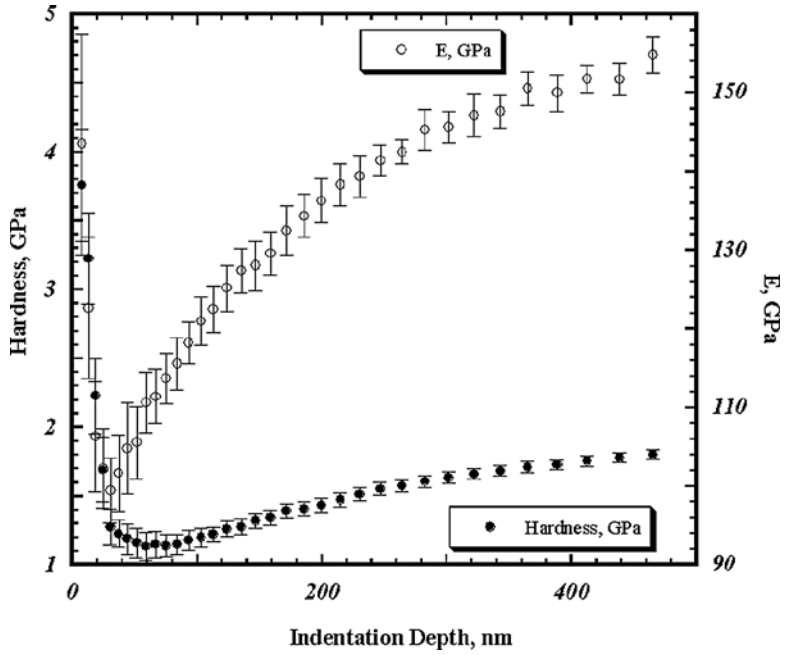


FIGURE 4.20. Elastic modulus and hardness for a 1 μm thick electroplated Cu film on Si substrate measured using continuous stiffness method.

Here, the substrate effects can be seen at a depth of 50 nm for the modulus, and 100 nm for the hardness. On the one hand, this is an advantage, as composite properties of the film and the substrate are measured. On the other hand, extrapolating "true" film mechanical properties from such data can be challenging. Here, the free-standing film tensile techniques are advantageous. A comprehensive review of the method applied for magnetic storage and MEMS was reported by Li and Bhushan [83].

4.4.3. Other Techniques

4.4.3.1. Wafer Curvature The basic principle of the wafer curvature technique is that differential thermal expansion between a specimen film and a silicon substrate produce measurable curvature of the substrate (the wafer); the curvature is related directly to the product of stress and thickness in the film, through the Stoney equation, discussed above. This phenomenon is used in evaluating and adjusting film deposition procedures, to measure residual stress in the deposited films. High values of residual stress, especially tension, may make a film less resistant to delamination from the substrate.

Wafer curvature measurement was adapted for characterization of mechanical behavior by Nix [84]. The substrate with its film is placed in a furnace equipped for measurement of the substrate curvature. The temperature is cycled, while the curvature is recorded. Given the film thickness, the film stress can be plotted against temperature. The accessible range of temperature is limited only by the eventual breakdown of the specimen film by melting or chemical reaction. The stress depends in turn on the difference in thermal expansion between the specimen film and the substrate, and the elastic constants of the specimen film. Deviations from linear behavior with temperature imply plastic deformation of the specimen film; the nature of this deformation is confirmed by the hysteresis loop observed at

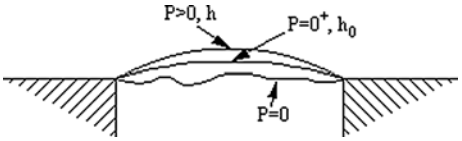


FIGURE 4.21. Schematic diagram of the bulge test specimen, showing a slack film loaded with zero, infinitesimal, and finite pressure.

least on the first temperature cycle. The advantages of the wafer curvature technique include the simplicity (in principle) of both the experimental technique and the specimen, which is a film on the same substrate used in actual manufactured products, without the necessity of selectively removing the substrate beneath the film. Analysis of the results does not require knowledge of the elastic properties of the deposited film, only those of the substrate. The disadvantages are that (1) the ultimate tensile strength and elongation to failure cannot be measured, and that (2) only certain combinations of Young's modulus, flow stress, and temperature are accessible. This technique has been very successful in providing insight and data on deformation mechanisms particularly in aluminum films [84].

4.4.3.2. Pressurized Bulge Testing The name of the pressurized bulge test is descriptive: by etching away the substrate beneath a region of the specimen film, the film can be exposed to stress by a pressurized fluid introduced beneath the substrate. The mechanics of a pressurized membrane can be used to analyze the observed behavior. The shape of the pressurized region is chosen intentionally; circular, square, and rectangular shapes have been explored. The out-of-plane deformation of the membrane can be measured by interferometry or related optical techniques. This technique has been used to explore the elasticity of thin films; care must be taken to properly characterize the initial state of the film, including the possibility of residual stress, Figure 4.21 [85,86]. It has also been used to measure the adhesion between the film and the substrate [87].

4.4.3.3. Deformed and Resonant Cantilever Micromachined cantilevers have been used as specimens in thin film properties measurements [29,88]. Photolithography can be used to define the cantilever geometry. Cantilevers can be deformed by loading with, for example, an instrumented indenter, or can be excited to resonance, to measure film elastic properties. The relationship between the mechanical stiffness or the resonant frequency and the elastic constant of the film depends sensitively on the dimensions of the cantilever [89]. The ideas of the bulge test and resonance can be combined in the resonant membrane test, which can be used to determine the product of film elastic modulus and mass per unit area. If the thickness and mass density of the film are known, the elastic modulus can be extracted.

4.4.4. Adhesion Tests

Adhesion between layers of different materials is a critical issue in microelectronic packages, and also within the chips. While the time-honored "scotch tape" adhesion test is still in use, quantitative tests, developed in recent years based on the concepts of fracture mechanics, provide material characteristics that can be compared to calculable stress- and strain-based driving forces, and are therefore suitable for use in lifetime predictions [90]. Reviews by Volinsky et al. [91] and by Lane [92] provide useful summaries. The basic idea, as in macroscale fracture mechanics, is that it is useful to quantify the conditions under which an existing crack may advance. The crack, in this case, is assumed to be a small delamination of the film from the substrate. The driving force for crack propagation

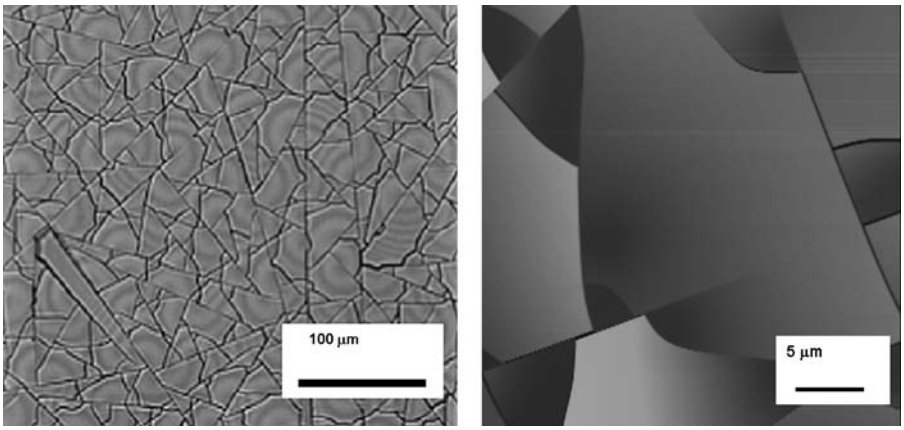


FIGURE 4.22. Optical and AFM images of cracks in low-k dielectric thin film.

is taken as the strain energy release rate, which depends on the geometry and the stress state. Both tensile and compressive stresses in thin films promote adhesion failures; a thin film in compression buckles, delaminates and spalls from the substrate when its strain energy release rate exceeds a critical value that is characteristic of the adhesion between film and substrate [93]. A general, simplified form of the strain energy release rate, G, in a stressed film, regardless of the algebraic sign of the stress, is:

$$G = Z\frac{\sigma_f^2 h}{E_f}, \tag{4.10}$$

where σ_f is the stress in the film, h is the film thickness, E_f is the modulus of elasticity, and Z is a dimensionless cracking parameter. More accurately, the energy release rate averaged over the front of advancing isolated crack is

$$G = g(\alpha, \beta)\frac{\pi(1-\nu^2)\sigma_f^2 h}{2E_f}, \tag{4.11}$$

where $g(\alpha, \beta)$ is a function of the Dundurs parameters α and β, and can be found in [94,95]. This strain energy release rate is the driving force for fracture. Film fracture or delamination is observed when the strain energy release rate exceeds the toughness of the film, G_f, or the interfacial toughness, Γ_I respectively ($G > G_f$, or $G > \Gamma_I$). One can avoid these types of failures by either reducing the film thickness, or the stress, or by increasing adhesion. Practically, the film thickness is easier to control. For a given stress level, there is a certain critical film thickness, at which failures are observed. As an example, Figure 4.22 shows through-thickness cracks in a low-k dielectric film 2 μm thick. Thinner films showed no signs of failure.

In case a film has fractured, and its residual stress and thickness are known, Equations (4.10) and (4.11) can be used as upper bound estimates for adhesion. For example, the 1 μm TiWN film on a GaAs wafer had adhesion of at least 5 J/m^2, having the maximum residual stress of 1 GPa, as seen in the stress map in Figure 4.3.

In the case of a compressed films, telephone cord delamination is commonly observed (Figure 4.23). The geometry of the buckles can be used to asses thin film adhesion.

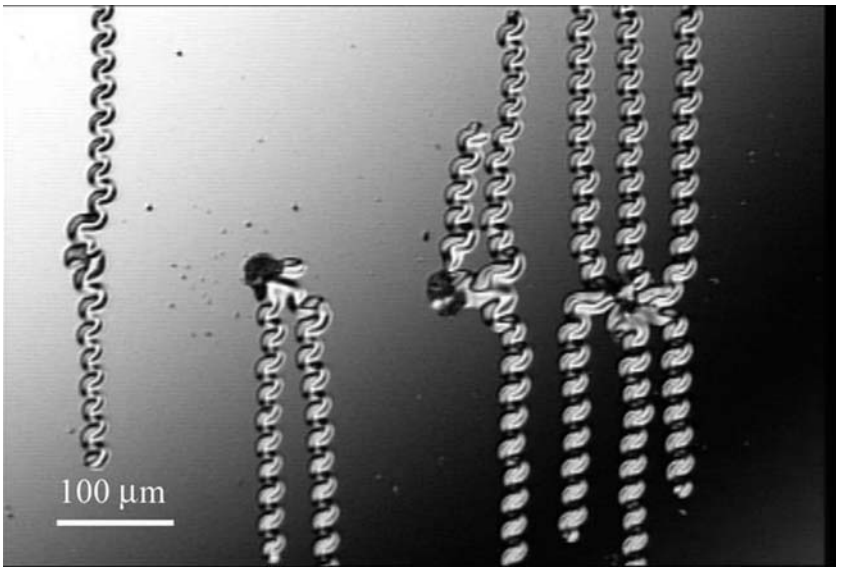


FIGURE 4.23. Telephone cord delamination in a 1 μm tungsten film.

The following analysis is based on Hutchinson's and Suo's developments for buckling-driven delamination of thin films [96]. Upon buckling, the stress in the film, σ_B, is estimated as

$$\sigma_B = \frac{\pi^2}{12}\frac{E}{(1-\nu^2)}\left(\frac{h}{b}\right)^2, \tag{4.12}$$

where h is the film thickness, b is the blister half-width, and E and ν are Young's modulus and Poisson's ratio, respectively. The buckling stress is acting in the vertical direction, perpendicular to the straight blisters shown in Figure 4.24(a). Here, the buckling stress of the W film is estimated to be 275 MPa.

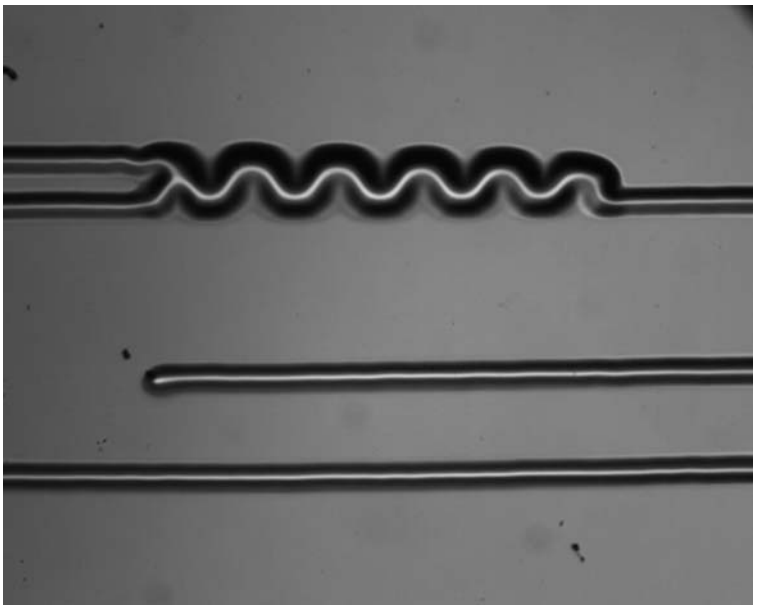
The compressive residual stress, σ_r, responsible for producing buckling delamination is

$$\sigma_r = \frac{3}{4}\sigma_B\left(\frac{\delta^2}{h^2}+1\right), \tag{4.13}$$

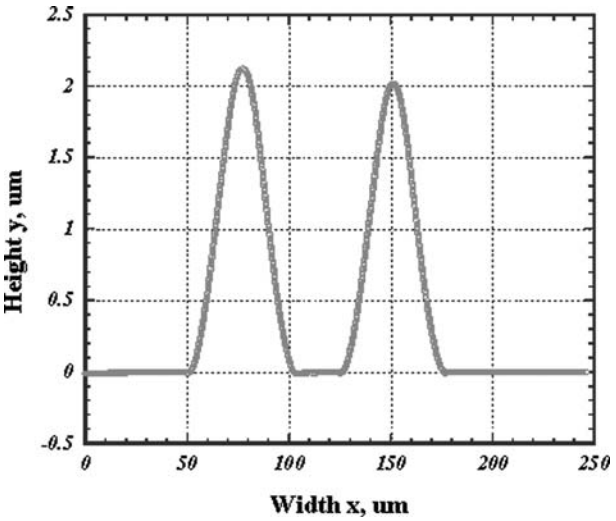
where δ is the blister height [Figure 4.24(b)]. For the example presented in Figures 4.23–4.24, a 1.9 GPa residual compressive stress is estimated, in good agreement with the measurements of wafer curvature and X-ray diffraction stress performed on the non-delaminated samples.

The film steady state interfacial toughness in the direction of blister propagation [Figure 4.24(a)] was estimated as

$$\Gamma_{SS} = \frac{(1-\nu^2)h\sigma_r^2}{2E}\left(1-\frac{\sigma_B}{\sigma_r}\right)^2, \tag{4.14}$$



(a)



(b)

FIGURE 4.24. Analysis of the telephone-cord delamination of a tungsten film shown in the previous figure. (a) Telephone cord delamination in a 1 μm tungsten film on top of a 2 nm diamond-like carbon (DLC) film on Si. (b) Corresponding blister heights profile.

which gave 3.6 J/m^2 in this case. Mode-dependent interfacial toughness in the buckling direction, perpendicular to blister propagation is

$$\Gamma(\Psi) = \frac{(1-\nu^2)h}{2E}(\sigma_r - \sigma_B)(\sigma_r + 3\sigma_B), \tag{4.15}$$

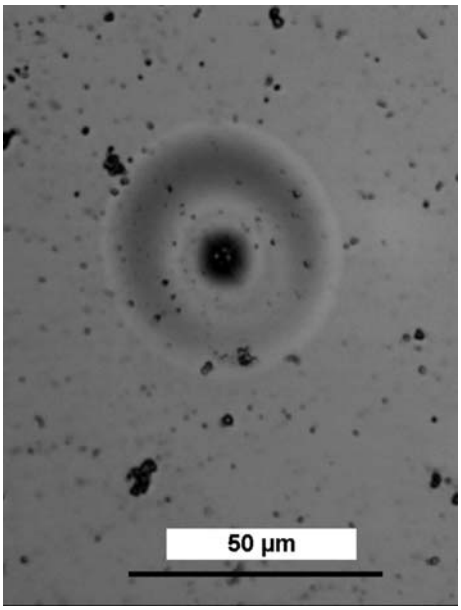
giving 6 J/m^2. This is in agreement with the superlayer indentation adhesion test result of 6.4 ± 0.4 J/m^2 [97].

The superlayer indentation test provides information on local film adhesion at the microscale. A superlayer film, selected for high stress, high strength, and high adhesion, is deposited on top of the film to be tested. Indentation is used to initiate delamination. The highly stressed hard superlayer provides additional driving force for interfacial crack propagation, and prevents plastic deformation of the tested film around the indenter. As the indenter tip is pressed against the superlayer film stack, it supplies additional energy necessary for crack initiation and propagation. The blister radius is measured optically, Figure 4.25(a). The indentation volume is obtained from the plastic depth of the load-displacement curve [Figure 4.25(b)] and the tip geometry. Both the blister radius and the indentation volume are then used to calculate the strain energy release rate (measure of the practical work of adhesion). Calculations for adhesion measurements were made by following the solution developed by Marshall and Evans [98] that was further expanded by Kriese and Gerberich et al. for multilayer films [99,100]. Figure 4.25(a) shows a typical delamination blister seen from making indents with a conical tip at 300 mN maximum load and a corresponding load-displacement curve. From Figure 4.25(b), the plastic indentation depth is obtained by using the power law fit of the top 65% of the unloading curve [76], and used to calculate the indentation volume, based on the tip geometry. It is assumed that the volume is conserved, and plastic deformation around the indenter results in the elastic displacement at the crack tip, allowing calculation of the indentation stress, and ultimately the strain energy release rate, a measure of the practical work of adhesion. Adhesion results for several film materials relevant to microelectronics are summarized in [101].

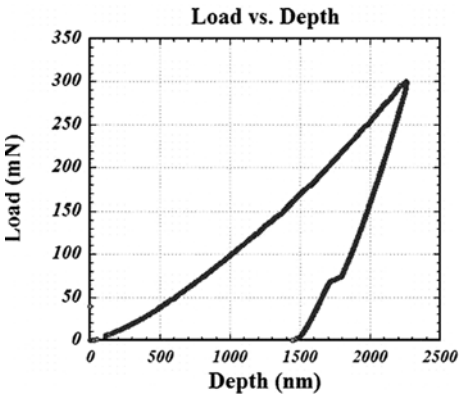
Moisture has been known to reduce the fracture toughness in bulk glasses [102], as well as causing failures in mirrors [103]. Figure 4.26 shows buckling delamination in a mirror backing layer that is over 200 years old. Similar effects are observed in thin films [97]. This becomes extremely important, as the films are exposed to moist and corrosive environments during processing (e.g., CMP). Figure 4.27 shows water-induced delamination of a tungsten film; telephone cord propagation of the delamination is seen in the upper portion *vs* no cracking on the dry lower portion. The modified superlayer indentation test conducted in moist environment showed an adhesion decrease up to 60 times compared to the tests in dry lab environment. In this case environmentally assisted fracture is observed; the Γ_I interfacial toughness term is decreased by the presence of water, so that the failure condition $G > \Gamma_I$ is met.

Because the interfacial energies found in films are numerically much lower than those in bulk metals, for which fracture toughness testing was developed, the four-point-bend bar with a cracks propagating along its length from a central notch has been found useful [90,104–106]. Below we briefly describe this technique, to show a specific application of fracture mechanics in thin film adhesion. The many reports of adhesion measurement methods in the literature are a testament to the importance of the problem, the difficulty of the measurement, and the ingenuity of the researchers. However, a detailed review is beyond the scope of this article.

The delaminating beam test specimen, Figure 4.28, is a four-point-bend bar with an interface of interest built into the interior of the beam along the whole length. A "sandwich" beam made with the substrate on the top and bottom, and the surface layers bonded together in the center, is a typical geometry. The substrate layers are much thicker than the interface layer, and give the assembly sufficient stiffness to handle. In the bending beam, the outer fiber in tension is often located on the upper side, and is conventionally referred



(a)



(b)

FIGURE 4.25. (a) Indentation-induced delamination blister in tungsten film; and (b) corresponding load-displacement curve.

to as the top of the specimen. The bottom fiber is in compression. The top section is carefully cut, without notching the bottom section. Cracks are intentionally nucleated to grow away from the notch along the interface layer being tested. While the crack length significantly exceeds the thickness of the cut layer, steady state energy release rate obtains, until the cracks reach the inner loading points of the four-point-bend specimen. The energy release rate is evaluated from the load and displacement, specimen geometry, and elastic properties of the support layers of the specimen. An advantage of this test is that the parameters needed to evaluate the adhesion do not include the residual stress on the film, which may be difficult to measure. Becker et al. [106] point out that properties measured with this specimen may depend on the specific geometry, contrary to the case for standard-

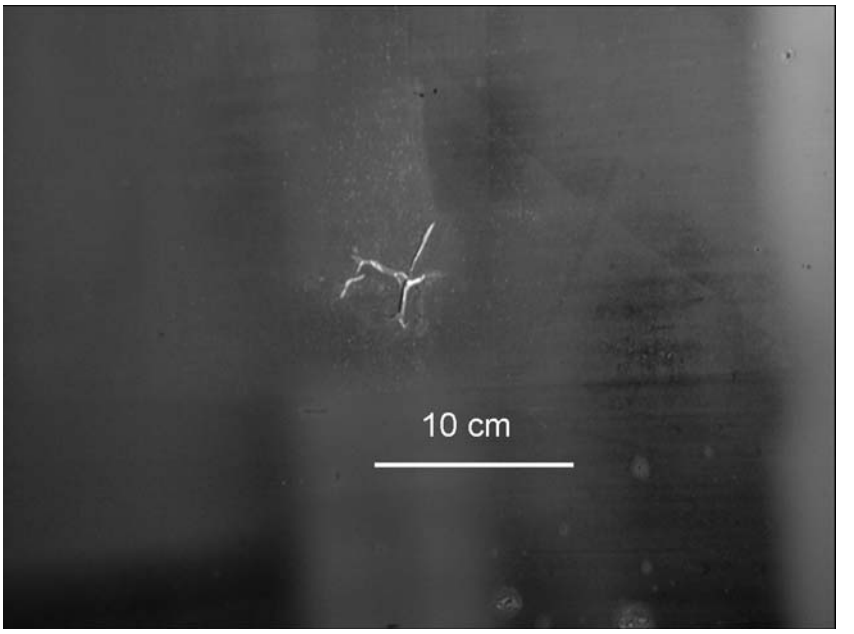


FIGURE 4.26. Environmentally assisted fracture in an antique mirror 200 years old.

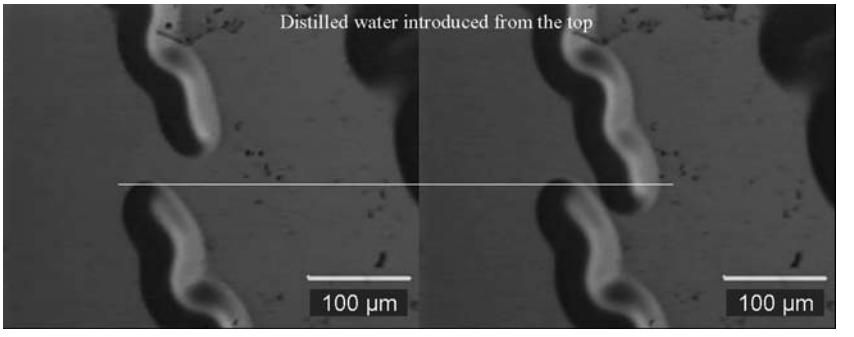


FIGURE 4.27. Water-induced blister growth. The image on the left shows blisters before water was introduced; the image on the right, after. The water was introduced from the direction of the top of the figure. The blister at the top (wet side) grew noticeably toward the lower side of the figure, while the blister at the bottom (dry side) did not grow.

ized fracture toughness specimens. This is not considered to be a serious disadvantage for testing materials for chips and electronic packages, because actual-size specimens can be tested.

All the fracture toughness techniques highlight a critical problem in the design of electronic packages and chips: some commonly used interfaces, such as polymer–metal interfaces, have very low fracture toughness, around 10 J/m^2 [90].

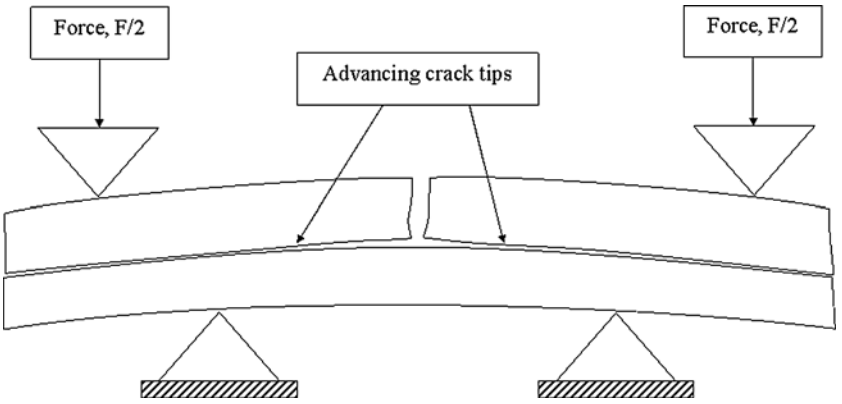


FIGURE 4.28. Delaminating beam specimen for measuring the energy required to separate an adhesive interface.

4.5. MATERIALS AND PROPERTIES

Reviews of the mechanical properties of thin films have appeared over the years [61,69]. The general conclusions formulated by Brotzen [69] are still relevant:

1. The elastic properties of thin films ordinarily do not differ greatly from bulk properties, and Young's moduli measured by various techniques do not differ greatly.
2. The thinner the films, the greater are the hardness and strength.
3. Specimen preparation methods and the structure of the films have a significant effect on the mechanical properties of the films.

Item 3 above should be applied as a cautionary remark to both previous items. Brotzen's conclusions will be adopted as "rules of thumb" for discussing and interpreting the measured properties of thin films. Application of these rules to some example cases will be discussed below.

4.5.1. Grain Size and Structure Size Effects

4.5.1.1. Hall-Petch Rule A useful general rule for strength properties of ductile metals is the Hall-Petch rule [107], which states that strength increases with the inverse square root of the grain size. In thin films the grain size often varies in the same direction as the film thickness, so rule 2 above can be seen as a special case of Hall-Petch behavior. Additional effects produced by the film-substrate interaction have been described, which tend in the same direction, that is, strength increases as the controlling size scale decreases. The small grain size usually present in thin films explains the common observation that thin films have higher strengths than bulk materials of the same chemical composition. Film strengths are usually portrayed as consistent with those of bulk materials when the grain sizes are considered.

Clearly the Hall-Petch rule must break down at some small grain size, because the size cannot be smaller than a single atom. Such a change in behavior has been detected recently at grain sizes of a few nanometers, although it is not yet clear whether the strength peaks or reaches a plateau. A set of data showing a clear peak in hardness at 11 nm was reported by Conrad and Narayan [108]. A molecular dynamics study by Schiotz and Jacobsen [55] provides some insight into the mechanism.

4.5.1.2. Low Elongation to Failure The effect of film thickness on elongation to failure in microtensile tests has not been explained in detail. Elongation is one of the measurements recorded in tensile tests of bulk materials; it is an indicator of a material's ductility and capacity for absorbing energy before fracture, and hence its fracture toughness. In bulk materials, values of 10% or more are common, and values exceeding 100% are possible. Metal thin films, often made using very pure starting materials, commonly exhibit very low elongation to failure in tensile tests. An elongation to failure of 1% for a specimen of copper or aluminum 1 μm thick is typical. This result is not commonly interpreted to mean that the film lacks the capacity for plastic deformation, but rather is considered to be a geometric effect. Deformed regions, once formed, reduce the area of the film locally, and do not work-harden enough to prevent additional deformation from accumulating. Specimens of polyimide [109], and of aluminum deposited under unusual conditions [68], have shown substantial elongation. Microscopic examination of the failed region in metal films, for example, by scanning electron microscopy (SEM), often shows fine-scale features characteristic of high local ductility, such as ligaments [110].

4.5.1.3. Young's Modulus Brotzen [69] comments at length on modulus values reported for bimaterial multilayer films with layer thicknesses of the order of nanometers. The situation appears to be that these bilayer films exhibit variable elastic properties as a function of layer thickness, but the variations are more modest than spectacular. A different type of modulus variation, the modulus deficit, was reported by Huang and Spaepen [111]. Here the relevant rule is number 3; film microstructure and morphology are the controlling features. Measurements of the density of thin films are difficult and are not commonly made. Porosity or voiding between grains is possible for some film deposition methods. Clearly a porous film may have an elastic modulus lower than a fully dense film of the same thickness. However, there are recent indications [112] that the typical expectation that the modulus deficit should be related to the porosity by a numerical factor near unity, as given by the results of Kachanov et al. [113], may underestimate the effect.

4.6. PROPERTIES OF SPECIFIC MATERIALS

Fabrication methods and test techniques for some widely used thin film materials have become sufficiently widespread that specific values of the properties can be usefully given. Of course, Brotzen's rule 3 still holds: the microstructure and chemical composition of the specific material at hand determine its properties.

Mechanical properties of selected thin materials as measured by microtensile testing or nanoindentation are listed in Table 4.2.

The practicality of the instrumented indentation test is a main reason for its widespread implementation. Of the two mechanical property values produced by the NI test, Young's modulus and hardness, only the Young's modulus is comparable to values produced by microtensile testing; even for the Young's modulus, conversion from the measured biaxial modulus is needed for comparison to handbook values. Limitations of applicability of the NI test, related to the film thickness and the relative stiffness of the film and the substrate, are being studied in current research. The high values of hardness produced in the NI tests, commonly reaching over 1 GPa, and occasionally over 10 GPa, are strikingly different from the strength values obtained by microtensile testing, Table 4.2. A basic rule of thumb is that the hardness value produced by indentation should be three times the stress value at some selected value of strain in the tensile test. The relevant strain value is sometimes given as 8%, but, as shown in Table 4.2 above, thin films rarely reach 8% strain in

TABLE 4.2.
Mechanical properties of selected thin films as measured by microtensile testing or nanoindentation. See notes below.

Material	Fabrication method	Thickness, μm	Yield strength, MPa	Ultimate tensile strength, MPa	Young's modulus, GPa	Elongation to failure, %	References
Microtensile results							
Al	sputtered	0.05	327		62[a]		[114]
Al	sputtered	0.1	700				[114]
Al	sputtered	0.2	330		70		[114]
Al	e-beam evaporated	0.2	205	375[b]	65–70		[72]
Al	e-beam evaporated	1	150		65–70		[72]
Al	e-beam evaporated	1	94	151	24–30	22.5	[68]
Al-0.5%Cu	MOSIS[d]	1.5 and 2.4	65	74	40	1.4	[115]
Cu	e-beam evaporated	0.2	345		125–129		[72]
Cu	e-beam evaporated	1	160[c]		125–129		[72]
Cu	electrodeposited	9.7	253	311	67		[116]
Au	e-beam evaporated	0.3, 0.5	220		53–55		[72]
Au	e-beam evaporated	1	90		53–55		[72]
Ni	electrodeposited	4.7		1516	102–114		[117]
Polyimide	Spun on, baked	0.6	103	181	5.5	24	[109]
PolySi	MUMPS 25, 30[d]	3.5	NA	0.95 GPa	157		[71]
PolySi	SUMMiT[d]	2.5	NA	3 GPa			[30]
Nanoindentation results							
Material	Fabrication method	Thickness, μm	Yield strength, MPa	Hardness, GPa	Young's modulus, GPa		References
Cu	electrodeposited, annealed	0.5		1.04–1.27	121–132		[118]
Cu	electrodeposited, annealed, polished	0.5		1.21–1.29	131–138		[118]
Cu	electrodeposited	0.2 to 2	530 to 330				[23]

The study by Espinosa et al. [72] is quoted extensively here because it is recent, uses consistent methods on gold, aluminum, and copper films, includes specimen thickness and width effects, and includes considerable microstructural characterization and post-test observations of the specimens.

[a] The authors remark that this value shows the effect of film thickness on Young's modulus. If correct, this would be one of the first experimental demonstrations of this theoretically predicted effect.

[b] Unusually high value for this material and thickness.

[c] Unusually low value for this material and thickness.

[d] Special proprietary deposition process; sources identified in the references.

microtensile tests. A more detailed study of the extraction of yield strength values from nanoindentation tests was reported by Kramer et al. [119].

4.7. FUTURE RESEARCH

4.7.1. Techniques

The experimental convenience of the nanoindentation test, combined with its focus on key properties, modulus and hardness, and its ability to produce a quantitative and reproducible force-displacement record, has made it widely popular. Outstanding problems include the derivation of design properties, such as yield strength, from nanoindentation data. Detailed data on typical features of the nanoindentation test, such as sink-in and pile-up, are being sought to help clarify the interpretation of the results. The microtensile test requires patterning of a test structure, but offers less ambiguous measurements of the yield and ultimate tensile strengths, along with an indication of elongation to failure and a fracture surface which can be examined by TEM. More generally, the perennial problem of the expense of fabrication of test structures seems to weigh heavily against the routine measurement of mechanical property values of thin films.

4.7.2. Properties

The difficulty of measuring adhesion in multilayer thin film structures, and criticality of this property for reliability, seem to offer an opportunity for improved measurement techniques. Nanoindentation has been applied to the measurement of adhesion [91], but more extensive data are needed to demonstrate the generality and usefulness of this approach to adhesion. The fracture mechanics based tests appear to be easier to analyze, at the cost of preparation of the bend bar specimens.

Some test techniques have found Young's modulus values that differ from bulk values in some cases. The scientific issue of whether or not thin films have Young's modulus values different from bulk materials has been settled; any film with a substantial number of atomic layers, meaning, 20 or more, should have elastic properties indistinguishable from those of a bulk specimen of the same microstructure. This implies that the elastic properties of thin films can now be regarded as probes of the microstructure. Microstructural features that may affect elastic behavior include mass density and, for metals, the volume and behavior of material effectively in grain boundaries.

4.7.3. Length Scale

The effects of "length scale" on mechanical behavior have received intense study recently, e.g., [120]. A complementary view is that fine-scale microstructural variation may become significant when an experiment samples the behavior of very small volumes of material, for example, the material under the sharp tip of a nanoindenter [121].

The ultimate length scale for mechanical properties measurements is the few-atom scale, which is clearly outside the main scope of this chapter, but may be a source of insight into the behavior of practical materials, and may also be important in establishing the accuracy of atomistic modeling approaches such as molecular dynamics. In this regard, the results reported by Rubio-Bollinger et al. [122], which seem to show that a chain of gold atoms can have a strength value (calculated by the authors from the results in [122])

surpassing that of high-strength steel, stand as a challenge to material testers and material developers alike.

REFERENCES

1. Sematech Internatonal Roadmap Committee, International Technology Roadmap for Semiconductors, 2003 Edition, Interconnect, 2003.
2. L.B. Freund and S. Suresh, Thin Film Materials: Stress, Defect Formation and Surface Evolution, Cambridge University Press, Cambridge, UK, 2003.
3. E.S. Machlin, Materials Science in Microelectronics: The Relationships between Thin Film Processing and Structure, Giro Press, Croton-on-Hudson, NY, 1995.
4. D.L. Smith, Thin Film Deposition: Principles and Practice, McGraw-Hill, Boston, MA, 1995.
5. M. Ohring, Materials Science of Thin Films, Deposition and Structure, Academic Press, San Diego, CA, 2002.
6. J.F. Nye, Physical Properties of Crystals, Clarendon, Oxford, 1985.
7. G. Simmons and H. Wang, Single Crystal Elastic Constants and Calculated Aggregate Properties: A HANDBOOK, MIT Press, Cambridge, MA, 1971.
8. R.G. Munro, Elastic Moduli Data for Polycrystalline Ceramics, NISTIR 6853, National Institute of Standards and Technology, Gaithersburg, MD, 2002.
9. Y. Okada and Y. Tokumaru, Precise determination of lattice-parameter and thermal-expansion coefficient of silicon between 300-K and 1500-K, Journal of Applied Physics, 56(2), pp. 314–320 (1984).
10. C.A. Harper, Handbook of Materials and Processes for Electronics, McGraw-Hill, New York, 1970.
11. D.R. Lide, CRC Handbook of Chemistry and Physics, CRC Press, Boca Raton, Florida, USA, 2004.
12. Y.S. Touloukian, R.K. Kirby, R.E. Taylor, and T.Y.R. Lee, Thermophysical Properties of Matter: Thermal Expansion, Nonmetallic Solids, Vol. 13, IFI/Plenum, New York, 1977.
13. A. Ballato, Elastic properties of crystalline quartz, in M. Levy, ed., Handbook of Elastic Properties of Solids, Liquids, and Gases, Volume II, Academic Press, San Diego, 2001.
14. J.V. Atanasoff and P.J. Hart, Dynamical determination of the elastic constants and their temperature coefficients for quartz, Physical Review, 59(1), pp. 85–96 (1941).
15. A.W. Lawson, Comment on the elastic constants of alpha-quartz, Physical Review, 59(9), pp. 838–839 (1941).
16. R. Geiss, Personal communication, SEM and TEM micrographs, 2005.
17. C. Moglestue, J. Rosenzweig, J. Kuhl, M. Klingenstein, M. Lambsdorff, A. Axmann, J. Schneider, and A. Hulsmann, Picosecond pulse response characteristics of GaAs metal-semiconductor-metal photodetectors, Journal of Applied Physics, 70(4), pp. 2435–2448 (1991).
18. M. Zirngibl and M. Ilegems, High-speed photodetectors on InGaAs/GaAs-on-GaAs superlattices, Journal of Applied Physics, 69(12), pp. 8392–8398 (1991).
19. R.R. Keller, A. Roshko, R.H. Geiss, K.A. Bertness, and T.P. Quinn, EBSD measurement of strains in GaAs due to oxidation of buried AlGaAs layers, Microelectronic Engineering, 75(1), pp. 96–102 (2004).
    Personal communication, SEM and TEM micrographs, 2005.
20. M. Morgen, E.T. Ryan, J.H. Zhao, C. Hu, T.H. Cho, and P.S. Ho, Low dielectric constant materials for ULSI interconnects, Annual Review of Materials Science, 30 pp. 645–680 (2000).
21. M.B. Anand, M. Yamada, and H. Shibata, Use of gas as low-*k* interlayer dielectric in LSI's: demonstration of feasibility, IEEE Transactions on Electron Devices, 44(11), pp. 1965–1971 (1997).
22. Y.H. Wang, M.R. Moitreyee, R. Kumar, S.Y. Wu, J.L. Xie, P. Yew, B. Subramanian, L. Shen, and K.Y. Zeng, The mechanical properties of ultra-low-dielectric-constant films, Thin Solid Films, 462-63, pp. 227–230 (2004).
23. A.A. Volinsky and W.W. Gerberich, Nanoindentation techniques for assessing mechanical reliability at the nanoscale, Microelectronic Engineering, 69(2-4), pp. 519–527 (2003).
24. J.B. Vella, I.S. Adhihetty, K. Junker, and A.A. Volinsky, Mechanical properties and fracture toughness of organo-silicate glass (OSG) low-*k* dielectric thin films for microelectronic applications, International Journal of Fracture, 119(4-2), pp. 487–499 (2003).
25. A.A. Volinsky, J.B. Vella, and W.W. Gerberich, Fracture toughness, adhesion and mechanical properties of low-K dielectric thin films measured by nanoindentation, Thin Solid Films, 429(1-2), pp. 201–210 (2003).
26. Y.L. Loo, T. Someya, K.W. Baldwin, Z.N. Bao, P. Ho, A. Dodabalapur, H.E. Katz, and J.A. Rogers, Soft, conformable electrical contacts for organic semiconductors: high-resolution plastic circuits by lamination,

Proceedings of the National Academy of Sciences of the United States of America, 99(16), pp. 10252–10256 (2002).
27. H. Koezuka, A. Tsumura, H. Fuchigami, and K. Kuramoto, Polythiophene field-effect transistor with polypyrrole worked as source and drain electrodes, Applied Physics Letters, 62(15), pp. 1794–1796 (1993).
28. A. Lodha and R. Singh, Prospects of manufacturing organic semiconductor-based integrated circuits, IEEE Transactions on Semiconductor Manufacturing, 14(3), pp. 281–296 (2001).
29. P.M. Osterberg and S.D. Senturia, M-TEST: a test chip for MEMS material property measurement using electrostatically actuated test structures, Journal of Microelectromechanical Systems, 6(2), pp. 107–118 (1997).
30. D.A. La Van, T. Tsuchiya, G. Coles, W.G. Knauss, I. Chasiotis, and D.T. Read, Cross comparison of direct strength testing techniques on polysilicon films, in C. Muhlstein and S.B. Brown, Eds., Mechanical Properties of Structural Films, American Society for Testing and Materials, West Conshohoken, Pennsylvania, 2001, pp. 16–27.
31. Y.D. Jeon, K.W. Paik, A. Ostmann, and H. Reichl, Effects of Cu contents in Pb-free solder alloys on interfacial reactions and bump reliability of Pb-free solder bumps on electroless Ni-P under-bump metallurgy, Journal of Electronic Materials, 34(1), pp. 80–90 (2005).
32. S.W. Russell, S.A. Rafalski, R.L. Spreitzer, J. Li, M. Moinpour, F. Moghadam, and T.L. Alford, Enhanced adhesion of copper to dielectrics via titanium and chromium additions and sacrificial reactions, Thin Solid Films, 262(1-2), pp. 154–167 (1995).
33. M.Y. Kwak, D.H. Shin, T.W. Kang, and K.N. Kim, Characteristics of TiN barrier layer against Cu diffusion, Thin Solid Films, 339(1-2), pp. 290–293 (1999).
34. M. Ueki, M. Hiroi, N. Ikarashi, T. Onodera, N. Furutake, N. Inoue, and Y. Hayashi, Effects of Ti addition on via reliability in Cu dual Damascene interconnects, IEEE Transactions on Electron Devices, 51(11), pp. 1883–1891 (2004).
35. O. Chyan, T.N. Arunagiri, and T. Ponnuswamy, Electrodeposition of copper thin film on ruthenium—A potential diffusion barrier for Cu interconnects, Journal of the Electrochemical Society, 150(5), C347–C350 (2003).
36. M.F. Doerner and W.D. Nix, Stresses and deformation processes in thin-films on substrates, Crc Critical Reviews in Solid State and Materials Sciences, 14(3), pp. 225–268 (1988).
37. J.A. Floro, S.J. Hearne, J.A. Hunter, P. Kotula, E. Chason, S.C. Seel, and C.V. Thompson, The dynamic competition between stress generation and relaxation mechanisms during coalescence of Volmer-Weber thin films, Journal of Applied Physics, 89(9), pp. 4886–4897 (2001).
38. J.G. Swadener, B. Taljat, and G.M. Pharr, Measurement of residual stress by load and depth sensing indentation with spherical indenters, Journal of Materials Research, 16(7), pp. 2091–2102 (2001).
39. A.L. Shull and F. Spaepen, Measurements of stress during vapor deposition of copper and silver thin films and multilayers, Journal of Applied Physics, 80(11), pp. 6243–6256 (1996).
40. S.G. Malhotra, Z.U. Rek, S.M. Yalisove, and J.C. Bilello, Analysis of thin film stress measurement techniques, Thin Solid Films, 301(1–2), pp. 45–54 (1997).
41. M.M. de Lima, R.G. Lacerda, J. Vilcarromero, and F.C. Marques, Coefficient of thermal expansion and elastic modulus of thin films, Journal of Applied Physics, 86(9), pp. 4936–4942 (1999).
42. Y.L. Shen, S. Suresh, and I.A. Blech, Stresses, curvatures, and shape changes arising from patterned lines on silicon wafers, Journal of Applied Physics, 80(3), pp. 1388–1398 (1996).
43. H.T.G. Hentzell, C.R.M. Grovenor, and D.A. Smith, Grain-structure variation with temperature for evaporated metal-films, Journal of Vacuum Science & Technology, A-Vacuum Surfaces and Films, 2(2), pp. 218–219 (1984).
44. W. Zhang, S.H. Brongersma, T. Clarysse, V. Terzieva, E. Rosseel, W. Vandervorst, and K. Maex, Surface and grain boundary scattering studied in beveled polycrystalline thin copper films, Journal of Vacuum Science & Technology B, 22(4), pp. 1830–1833 (2004).
45. B.D. Cullity, Elements of X-Ray Diffraction, Second ed., Addison-Wesley, Reading, Massachusetts, 1978.
46. J.M.E. Harper, C. Cabral, P.C. Andricacos, L. Gignac, I.C. Noyan, K.P. Rodbell, and C.K. Hu, Mechanisms for microstructure evolution in electroplated copper thin films near room temperature, Journal of Applied Physics, 86(5), pp. 2516–2525 (1999).
47. Y. Hanaoka, K. Hinode, K. Takeda, and D. Kodama, Increase in electrical resistivity of copper and aluminum fine lines, Materials Transactions, 43(7), pp. 1621–1623 (2002).
48. D. Josell, C. Burkhard, Y. Li, Y.W. Cheng, R.R. Keller, C.A. Witt, D.R. Kelley, J.E. Bonevich, B.C. Baker, and T.P. Moffat, Electrical properties of superfilled sub-micrometer silver metallizations, Journal of Applied Physics, 96(1), pp. 759–768 (2004).
49. C.E. Schuster, M.G. Vangel, and H.A. Schafft, Improved estimation of the resistivity of pure copper and electrical determination of thin copper film dimensions, Microelectronics Reliability, 41(2), pp. 239–252 (2001).

50. B.C. Johnson, Electrical-resistivity of copper and nickel thin-film interconnections, Journal of Applied Physics, 67(6), pp. 3018–3024 (1990).
51. F.C. Brown, The Physics of Solids, Benjamin, New York, 1967.
52. R.H. Geiss, R.R. Keller, D.T. Read, and Y.-W. Cheng, TEM-based analysis of defects induced by AC thermo-mechanical versusmicrotensile deformation in aluminum thin films, Materials Research Society Conference Proceedings, Symposium B, 2005, to be published.
53. J.R. Greer and W.D. Nix, Size dependence of mechanical properties of gold at the sub-micron scale, Applied Physics A-Materials Science & Processing, 80(8), pp. 1625–1629 (2005).
54. M.A. Haque and M.T.A. Saif, In situ tensile testing of nanoscale freestanding thin films inside a transmission electron microscope, Journal of Materials Research, 20(7), pp. 1769–1777 (2005).
55. J. Schiotz and K.W. Jacobsen, A maximum in the strength of nanocrystalline copper, Science, 301(5638), pp. 1357–1359 (2003).
56. E. Arzt, O. Kraft, R. Spolenak, and Y.C. Joo, Physical metallurgy of electromigration: failure mechanisms in miniaturized conductor, Zeitschrift fur Metallkunde, 87(11), pp. 934–942 (1996).
57. M. Ohring, Failure and Reliability of Electronic Materials and Devices, Academic Press, Boston, MA, 1998.
58. N. Tamura, A.A. MacDowell, R. Spolenak, B.C. Valek, J.C. Bravman, W.L. Brown, R.S. Celestre, H.A. Padmore, B.W. Batterman, and J.R. Patel, Scanning X-ray microdiffraction with submicrometer white beam for strain/stress and orientation mapping in thin films, Journal of Synchrotron Radiation, 10 pp. 137–143 (2003).
59. C.K. Hu, K.P. Rodbell, T.D. Sullivan, K.Y. Lee, and D.P. Bouldin, Electromigration and stress-induced voiding in fine Al and Al-alloy thin-film lines, Ibm Journal of Research and Development, 39(4), pp. 465–497 (1995).
60. IBM, http://www.ibmchips.com\cmos7s.htm, Photo courtesy of International Business Machines Corporation. Unauthorized use not permitted.
61. D.A. Hardwick, The mechanical-properties of thin-films—a review, Thin Solid Films, 154(1-2), pp. 109–124 (1987).
62. X.Y. Ding, W.H. Ko, and J.M. Mansour, Residual-stress and mechanical-properties of boron-doped P+-silicon films, Sensors and Actuators A-Physical, 23(1-3), pp. 866–871 (1990).
63. D.T. Read and J.W. Dally, A new method for measuring the strength and ductility of thin-films, Journal of materials research, 8(7), pp. 1542–1549 (1993).
64. D.T. Read, Piezo-actuated microtensile test apparatus, Journal of Testing and Evaluation, 26(3), pp. 255–259 (1998).
65. W.N. Sharpe, B. Yuan, and R.L. Edwards, A new technique for measuring the mechanical properties of thin films, Journal of Microelectromechanical Systems, 6(3), pp. 193–199 (1997).
66. D.T. Read, Young's modulus of thin films by speckle interferometry, Measurement Science & Technology, 9(4), pp. 676–685 (1998).
67. J.C. Fox, R.L. Edwards, and W.N. Sharpe, Thin-film gage markers for laser-based strain measurement on MEMS materials, Experimental Techniques, 23(3), pp. 28–30 (1999).
68. D.T. Read, Y.W. Cheng, R.R. Keller, and J.D. McColskey, Tensile properties of free-standing aluminum thin films, Scripta Materialia, 45(5), pp. 583–589 (2001).
69. F.R. Brotzen, Mechanical testing of thin-films, International Materials Reviews, 39(1), pp. 24–45 (1994).
70. J.A. Ruud, D. Josell, F. Spaepen, and A.L. Greer, A new method for tensile testing of thin films, Journal of Materials Research, 8(1), pp. 112–117 (1993).
71. W.N. Sharpe, K.M. Jackson, G. Coles, M.A. Eby, and R.L. Edwards, Tensile tests of various thin films, in C. Muhlstein and S.B. Brown, Eds., Mechanical Properties of Structural Films, American Society for Testing and Materials, West Conshohoken, Pennsylvania, 2001, pp. 229–247.
72. H.D. Espinosa, B.C. Prorok, and B. Peng, Plasticity size effects in free-standing submicron polycrystalline FCC films subjected to pure tension, Journal of the Mechanics and Physics of Solids, 52(3), pp. 667–689 (2004).
73. M.R. Van Landingham, Review of instrumented indentation, Journal of Research of the National Institute of Standards and Technology, 108(4), pp. 249–265 (2003).
74. J.L. Loubet, J.M. Georges, O. Marchesini, and G. Meille, Vickers indentation curves of magnesium-oxide (MgO), Journal of Tribology-Transactions of the Asme, 106(1), pp. 43–48 (1984).
75. M.F. Doerner and W.D. Nix, A method for interpreting the data from depth-sensing indentation measurements, Journal of Materials Research, 1(4), pp. 601–616 (1986).
76. W.C. Oliver and G.M. Pharr, An improved technique for determining hardness and elastic-modulus using load and displacement sensing indentation experiments, Journal of Materials Research, 7(6), pp. 1564–1583 (1992).

77. R.B. King and T.C. Osullivan, Sliding contact stresses in a two-dimensional layered elastic half-space, International Journal of Solids and Structures, 23(5), pp. 581–597 (1987).
78. T.Y. Tsui, W.C. Oliver, and G.M. Pharr, Influences of stress on the measurement of mechanical properties using nanoindentation. 1. Experimental studies in an aluminum alloy, Journal of Materials Research, 11(3), pp. 752–759 (1996).
79. A. Bolshakov, W.C. Oliver, and G.M. Pharr, Influences of stress on the measurement of mechanical properties using nanoindentation. 2. Finite element simulations, Journal of Materials Research, 11(3), pp. 760–768 (1996).
80. S.V. Hainsworth, H.W. Chandler, and T.F. Page, Analysis of nanoindentation load-displacement loading curves, Journal of Materials Research, 11(8), pp. 1987–1995 (1996).
81. R. Berriche, Vickers hardness from plastic energy, Scripta Metallurgica et Materialia, 32(4), pp. 617–620 (1995).
82. J.B. Pethica and W.C. Oliver, Tip surface interactions in Stm and Afm, Physica Scripta, T19A, pp. 61–66 (1987).
83. X.D. Li and B. Bhushan, A review of nanoindentation continuous stiffness measurement technique and its applications, Materials Characterization, 48(1), pp. 11–36 (2002).
84. W.D. Nix, Mechanical-properties of thin-films, Metallurgical Transactions, A-Physical Metallurgy and Materials Science, 20(11), pp. 2217–2245 (1989).
85. A.F. Jankowski and T. Tsakalakos, Effects of deflection on bulge test measurements of enhanced modulus in multilayered films, Thin Solid Films, 291, pp. 243–247 (1996).
86. M.K. Small and W.D. Nix, Analysis of the accuracy of the bulge test in determining the mechanical-properties of thin-films, Journal of Materials Research, 7(6), pp. 1553–1563 (1992).
87. K.M. Liechti and A. Shirani, Large-scale yielding in blister specimens, International Journal of Fracture, 67(1), pp. 21–36 (1994).
88. K.E. Petersen and C.R. Guarnieri, Youngs modulus measurements of thin-films using micromechanics, Journal of Applied Physics, 50(11), pp. 6761–6766 (1979).
89. T.P. Weihs, S. Hong, J.C. Bravman, and W.D. Nix, Mechanical deflection of cantilever microbeams—a new technique for testing the mechanical-properties of thin-films, Journal of Materials Research, 3(5), pp. 931–942 (1988).
90. R. Dauskardt, M. Lane, Q. Ma, and N. Krishna, Adhesion and debonding of multi-layer thin film structures, Engineering Fracture Mechanics, 61(1), pp. 141–162 (1998).
91. A.A. Volinsky, N.R. Moody, and W.W. Gerberich, Interfacial toughness measurements for thin films on substrates, Acta Materialia, 50(3), pp. 441–466 (2002).
92. M. Lane, Interface fracture, Annual Review of Materials Research, 33, pp. 29–54 (2003).
93. M.D. Thouless, Cracking and delamination of coatings, Journal of Vacuum Science & Technology A-Vacuum Surfaces and Films, 9(4), pp. 2510–2515 (1991).
94. J.W. Hutchinson and Z. Suo, Mixed-mode cracking in layered materials, Advances in Applied Mechanics, 29, pp. 63–191 (1992).
95. J.L. Beuth, Cracking of thin bonded films in residual tension, International Journal of Solids and Structures, 29(13), pp. 1657–1675 (1992).
96. J.W. Hutchinson and Z. Suo, Mixed-mode cracking in layered materials, Advances in Applied Mechanics, 29, pp. 63–191 (1992).
97. A.A. Volinsky, P. Waters, J.D. Kiely, and E.C. Johns, Sub-Critical Telephone Cord Delamination Propagation and Adhesion Measurements, Materials Research Society, Warrendale, PA, 2004, pp. U9.5.
98. D.B. Marshall and A.G. Evans, Measurement of adherence of residually stressed thin-films by indentation. 1. Mechanics of interface delamination, Journal of Applied Physics, 56(10), pp. 2632–2638 (1984).
99. M.D. Kriese, W.W. Gerberich, and N.R. Moody, Quantitative adhesion measures of multilayer films: Part I. Indentation mechanics, Journal of Materials Research, 14(7), pp. 3007–3018 (1999).
100. M.D. Kriese, W.W. Gerberich, and N.R. Moody, Quantitative adhesion measures of multilayer films: Part II. Indentation of W/Cu, W/W, Cr/W, Journal of Materials Research, 14(7), pp. 3019–3026 (1999).
101. A.A. Volinsky, N.R. Moody, and W.W. Gerberich, Interfacial toughness measurements for thin films on substrates, Acta Materialia, 50(3), pp. 441–466 (2002).
102. S.M. Wiederhorn, S.W. Freiman, E.R. Fuller, and C.J. Simmons, Effects of water and other dielectrics on crack-growth, Journal of Materials Science, 17(12), pp. 3460–3478 (1982).
103. V.P. Burolla, Deterioration of the silver-glass interface in 2 surface solar mirrors, Solar Energy Materials, 3(1-2), pp. 117–126 (1980).
104. M. Charalambides, Fracture mechanics specimen for interface toughness measurement, Journal of Applied Mechanics, 56(0), pp. 77–82 (1989).

105. I. Hofinger, M. Oechsner, H.A. Bahr, and M.V. Swain, Modified four-point bending specimen for determining the interface fracture energy for thin, brittle layers, International Journal of Fracture, 92(3), pp. 213–220 (1998).
106. T.L. Becker, J.M. McNaney, R.M. Cannon, and R.O. Ritchie, Limitations on the use of the mixed-mode delaminating beam test specimen: effects of the size of the region of K-dominance, Mechanics of Materials, 25(4), pp. 291–308 (1997).
107. G. Dieter, Mechanical Metallurgy, McGraw-Hill, New York, 1986.
108. H. Conrad and J. Narayan, Mechanism for grain size softening in nanocrystalline Zn, Applied Physics Letters, 81(12), pp. 2241–2243 (2002).
109. D.T. Read, Y.-W. Cheng, and J.D. McColskey, Microtensile behavior of a commercial photodefinable polyimide, in Proceedings of the 2002 SEM Annual Conference, Society for Experimental Mechanics, Bethel, Connecticut, 2002, pp. 64–67.
110. R.R. Keller, J.M. Phelps, and D.T. Read, Tensile and fracture behavior of free-standing copper films, Materials Science and Engineering A-Structural Materials Properties Microstructure and Processing, 214(1-2), pp. 42–52 (1996).
111. H.B. Huang and F. Spaepen, Tensile testing of free-standing Cu, Ag and Al thin films and Ag/Cu multilayers, Acta Materialia, 48(12), pp. 3261–3269 (2000).
112. D.T. Read, Atomistic simulation of modulus deficit in an aggregate of metal spheres, Journal of Applied Physics, 97(1), (2005).
113. M. Kachanov, I. Tsukrov, and B. Shafiro, Effective moduli of solids with cavities of various shapes, Applied Mechanics Reviews 47(1), S151–S174 (1994).
114. M.A. Haque and T.A. Saif, Uniaxial tensile and bending experiments on nanoscale metal films, in A. Shukla, R.M. French, A. Andonian, and K. Ramsey, Eds., Proceedings of the 2002 SEM Annual Conference & Exposition on Experimental and Applied Mechanics, Society of Experimental Mechanics, Bethel, Connecticut, 2002, pp. 134–138.
115. D.T. Read, Y.-W. Cheng, J.D. McColskey, and R.R. Keller, Mechanical behavior of contact aluminum alloy, in C.S. Ozkan, L.B. Freund, R.C. Cammarata, and H. Gao, Eds., Thin Films: Stressses and Mechanical Properties IX, Materials Research Society, Warrendale, Pennsylvania, 2002, pp. 263–268.
116. D.T. Read, R. Geiss, J. Ramsey, T. Scherban, G. Xu, J. Blaine, B. Miner, and R.D. Emery, Nanoindentation and tensile behavior of copper films, in D.F. Bahr, Ed., Mechanical Properties Derived from Nanostructuring Materials, Materials Research Society, Warrendale, PA, 2003, pp. 93–98.
117. B. Yeung, W. Lytle, V. Sarihan, D.T. Read, and Y. Guo, Applying a methodology for microtensile analysis of thin films, Solid State Technology, 125–129 (2002). Ref Type: Magazine Article.
118. A.K. Sikder, A. Kumar, P. Shukla, P.B. Zantye, and M. Sanganaria, Effect of multistep annealing on mechanical and surface properties of electroplated Cu thin films, Journal of Electronic Materials, 32(10), pp. 1028–1033 (2003).
119. D. Kramer, H. Huang, M. Kriese, J. Robach, J. Nelson, A. Wright, D. Bahr, and W.W. Gerberich, Yield strength predictions from the plastic zone around nanocontacts, Acta Materialia, 47(1), pp. 333–343 (1998).
120. M.A. Haque and M.T.A. Saif, Strain gradient effect in nanoscale thin films, Acta Materialia, 51(11), pp. 3053–3061 (2003).
121. J.R. Greer, W.C. Oliver, and W.D. Nix, Size dependence of mechanical properties of gold at the micron scale in the absence of strain gradients, Acta Materialia, 53(6), pp. 1821–1830 (2005).
122. G. Rubio-Bollinger, S.R. Bahn, N. Agrait, K.W. Jacobsen, and S. Vieira, Mechanical properties and formation mechanisms of a wire of single gold atoms, Physical Review Letters, 8702(2) (2001).

5

Carbon Nanotube Based Interconnect Technology: Opportunities and Challenges

Alan M. Cassell and Jun Li

Center For Nanotechnology NASA Ames Research Center, Moffett Field, CA 94035, USA

Abstract As candidate materials for future wiring technologies, carbon nanotubes possess extraordinary physical and electrical characteristics. Carbon nanotubes have high current carrying capacity, excellent thermal conductivity, low thermal expansion coefficients, and are less susceptible to electromigration than conventional interconnect materials such as copper, tungsten and aluminum. It is likely that carbon nanotubes in combination with conventional materials will be implemented as a hybrid solution in on-chip interconnect technologies. Contact resistance at the nanotube–metal interface becomes a primary area for reliability engineering. Recent improvements in plasma based processing have demonstrated that individual, high-length-to-diameter ratio, vertically oriented carbon nanotubes can be fabricated to achieve architectures useful for advanced technologies. In this chapter, we present an overview of carbon nanotubes based electronics and describe our recent works in the development of carbon nanotube as a candidate interconnect material. The overview is limited to the fundamental characteristics of carbon nanotubes as implemented in wiring applications. We address the challenges and opportunities facing carbon nanotube implementation in CMOS semiconductor processing, as well as other possible nanoelectromechanical applications.

5.1. INTRODUCTION: PHYSICAL CHARACTERISTICS OF CARBON NANOTUBES

5.1.1. Structural

The term carbon nanotube (CNT) has become a broadly descriptive term encompassing tubular nanostructures composed of graphitic carbon. However, to understand CNT properties as well as those of closely related structures, it becomes necessary to understand the bonding of the carbon atoms, the chemical nature of the bonds, and how the crystallites formed by these graphitic carbon sheets arrange themselves into different topological forms. Graphitic carbon is a planar hexagonal network arrangement of sp^2 bonded (covalent organic bonding nomenclature) carbon atoms [Figure 5.1(A)]. For a detailed de-

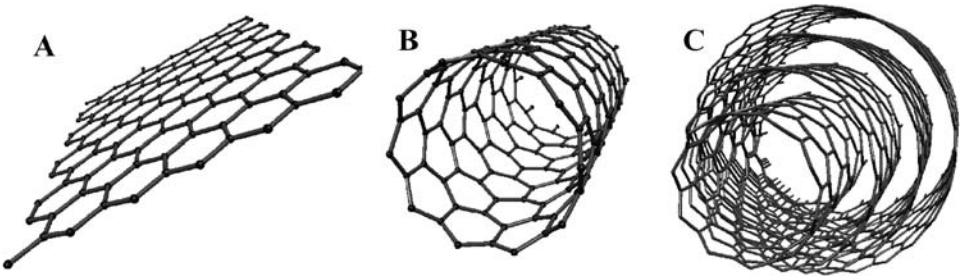


FIGURE 5.1. Nanostructures of graphitic carbon. A single graphene sheet is depicted in panel A. Single-walled CNT (10, 0) is shown in panel B, and a multiwalled CNT is shown in panel C.

scription of the atomic arrangement and bonding character of graphitic carbon, the reader is referred to references [1,2]. Graphitic carbon is also known as graphene, and it is the curvature, helical twist and the amount of crystalline defects or impurities in the graphene sheet(s) that largely determines the properties of CNTs and related structures, namely electrical resistivity, modulus, and thermal conductivity.

Structurally perfect CNTs are seamless cylinders composed of rolled graphene sheets. These sheets occur as single layers [single walled CNT, SWCNT, Figure 5.1(B)] or concentric layers [multiwalled CNT, MWCNT, Figure 5.1(C)]. SWCNTs can have diameters between 0.4 nm and 5 nm. Typical diameters are between 0.7 nm and 3.0 nm. SWCNTs can be very long. Experimental observations have reported lengths over several mm [3]. The arrangement of the atoms in SWCNTs gives rise to "armchair," "zigzag" or "chiral" tubes. MWCNTs (two or more concentric layers of tubular graphene sheets) have tubes with inner diameters as small as a 0.4 nm and outer diameters up to hundreds of nanometers. The individual layers in MWCNTs are spaced by approximately 0.34 nm. The lengths of MWCNTs can also span several hundred μm.

As has been indicated, the term carbon nanotube has become a catchword for describing many of the non-planar, tubular forms of graphitic carbon. Other less crystalline morphologies, such as the bamboo [4], nanofiber [5], nanoscroll [6] and nanotubule [7] are related to structurally perfect MWCNTs. There are some distinct differences that dramatically impact the properties and hence the potential uses of these alternate tube embodiments. The vast number of carbon nanostructure morphologies occurs from the incorporation of various defects, such as disclinations, dislocations and pentagonal and heptagonal rings, during growth of the graphitic crystallites. Defective graphene stacking arrangements along the tube/fiber axis influence the CNT properties, namely, electron/phonon transport characteristics and mechanical characteristics. When considering tubular carbon nanostructures for many applications, it is imperative to delineate their microstructural attributes and correlate these attributes with the observed material properties. Ultimately it is the amount of defects, impurities and crystalline order that determine the scope of the CNT applicability.

5.1.2. Electrical

Most of the recent research activities have been focused on the electrical characteristics of CNTs. Theoretical investigations [8–10] that have been confirmed by experimental observations have revealed many extraordinary properties, such as ballistic electron transport, as well as the metallic and semiconducting properties of SWCNTs. Generally, the

electronic properties of an isolated defect-free SWCNT are dependent on the helicity of the CNT structure. The helicity is defined as the vector **C**, along which a graphene sheet is rolled up into a hollow cylinder. The vector **C** can be expressed in terms of two integers (m, n) corresponding to the two primary lattice vectors $\mathbf{a}_1$ and $\mathbf{a}_2$ of the hexagonal graphite structure, i.e.,

$$\mathbf{C} = m\mathbf{a}_1 + n\mathbf{a}_2. \tag{5.1}$$

Upon forming a SWCNT, the two end-points of the vector **C** are exactly superimposed. Hence, the nominator (m, n) carries the information of both the helicity and the size of the SWCNT. It is known that a defect-free SWCNT presents semiconducting properties if

$$(m - n)/3 \neq \text{integer}, \tag{5.2}$$

and it presents metallic properties if

$$(m - n)/3 = \text{integer}. \tag{5.3}$$

In the present production technology, the helical arrangement of carbon atoms in SWCNTs is random during the growth process. As a result, statistically 2/3 of the randomly produced SWCNTs are semiconducting and 1/3 of them are metallic [11]. The band gap for a semiconducting tube is given by

$$E_g = 2d_{cc}\gamma/D, \tag{5.4}$$

where $d_{cc} = 0.142$ nm is the C−C bond length of the graphene, $\gamma = 2.5$–3.2 eV is the nearest-neighbor hopping parameter, and D is the diameter of the SWCNT [12].

For the application as interconnect materials, metallic SWCNTs are of greater interest. It has been experimentally confirmed [13,14] that a single SWCNT behaves as a ballistic quantum wire intrinsically due to the confinement effect on the tube circumference. The conductance of a single SWCNT is given by

$$G = G_o MT = (2e^2/h)MT, \tag{5.5}$$

where $G_o = (2e^2/h) = (12.9\ \text{k}\Omega)^{-1}$ is the quantized conductance [15], M is the number of conducting modes which is fixed at $M = 2$ for a perfect SWCNT, and T is the transmission probability for an electron through the contacts and the tube. Ideally this probability is near unity. In reality, T may be significantly lower than 1 due to electron–electron coupling, tube–environment coupling, scattering from defects and impurities, structural distortions, and poor metal–CNT contacts.

For interconnect wires, either on a solid surface or encapsulated in dielectric materials, the residual interactions with the environment induce disorder in the initially perfect metallic SWCNT. The theory of transport in such disordered materials predicts a transition from the quantum resistance $R_0 = (G_o M)^{-1} = 6.45\ \text{k}\Omega$ to a localized regime where resistance R increases exponentially with the length L:

$$R(L) = R_0 e^{L/\xi}, \tag{5.6}$$

where ξ is the *localization length* for electrons in the disordered wire. For wires with M open conduction channels, ξ is given by $\xi = Ml$ where l is the elastic mean free path for backward electron scattering [16].

For normal metallic quantum wires with fixed amount of disorder, as the transverse size increases, l remains fixed, but ξ increases due to the introduction of new channels. However, the motion of the electrons is largely diffusive for $M \gg 1$. In contrast, for weakly disordered SWCNTs, M is pinned at 2 and the motion of electrons remains ballistic. This property makes SWCNTs as potential long ballistic conductors, which are fundamentally different from current interconnects. Furthermore, White et al. [16] predict that, l, the elastic mean free path for backward electron scattering, increases linearly with the diameter D, or the number of carbon atoms in the cross-section ring, for example $2N_{\mathrm{B}}$ in an "armchair" (m,m) tube, i.e.,

$$l \propto N_{\mathrm{B}} \propto D. \tag{5.7}$$

This property is likely valid with other types of metallic SWCNTs as well. The importance of this property is that, by increasing the diameter of the SWCNT, one can achieve a long ballistic conductor with a localization length easily over 10 μm for normal SWCNTs, which is not attainable with conventional metallic quantum wires.

The quantum resistance $R_0 = (G_{\mathrm{o}}M)^{-1} = 6.45\ \mathrm{k}\Omega$ is too high for practical interconnects. Using SWCNT bundles or ropes, which consist of n SWCNTs, can solve this problem. It has been reported before that SWCNTs tend to form crystalline hexagonally packed ropes held by van der Waals interaction during growth processes [17]. A rope of about 20 nm diameter may consist of over 100 SWCNTs. The total resistance of such a parallel assembly dramatically drops to

$$R_{0,\mathrm{n}} = R_0/n = h/(4e^2n), \tag{5.8}$$

and the conductance is increased to

$$G = nG_{\mathrm{o}}M = 4ne^2/h. \tag{5.9}$$

Meanwhile, the electron wave propagation speed increases with $\sqrt{n}$, causing the inductance to be lowered by more than one order of magnitude, so that the latency is essentially defined by the RC charge-up time. The stacking of SWCNTs into a bundle or rope reduces the disorder attributed to tube–environment coupling and the total capacitance, both of which are advantageous for interconnects.

MWCNTs can be considered as parallel concentric SWCNTs. Each graphene shell may present electrical properties similar to an isolated SWCNT with the same helical structure. The helicity of each shell is fully random during growth, and there is little coupling to neighboring shells. Most electrical transport studies only make contact to the outermost shell of MWCNTs [18,19]. However, it is highly desired for interconnect applications that electrical contacts can be made to each graphene shell to form a parallel ballistic quantum conductor [20] with the conductance as

$$G = N * G_{\mathrm{o}}M * T = 4NTe^2/h, \tag{5.10}$$

where N is the number of graphitic shells in the MWCNT, and T is the transmission probability with an ideal value at unity similar to that in Equation (5.5).

Parallel assemblies of MWCNTs have fundamental advantages over SWCNT bundles/ropes. First, the diameter D is much larger. MWCNTs can be readily produced with a diameter from a few nanometers to the size of the technology node on the order of tens of nanometers, while the number of shells N can be varied from 2 to tens. For example, for a MWCNT with 30 nm diameter, $N = 30$ can be obtained. As indicated by Equation (5.7), the mean free path length of electron, l, can be dramatically increased as D increases. The bandgap is significantly reduced according to Equation (5.4), resulting in metallic properties in nearly every individual shell at room temperature.

Second, the graphene shells are closely stacked with only 0.34 nm separation between them. Unlike the relatively large interstitial space in SWCNT bundles, impurities such as gas molecules cannot access the tight space between the MWCNT shells. This will eliminate the scattering by adsorbates for all the graphene shells except the outermost one. As a result, the individual inner graphene shells in MWCNTs should behave almost as perfect defect-free CNTs except for a very small tube–tube coupling.

A recent study reports that 2D electronic transport within the stack of a few layers of atomically flat graphitic sheets on a SiO_2 surface remains ballistic up to nearly micron distances [21]. It is reasonable to speculate that by rolling them into MWCNTs with a diameter up to 100 nm, these graphene sheets may remain as ballistic conductors. The full exploration of such novel materials for interconnects has yet to be undertaken.

5.1.3. Mechanical

There have been many studies to understand and harness the tremendous mechanical properties of CNTs, which are highlighted by a large Young's modulus (>1 TPa) and tensile strength (150 GPa) [22]. The mechanical characteristics of CNTs and graphite come from the in plane C−C covalent bonding network. In-plane graphene has a high Young's modulus of $\sim$1 TPa, and CNT modeling and simulations predict a slight modulus increase over graphite from the axial bonding enhancement gained when rolling the sheet into a seamless cylinder [23]. Early experimental efforts confirmed these predictions, and it was found that indeed an individual single or multiwalled CNT possess Young's moduli near or exceeding 1 TPa [24]. Equally remarkable is that under deformation, CNTs elastically respond, and can sustain up to 15% tensile strain before fracture [22]. These high levels of strain are attributed to the high levels of elastic buckling through which the stress is released. It is postulated [25] that the sp^2 bonds in carbon undergo rehybridization to release the stress. Indeed, CNTs are one of the strongest materials known so far. A detailed overview of mechanical characteristics of CNTs can be found in [26,27].

These outstanding mechanical attributes have been particularly focused on the use of CNTs in bulk applications as reinforcements for composite materials, including polymers [28–34], ceramics [35–37] and metal-matrix [38–40] systems. Progress has been somewhat limited in the development of bulk composites due to the lack of sufficient quantities of raw nanotubes and also due to the limited application of models that successfully bridge the relevant length scales (micro, meso, macro) and give detailed conceptual insight as to the relevant failure mechanisms occurring at each level. More progress has been made at the nanoscale, for example, CNTs have been investigated as robust probe tips [41–44] in scanning probe microscopy, and as electromechanical devices for use as relays in circuitry [45–47]. In the context of interconnect wiring applications, processing individual or bundles of CNTs becomes paramount. Reliable electrical performance is coupled with reliable mechanical performance, especially when it is likely that integrating CNTs with conven-

tional semiconductor materials and processing schemes will introduce unknown effects at the interfaces between different materials.

5.1.4. Thermal

Both graphite and diamond exhibit extraordinary thermal conductivity and thus it is no surprise that CNTs also exhibit similar thermal properties. As in graphite, MWCNTs have highly anisotropic thermal conductivity values that differ by more than 5 orders of magnitude in the graphene planes vs between planes. Experimental results show a temperature dependence on the measured specific heat, which is consistent with weak interlayer coupling. Furthermore, recent measurements on individual CNTs have reported thermal conductivities as high as 3000 W/mK [48], while random oriented bulk samples have shown thermal conductivities of 200 W/mK or less [49]. Phonons dominate thermal transport in these materials, and weak intertube coupling will impact how these materials are processed for use in thermal management applications. Indeed, many groups are now pursuing various approaches for using CNTs as thermal management materials [50–54], and the control of CNT alignment, interfacial contact, and processability are prime considerations for finding niche solutions for these materials.

The thermal expansion property of CNTs is another fact making them attractive as interconnect materials, which has not been addressed in the literature. This property is very important to define the reliability of interconnects under thermal cycling. A MWCNT is expected to show similar highly anisotropic thermal expansion properties as do graphite crystals. The coefficient of thermal expansion (CTE) normal to the graphite basal plane, α_c, is a slowly varying parameter with a value $\alpha_c = \sim 20 \times 10^{-6}$ to 29×10^{-6} K^{-1} over a wide temperature range (from about -200 to 700°C) [55]. In contrast, the CTE parallel to the graphite basal plane, α_a, has a negative value from ~ -250 to $\sim$300°C with a minimum around room temperature of $\alpha_a = -1.2 \times 10^{-6}$ K^{-1} and becomes positive at higher temperatures [56]. A recent study [57] confirmed that the CTE of SWCNTs along the axial direction, $\alpha_{//}$, shows similar behavior as α_a along the graphite basal planes. MWCNTs are expected to exhibit similar properties.

5.2. CNT FABRICATION TECHNOLOGIES

The physical properties of CNTs are strongly dependent on structural defects and impurities present in the graphite sheets comprising the tubes. Depending upon the growth technique employed, a plethora of CNT sub types that show remarkably different physical characteristics can be obtained. Growth techniques that generate pure carbon vapor such as arc-discharge [59] and laser ablation [60] are known to produce CNT materials with the highest crystalline quality. This is largely due to the high process temperature, which can anneal defects in the graphene sheets. One major drawback of the arc-discharge and laser-ablation approaches is that many unwanted impurities (amorphous carbon, catalyst) remain in the deposited bulk material and can be quite cumbersome to purify. In addition, it is quite difficult to scale these growth techniques for mass production, although recently it has been shown that arc discharge can effectively be scaled [61] for kg scale production.

Chemical vapor deposition (CVD) growth approaches that take place at much lower temperatures have been recently developed. These approaches can be scaled for the production of bulk quantities. In addition, CVD technologies can be utilized with other materials

processing capabilities and afford broader process latitude, such as growth on patterned substrates [62,63]. The ability to precisely control the location of the CNTs is accomplished using a catalyst-mediated growth process from lithographically patterned metals. This technique has become the method of choice for most applications development due to its economy and ease [64]. CVD of CNTs is generally performed at temperatures between 600–900°C. In some cases the as-grown material has many structural defects. This drawback is far outweighed by the other processing benefits that CVD affords. These drawbacks are explored in detail below.

5.2.1. Chemical Vapor Deposition of Carbon Nanotubes

There are two main classes of CVD growth processes, namely, thermal and plasma enhanced. Both techniques share many similarities. The major difference is in the type of the heat source that is employed during growth. Thermal CVD (hot wall approach) uses conventional heat sources such as a tube furnace [65] to heat the walls of a process tube. In a typical tube furnace embodiment, substrates are placed inside the tube, and the atmosphere (temperature, pressure, gas composition, flow rate) inside the process tube is carefully controlled to affect the growth of the CNTs from transition metal catalyst placed on the substrate. Methane (or other appropriate carbon containing gas feedstock) is then decomposed over transition metal catalysts (Fe, Co, Ni, etc.) that nucleate the CNTs. The diameter of the CNTs is largely controlled by the diameter of the catalyst seed that nucleates the individual CNT, and the CNT length is controlled by the growth time. The prime advantage that CVD offers over other CNT growth approaches is the ability to grow from patterned substrates. This has been demonstrated in the successful integration of CNTs onto silicon substrates for the assembly of electronic device components such as transistors and field emission devices, as well as in mechanical devices, such as AFM probe tips. Many improvements and process enhancements have been developed for thermal CVD [66–69]. A key drawback is that it is not possible to generate, individual, vertically aligned, freestanding CNTs. Figure 5.2 shows patterned catalyst arrays (patterned using e-beam lithography) after thermal CVD growth [Figure 5.2(A)]. The 45° perspective view shows that the multiwalled CNTs nucleated from the particles are randomly oriented, and their lengths are quite non-uniform. It is advantageous to have vertical orientation control for optimizing downstream integration and processing in specific applications. Furthermore, utilizing individual, vertically oriented, freestanding CNTs eliminates intertube coupling effects that could adversely affect for example, the electrical performance characteristics.

A more advanced CNT growth approach that gives further process latitude is the use of plasma generating sources alone or in combination with conventional heat sources [70]. Plasma enhanced CVD (PECVD) allows for lower process temperatures because precursor dissociation is affected by high-energy electron impact reactions that would otherwise take place at much higher temperatures when using thermal CVD. In PECVD, there are a number of methods used to generate plasmas. High efficiency, high density plasmas from RF capacitive and inductive [71–78] and microwave sources [79–87] have been used in the growth of CNTs, as well as low density glow discharges generated using DC sources [Figure 5.3(A)] [88–104]. Regardless of the plasma source used, one of the key aspects is the generation of electric fields on the wafer surface in the plasma sheath. Electric fields present in the sheath orient the growth of CNTs normal to the substrate surface, which leads to the growth of individual, freestanding CNTs [Figure 5.2(B)]. While obtaining freestanding CNT structures is important for many applications (i.e., field emitters, relays,

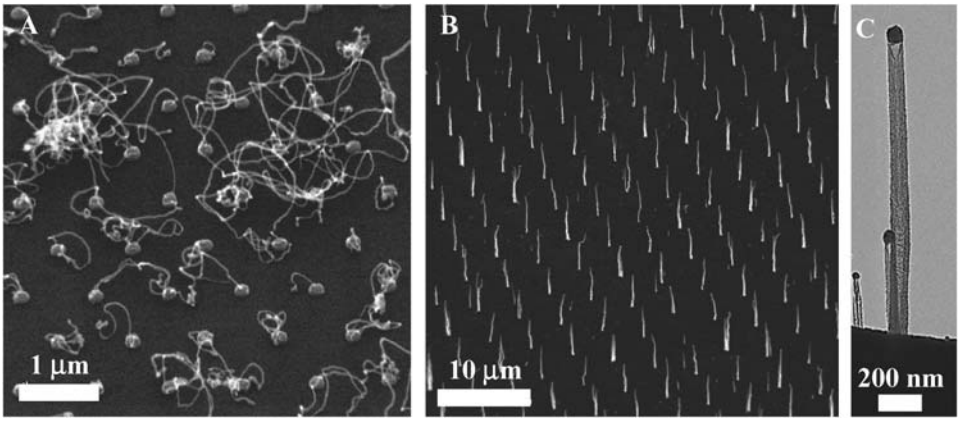


FIGURE 5.2. Comparing SEM images (taken at 45°) of thermal CVD derived individual CNTs (Panel A) with PECVD derived individual CNTs (Panel B). In the case of thermal CVD, it is difficult to control the length and the growth uniformity across the patterned catalyst arrays whereas in PECVD, growth uniformity is much better. Panel C shows a cross-sectional transmission electron micrograph of an individual, free-standing CNT obtained using PECVD.

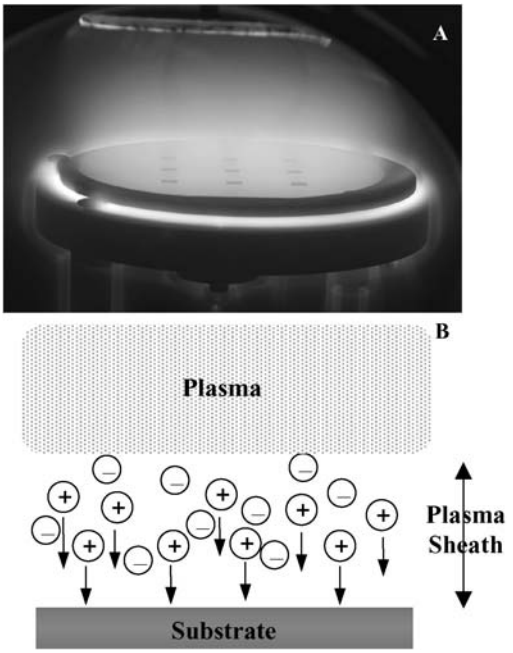


FIGURE 5.3. A glow discharge from a DC plasma used to grow carbon nanofibers (panel A). A plasma sheath forms over the wafer surface where high energy ion bombardment can occur.

interconnects, AFM tips), there are trade-offs between processing conditions and deleterious effects to the substrate. In particular, large cathode bias often leads to high-energy ion bombardment that cause damage to the growth substrate such as a silicon wafer and may pose adverse impacts to the as-grown CNTs. Another limitation of PECVD based growth

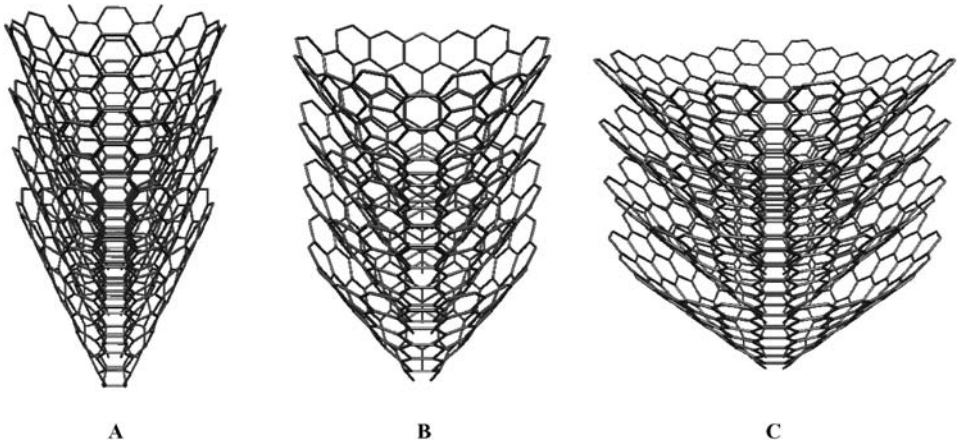


FIGURE 5.4. Typical carbon nanofiber morphologies achieved from PECVD showing different cone angles relative to the fiber axis. The ideal nanofiber would have the smallest cone angle possible to have nearly identical transport properties as multiwalled CNTs. Depicted are 20°, 30° and 42° cone angles (panels A, B, C respectively).

of CNTs is that current process temperatures are still too high for many common materials used in complementary metal oxide semiconductor (CMOS) process flows. Solutions are now being sought for achieving low temperature growth while retaining the high crystalline order of the CNTs. In general, lower growth temperatures lead to much lower CNT crystalline quality, and the growth of individual, freestanding structures gives carbon nanofiber morphologies (Figure 5.4). A cross-sectional transmission electron micrograph of a carbon nanofiber is shown in Figure 5.3(C).

5.2.2. Process Integration and Development

A number of reports have outlined the motivations for using CNTs in interconnect technology [105–111]. A few of these reports have introduced potential solutions for various layers in the hierarchical wiring architecture. Two main paths have been developed for achieving CNTs as vias at both the local and intermediate levels. The first method (top-down approach) closely resembles the conventional copper damascene path that follows a patterning, dielectric dry-etch (for via hole), catalyst deposition, and CNT growth pathway [106–108]. The second methodology (bottom-up approach) removes the dry etching step to achieve the via hole and instead follows a catalyst patterning, CNT growth, dielectric deposition, planarization route [110,111]. Both methods have their advantages and shortcomings, and ultimately the method that affords the most economically viable route will be adopted. A more detailed description of the process development challenges is given in Section 5.4.

In order to integrate CNTs as replacement materials for wiring in integrated circuits, three key areas need to be further explored: (1) more advanced PECVD growth processes and assembly techniques for vertical and horizontally oriented CNTs; (2) barrier layer and interface contact materials engineering; and (3) multilevel CNT processing/integration solutions utilizing new planarization and etch processes.

Figure 5.5 shows a cross-sectional view of a potential architecture whereby individual, vertically aligned CNTs would replace local copper vias, small bundles of CNTs would

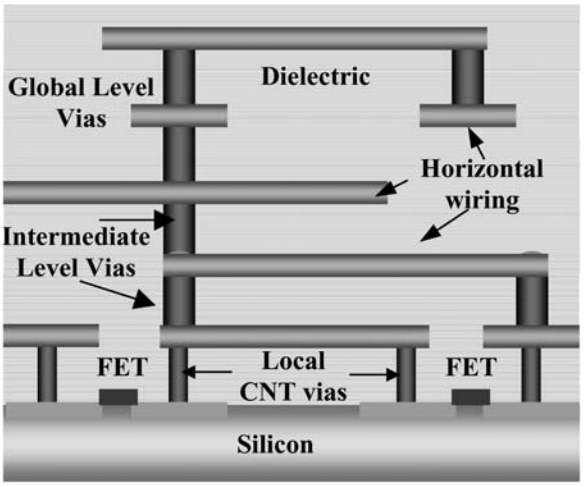
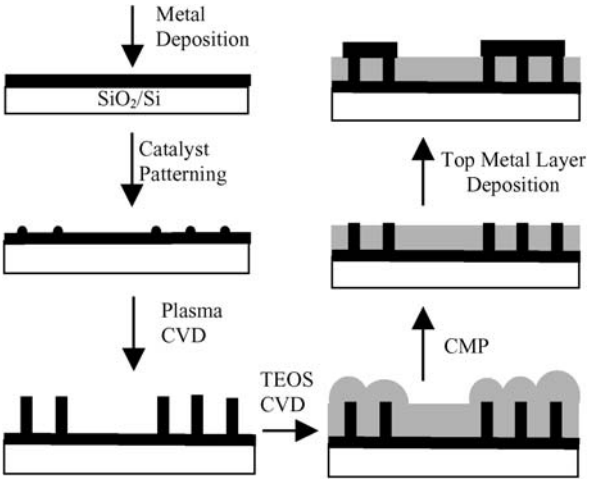


FIGURE 5.5. Cross-sectional view of envisioned semiconductor circuit whereby CNTs are implemented as replacements for Cu interconnects [107].



Scheme 5.1. Fabrication process sequence for a local level CNT interconnect test structure [111].

replace intermediate level vias and transmission lines, and large CNT bundles would replace global vias and transmission lines. The most likely near term implementation would be where individual vertically aligned CNTs would replace local Cu vias, and a longer term embodiment would be the use of bundles of vertically oriented CNTs as intermediate level vias. It is still unclear how to implement horizontally oriented bundles in a multilevel embodiment as described in Figure 5.5.

As a potential solution for local vias, a generic fabrication process flow would start from the MOSFET layer, whereby vertical CNTs would be grown in a bottom-up fashion by PECVD to the desired height [107]. Next, interlayer dielectric would be deposited and polished (exposing the ends of the CNTs), then patterning and deposition of the horizontal Cu wiring lines would complete the first level (see Scheme 5.1). Thereafter, conventional

copper damascene process technology would be used to build the successive wiring layers. Successive generations then would make use of the intermediate and global level CNT wiring solutions outlined earlier.

5.3. CARBON NANOTUBES AS INTERCONNECTS

The ITRS predicts that in 2013, current densities of $3.3 \times 10^6/\text{cm}^2$ will be needed in conventional ICs, which can not be obtained with current interconnect material technology. As is well documented [112], copper based interconnects suffer from many drawbacks as features downsize. Namely, grain and surface scattering effects drastically impact the resistivity as the cross-section of metal wires reaches the mean free path for electron scattering. Furthermore, electromigration (current induced displacement of atoms) can lead to voids, producing an open circuit or accumulation points, giving shorts. The shortcomings of traditional interconnect materials and processing technology is leading many to consider new material solutions and process integration technologies. As outlined above, CNTs are extremely promising candidate wiring materials for ICs. A unique combination of high current carrying capacity, high thermal conductivity, mechanical integrity and electron mean free paths of several microns have initiated numerous investigations to study potential solutions for implementing these nanoscale materials as interconnects.

5.3.1. Limitations of the Current Technology

Designers have taken the performance gains afforded by low resistivity copper interconnects (resistivity assumption of $1.7 \times 10^{-6}\ \Omega\,\text{cm}$) and low-dielectric constant materials to increase the speed of interconnects while reducing the capacitance per unit length. It has now become recognized that material solutions only partially solve the problem, and are likely to only extend the life of the existing paradigm for a few years [113,114]. Dimensional constraints will pose even more challenges on materials (interface quality, operation temperature, and barrier layer profile) for maintaining low resistivity in the near-term. On the processing side, the current interconnect fabrication pathway including dielectric etch for trenches/vias, deposition to fill metal plugs and planarization is reaching certain limitations. The aspect ratio for contact holes for DRAM stacked capacitors is currently 12:1 and is expected to increase to 23:1 by 2016 [107]. Fabricating high aspect ratio features with straight sidewalls is difficult and will become even more complex in the future. Furthermore, aspect ratio dependent etching as via hole diameters decrease will be difficult to overcome [115]. These alarming trends warrant the exploration of new materials and processing schemes to alleviate these concerns and to continue on the path of Moore's Law. Although CNTs themselves pose many technical challenges, their fundamental materials properties call for a concerted effort to fully investigate the potential solutions CNTs offer in future wiring technologies. Future development then should focus on the key technical hurdles for introduction of CNT interconnect technology and the development of reliable methods for predicting performance for future technology needs.

5.3.2. Architecture, Geometry and Performance Potential Using Carbon Nanotubes

5.3.2.1. Vertical Geometry At present, it appears that for local vias, there are potential CNT solutions for ultra large-scale integration (ULSI) [105,107,108,111]. It is still unclear at present how to achieve horizontally oriented bundles of CNTs at the wafer-scale

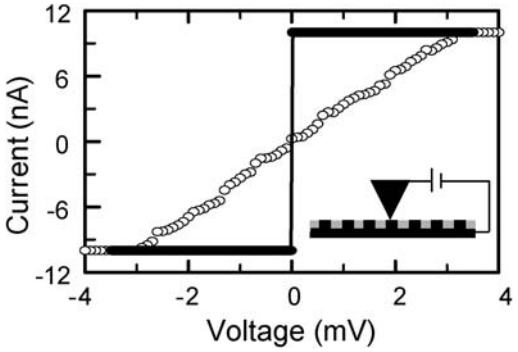


FIGURE 5.6. *I–V* curve of an individual CNT embedded in SiO_2 (open circles). The solid line represents a bundle of a few tubes [111].

for higher level wiring solutions, although laboratory demonstrations have shown promise [68,69]. Various fabrication approaches have been developed for achieving individual or bundles of vertically aligned MWCNTs [106,107,110,111] for interconnects. An example of such a process for a CNT interconnect test structure is outlined in Scheme 5.1. Onto silicon <100> wafer with 500 nm of thermal grown oxide is deposited 200 nm thick Ta or Cr contact lines. Next, 20 nm Ni catalyst films are deposited onto pre-patterned sites. Then plasma CVD is used to grow a MWNT array by an inductively coupled plasma process or dc plasma-assisted hot filament CVD as reported previously [71,94]. Each CNT is vertically aligned and freestanding on the surface [see Fig. 5.2(B) and (C)]. Such individual freestanding CNT structures are not possible by thermal CVD but are enabled in PECVD due to the electric field normal to the substrate. Next, the free space between the individual CNTs is filled with SiO_2 by CVD using tetraethoxysilane (TEOS) [110]. This is followed by chemical mechanical planarization (CMP) to produce a CNT array embedded in SiO_2 with only the CNT ends exposed over the planarized solid surface [110]. The top metal lines are then deposited for resistance measurement.

5.3.2.2. Electron Transport Characteristics To demonstrate the viability of implementing CNTs grown using the PECVD process in interconnect systems, we present results of electrical characterization of individual, small bundles, and large bundles of MWCNTs connected in parallel prepared using the process outlined in Scheme 5.1.

The transport characteristics of individual carbon nanofibers can be assessed on the planarized SiO_2–CNT structure without the top metal line using atomic force microscopy (AFM) modified with a current sensing module (CSAFM). This technique can measure electrical properties of individual CNTs. The Si_3N_4 cantilever was coated with a Pt film so that a voltage bias can be applied. The technique can simultaneously generate the topography, deflection, and current images of an embedded CNT sample. The topography signal (not shown) indicates that CNTs protrude out of the SiO_2 matrix consistent with SEM images of the same surface. Corresponding to the CSAFM images generated, the AFM tip can be easily positioned over different conducting spots to generate *I–V* curves of individual CNTs quickly. Figure 5.6 shows typical *I–V* curves of a single MWCNT and a compact bundle ($\sim 250 \times 500\ \mu m^2$) in the embedded array. The *I–V* curve for the single MWCNT is a straight line within the instrumental limits of ± 10 nA. The resistance of the single MWCNT is about 300 kΩ but that of the bundle is much lower than the 2 kΩ instrumental limit. The *I–V* curve of the insulating SiO_2 shows a flat line at zero with a 1 pA rms

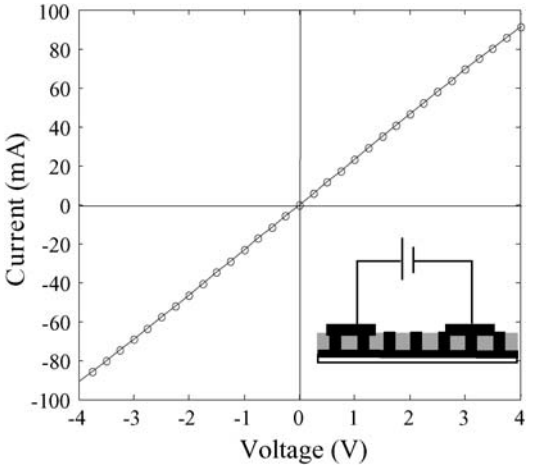


FIGURE 5.7. Typical *I–V* curve for parallel MWNTs connected in series through two 10×10 μm chromium contact pads.

noise (data not shown). Further two-terminal *I–V* measurements using a four-probe station linked with a semiconductor parameter analyzer were carried out to inspect the bundles in the ±5.0 V range. In these experiments, repeatedly applying $\sim 1 \times 10^6$ A/cm^2 current density over many hours did not show any damage indicated by *I–V* measurements. Note that ref. [108] already reports no degradation even at 10^{10} A/cm^2 with only loose thermal contacts. So, it is expected that CNTs embedded in a SiO_2 matrix would withstand current densities far higher than that desired by the ITRS [112].

Figure 5.7 shows a typical *I–V* curve of MWCNTs contacted through two contact pads. The near-ohmic behavior of the curve suggests that parallel MWCNTs ($R = 43\ \Omega$) may provide performance comparable to copper vias for on-chip interconnect architectures without the processing difficulties presented by current state-of-the-art copper deposition technology. The sheet resistance of the chromium underlayer was also considered in our study. An area of bare chromium, subjected to the same growth conditions as the CNTs, was characterized using two- and four-terminal measurements. The resulting chromium resistance for the two-terminal measurement was ∼10–15 Ω, while the four-terminal measurement yielded ∼4–6 Ω of resistance, most of which can be attributed to the resistance of the contact at the probe–Cr interface. While these resistance values seem small, if we compare the two-terminal *I–V* measurement of the nanotubes to the contact–Cr resistance, we can see that the proportion of contact resistance to overall resistance is quite high.

The importance of depositing a quality contact for assessing transport characteristics becomes apparent when we compare directly contacting parallel MWCNTs with a 25 μm diameter tungsten probe tip to contacting nanotubes through the patterned 10×10 μm^2 contacts. Figure 5.8 shows the stark contrast in resistance values when considering the deposition of metal contacts over multiple nanotubes. This result suggests that there is a distinct chemical and physical interaction between the deposited top metal layer and the end of the MWCNT that creates a less resistive metal–CNT junction. Dangling bonds at the nanotube tip after chemical mechanical polishing (CMP) enhance the possibility of creating a quality metal–CNT contact using metal deposition. A similar study using thermal CVD to grow nanotubes has been conducted in [105], however, the structural and electrical integrity [111] of such devices under standard microelectronic fabrication techniques such as CMP

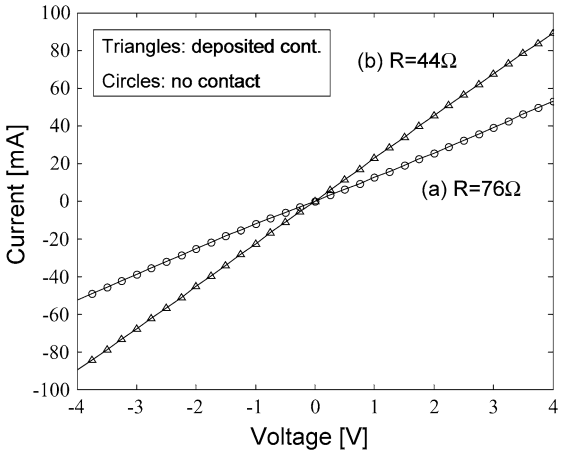


FIGURE 5.8. *I–V* curves comparing parallel MWNTs (a) contacted directly with a 25 μm contact (circles) and (b) with deposited 10 × 10 μm contact pads (triangles).

must also be considered. How to improve the metal–CNT contact will be further discussed in the next section.

5.4. DESIGN, MANUFACTURE AND RELIABILITY

In this section we discuss the challenges facing the implementation of CNTs as vias in local and intermediate level interconnect wiring solutions. Attention is placed on the microstructural properties of CNTs and the effects on electrical characteristics as well as interface materials integration challenges.

5.4.1. Microstructural Attributes and Effects on Electrical Characteristics

As described in the introductory section, the quality of the CNT itself contributes to the observed resistance. It is known that most plasma-grown structures are somewhat defective, characterized by periodic bamboo-like or ice cream cone-like closed shells along the axis, as confirmed by TEM images. True ballistic behavior is possible only with ideal MWCNTs. Whereas an ideal MWCNT will have all the walls parallel to the central axis ($\theta = 0$), most plasma-grown structures exhibit small θ values and hence, they are sometimes referred to as multiwalled carbon nanofibers (CNFs). In any case, the electrons have to cross the graphitic layers in such structures in order to be transported from one end to the other. This gives a much larger resistance similar to what one would get perpendicular to the basal plane of graphite. Our best two-terminal *I–V* measurements with the embedded vertical structure shown in Scheme 5.1 after minimizing the metal–CNT contact resistance gives about 3 kΩ per CNT (50 nm diameter and a length of ∼4 μm). This value is close to the estimated resistance of ∼2 kΩ expected for classical (diffuse) conduction through a high-quality carbon fiber (with $\rho \approx 1 \times 10^{-4}$ Ω cm) [116]. This is consistent with the bamboo-type CNT structure produced by PECVD in our experiments. Such microstructural attributes needs to be elaborated for future interconnect development.

A charge transport model for bamboo-type carbon nanotubes (Figure 5.9) demonstrates the effect of cone angle on the electrical resistivity [117]. This empirical model

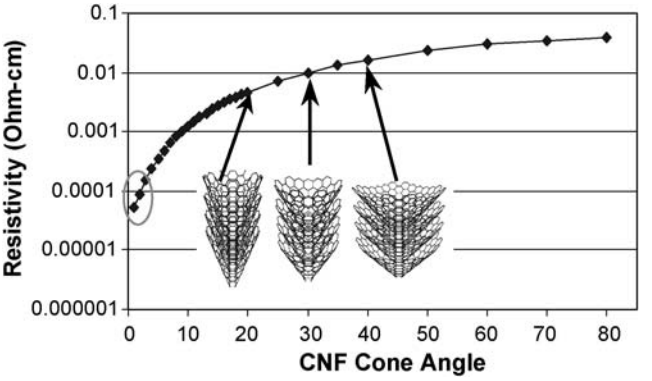


FIGURE 5.9. Effective resistivity as a function of CNT cone angle. Ideal multiwalled CNTs have cone angles of 0° and no defects.

serves as a first approximation for understanding charge transport as related to the microstructural arrangement of the graphene sheets/cones in the CNFs.

$$\rho(\theta) = \rho_a \sin^2\theta + \rho_c \cos^2\theta. \tag{5.11}$$

In Equation (5.11), $\theta = 90 - \alpha$, whereby α is the cone angle. Typical PECVD growth affords cone angles of 5–20°, which give a range of resistivities between 3.0×10^{-4}–6.0×10^{-3}. Even with low cone angles the resistivity is too high for implementation as interconnects as described earlier. However, increasing the number of contributing channels will drastically increase the conductance. In order to use carbon nanofibers then, we need to minimize cone angle and maximize the number of contributing channels.

While PECVD growth of vertically oriented CNTs may not achieve the ideal MWCNT structure ($\alpha = 0°$), it is possible to achieve defect-free low cone angle ($\alpha < 2°$) carbon nanofibers that would provide many ballistic transport channels at short lengths (<2 μm depending upon cone angle). For instance, if many nanocones (building blocks of carbon nanofibers) were stacked inside each other possessing small cone angles then we would obtain many conducting channels as long as the graphene sheets do not terminate prior to contacting the bottom and top contacts. To illustrate this concept further, we calculated nanocone heights achievable for given nanocone diameter and cone angle (see Figure 5.10). For cone angles $\alpha \leq 4°$ and carbon nanocone diameters between 30–100 nm, single nanocone heights within the CNFs are greater than 200 nm. Thus, it is conceivable to envision a CNF solution for the 45 nm node whereby a 30 nm diameter nanocone with a cone angle of 3° would have a height in excess of 280 nm. This would achieve an aspect ratio of nearly 10:1, which far surpasses the aspect ratio requirements for local level interconnects for the 45 nm node, which is currently identified as 1.8. To illustrate this concept more clearly as a solution, it is also necessary to calculate the number of contributing channels or graphene sheets that would traverse the bottom and top contacts in a local wiring implementation. In this case, we assume a cone angle of 3°, an outer nanocone diameter of 45 nm, and an inner nanocone diameter of 30 nm. Then, to obtain the number of nanocones we subtract the diameter of the innermost nanocone from the outermost nanocone and divide by the spacing between graphene sheets (∼0.34 nm). This gives ∼45 nanocones/contributing channels for this embodiment. Alternatively, the use of the top-down approach described earlier [106] whereby CNTs are grown from via

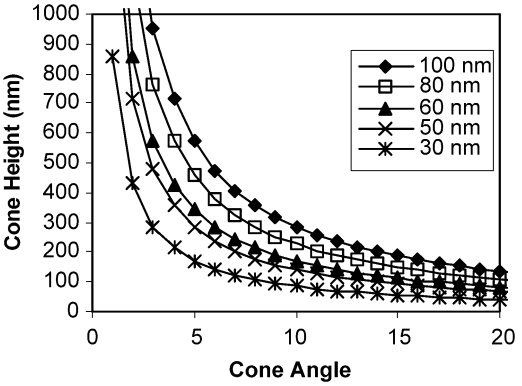


FIGURE 5.10. Calculated cone height vs cone angle for different carbon nanocone diameters.

holes could achieve a similar number of conducting channels, although it is not clear that there will indeed be dry etch processes for via hole diameters less than 45 nm. Regardless of the approach implemented (bottom-up vs top-down), maximizing the number of contributing channels and minimizing or eliminating defects is critical to achieve the intrinsic ballistic conductance needed for local wiring solutions. For intermediate and global wiring solutions, it appears that size constraints are less of a concern and instead line resistance is more of a concern.

5.4.2. Interfacial Contact Materials

A critical challenge facing the development of most nanostructure based electronic devices is the reliable engineering of electrical contacts and annealing of defects in nanomaterial building blocks. Repairing point defects that occur within low-dimensional nanomaterials as well as engineering the nanoscale interfacial contacts is crucial for the realization of high performance devices. Attaining reliable contacts to nanostructures allows for the detailed investigation of electron and phonon transport and heat dissipation phenomena intrinsic to the materials. Furthermore, understanding performance and achieving reliable operation is critical for the ultra large-scale integration (ULSI) of nanostructures.

The resistance of a single MWCNT measured with CSAFM such as that presented in Figure 5.6 is about 50 to 300 kΩ, which is dominated by poor contact. A larger probe of ∼25 μm diameter engaged with higher pressure significantly reduces the resistance to ∼20 kΩ per MWCNT (Figure 5.8). Using the 10×10 μm^2 Cr contact pads physically deposited on the planarized surface, the resistance is further reduced to 3–15 kΩ per MWCNT (Figures 5.7 and 5.8). Clearly, the metal–CNT contact needs to be optimized before the intrinsic properties of CNTs can be obtained. For most studies, the metal contacts are made to the sidewall of CNTs and thus only the outermost shell was measured. With the processing flow shown in Scheme 5.1, the contact could be made to all the shells which is favorable for interconnect applications. However, the contact is only an atomic band to each shell and thus is very sensitive to the interface integrity. Theoretical studies [118,119] show that even the contact length between the metal and the nanotube in the side-contact configuration is critical for achieving low resistance. Particularly, the conductance drops dramatically when the contact length is less than 10 nm. For the point contact geometry of CSAFM described in Figure 5.6, it is not surprising to observe a large resistance. In

practice, a catalyst metal such as Fe, Co, or Ni may be deposited on top of the MWNTs before the deposition of the top metal line. Thermal annealing with the presence of the transition metal can improve the electrical contact between CNTs and metal lines [120]. A recent study by A. Javey et al. [14] indicated that Pd can make good contact with SWCNTs in a side-contact configuration. It may be also applied to the end-contact schemes for interconnect fabrication. The bottom metal contact also needs to be optimized.

However, the optimization of the contacts to vertically aligned MWCNTs has to be considered by integrating metals, dielectrics and other functional materials, with the parametric space in the PECVD growth of CNTs which consists of a number of factors such as growth temperature, plasma power, bias on the substrate, pressure, nature of the gaseous feedstock as well as catalyst and substrate composition. Of primary importance in the latter category is the screening of underlayer contact materials that are compatible with the catalyst metal and the growth process. Generally, for electronic device applications, contact materials with a close work function match to CNTs $\sim$5 eV [121] should be explored to minimize contact resistance while maximizing mechanical stability, even though it is still unclear how the nature of the CNT/metal interface affects the electron transport. The systematic exploration of metal contact materials that are compatible with CNT growth will undoubtedly lead to uncovering the nature of these interfacial problems.

It is important to note that for side-contacted nanotubes and end-contacted nanotubes, we expect distinctly different mechanisms of electron transport across the metal–CNT interface. Side-contacted CNTs depend upon tunneling of electrons across a finite physical barrier created by van der Waals interaction at the metal–CNT interface [122] to facilitate electron injection into a metal–CNT–metal system. The contact quality for end-contacted nanotubes depends on chemical (chemisorption) and/or physical (physisorption) processes at the nanotube tip. Some novel ideas for optimizing measurement techniques and minimizing contact resistance have been presented in previous studies along these lines. For instance, electron bombardment has been used to create disorder at the metal–CNT interface, introducing scattering centers, thereby enhancing transmission of electrons between the metal and CNT [123], producing single MWCNT resistance in the 3 kΩ range. In another study, a rope of SWCNTs was used to conduct four-point probe measurements yielding contact resistance down to 750 Ω by taking advantage of top and bottom contacting of electrodes to the nanotube rope [124]. Reference [124] effectively demonstrates a method for de-embedding probe resistance, subsequently reducing measured resistance. For the type of geometry presented in [123,124], tight-binding calculations [118,125] have shown that contact area affects the transmission of electrons across the side-contacted metal–CNT junction, hence directly impacting contact resistance. The configurations presented in [123,124] can be characterized as being side-contacted, similar to state-of-the-art on-chip copper interconnects.

The transport mechanism across a side-contacted MWCNT with gold electrodes is explained in Reference [126] as tunneling across an energy barrier created by the finite separation between the gold electrode and MWCNT. Combining this theory with observations that the wave function orthogonal to the nanotube axis decays rapidly away from the CNT surface [118], we conclude that electron injection into the nanotube is accomplished by tunneling through this mechanically induced energy barrier at the metal–CNT junction. Due to the fact that the physical separation is on the order of the bond length, a few angstroms, between carbon and the metal electrode material, tunneling depends on the chemical composition and configuration of electronic states at the surface. Since the wave function overlap is reduced by this gap-induced energy barrier, transmission is decreased

for side-contacted carbon nanotubes with small contact area. For large contact area, electronic structure calculations suggest better electronic coupling between the electrode material and the MWNT [118,127].

5.4.3. End-contacted Metal–CNT Junction

Similar to the side-contacted geometry, the chemistry at the end-contacted metal–CNT junction must be well understood [125]. The TEM analysis of the MWNT tip structure allows us to understand how the physical properties of the nanotube affect electron movement through metal–CNT–metal systems. The presence of dangling bonds or the oxygen-containing groups at the end of a MWCNT provides an opportunity for chemisorption of the carbon nanotube by the electrode material to occur. The nature of this process and the number of shells (i.e., walls) that bond with the metal will determine the amount of charge transfer between the metal and the CNT. The end-contacted structure shown in Figure 5.5 represents a possible architecture for using MWCNTs as on-chip interconnects. The size and vertical nature of the aligned array make them good candidates for use as vias, bypassing processing problems for high-aspect-ratio structures with current silicon-based technology, while retaining desirable electronic properties. In addition, this bottom-up approach for synthesizing CNTs allows for silicon-compatible processing, ideal for integration into current wafer manufacturing processes. We also must discuss some of the drawbacks for implementing such a structure into IC design in addition to the bamboo-like structure of the nanotube grown by PECVD. As reported in References [109] and [110], these structures are limited by the mechanical strength of the nanotubes, restricting the length of the MWNT wires to $\sim$10 μm, too short for most practical wiring applications. As MWNT growth processes mature, the initial step for implementing them into silicon-based systems would be to replace copper via technology with a similar CNT-based via structure [105]. An alternative to more advanced growth processes is to more thoroughly investigate post-growth annealing as well as higher temperature growth to reduce the resistance.

5.4.4. Thermal Stress Characteristics

The fabrication of an IC chip requires many thermal cycles from room temperature to 350–450°C, during which large thermal stresses are introduced due to thermal expansion mismatches among the interconnect, dielectrics, and the substrate materials. The build up of large stresses at the Cu lines poises a big problem affecting the structural integrity and long term stability of current Cu damascene technology [128]. These problems can be significantly reduced with CNT interconnects. Table 5.1 lists the CTE and Young's modulus of CNTs in comparison to those of the Si<100> substrate, Cu interconnects, and silicon dioxide derived from TEOS, and methyl silasesquioxane (MSQ). Clearly, the difference in CTE between CNTs and the surrounding TEOS SiO_2 dielectrics in the longitudinal direction is 10 times smaller than that between Cu wires and TEOS SiO_2. CNTs are more compatible with TEOS SiO_2 dielectrics than the low-k materials MSQ. The high mechanical strength of CNTs and SiO_2 make the whole interconnect system much more stable than Cu interconnects. The CTE of longitudinal CNTs and SiO_2 are also much closer to that of the Si substrate, which is desired for a more reliable system.

TABLE 5.1.
CTE and Young's modulus of materials used in interconnects.

	CTE (10^{-6}/°C)	Young's modulus (GPa)
Si(100)	2.61	131.0
Cu	17.7	115.0
CNT, longitudinal	−1.2–0	∼1000[a]
CNT, transverse	20–29	∼1000[a]
MSQ	20.0	6.3
TEOS	0.51	71.4
Ta	6.5	185.7

[a] From [129].
[b] Most of other data are from [128].

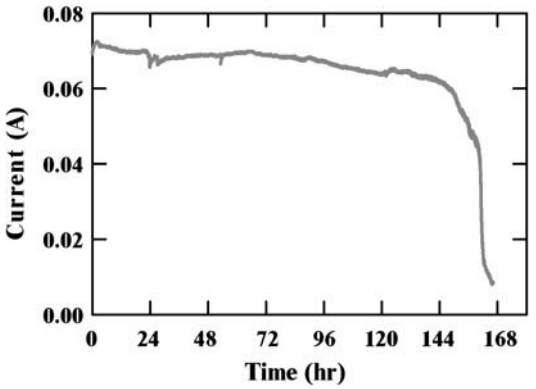


FIGURE 5.11. Current carrying capacity test for vertically oriented CNFs.

5.4.5. Reliability Test

We have also investigated the reliability of the vertically oriented MWCNT (or more precisely CNF) structures under a high current density in a two-terminal configuration. Figure 5.11 shows the current vs. time at a constant bias (5 V) over two 25 μm diameter probes contacted with the MWCNTs embedded in SiO_2 from the planarized top surface. At the initial state, the current density is conservatively estimated between 1×10^6–2×10^7 A/cm^2 by assuming 200–500 CNFs with an average diameter of ∼40 nm and length of 3–4 μm in parallel contact with the probes. The measurements are carried over a period of one week under ambient conditions. No severe degradation is seen until after day 6 when the current starts to decrease. We believe that the resistive heating gradually caused the degradation of the CNFs presumably at the top contact between the CNFs and the probe. The capability to carry such a high current density over so many days already surpasses any available interconnect materials and meets the requirements listed in the ITRS [112]. Further optimization may reach a stable current density as high as 1×10^{10} A/cm^2 as reported with the high-quality MWCNTs [109].

5.5. SUMMARY

We have reviewed the unique materials properties of CNTs and described the current status in developing CNTs as potential interconnects in IC chips. Although this technology is largely in the conceptual stage, there is growing interest on the use of CNTs for wiring. The key hurdles facing the implementation of a CNT interconnect solution fall on the processing end of the technology as it appears the promise of this technology is well documented. Future process improvements should balance the use of the highest quality CNT material with the economy, scale and reliability of their manufacture. Achieving optimum interfacial contacts between CNTs and other metals is of particular importance as is processing/self-assembly techniques to control the size, placement and geometry of the CNTs. With more effort from the microelectronics industries involved in this field, there is great potential that CNTs may move from materials research laboratories into real applications. It is hoped that this brief overview will serve as a stimulus for further development of CNT based interconnect technology.

REFERENCES

1. B.T. Kelly, Physics of Graphite, Applied Science, 1981.
2. R. Saito, G. Dresselhaus, and M.S. Dresselhaus, Physical Properties of Carbon Nanotubes, Imperial College Press, London, 1998.
3. K. Hata, D.N. Futaba, K. Mizuno, T. Namai, M. Yumura, and S. Iijima, Water-assisted highly efficient synthesis of impurity-free single-walled carbon nanotubes, Science, 306, p. 1362 (2004).
4. K.B.K. Teo, C. Singh, M. Chhowalla, and W.I. Milne, Catalytic synthesis of carbon nanotubes and nanofibers, Encyclopedia of Nanoscience and Nanotechnology, 1, pp. 665–686 (2004).
5. P.E. Nolan, D.C. Lynch, and A.H. Cutler, Carbon deposition and hydrocarbon formation on group VIII metal catalysts, J. Phys. Chem. B, 102, p. 4165 (1998).
6. L.M. Viculis, J.J. Mack, and R.B. Kaner, A chemical route to carbon nanoscrolls, Science, 299, p. 1361 (2003).
7. C.R. Martin, Nanomaterials—a membrane-based synthetic approach, Science, 266, p. 1961 (1994).
8. M. Dresselhaus, G. Dresselhaus, and P.C. Eklund, Science of Fullerenes and Carbon Nanotubes, Academic Press, San Diego, 1996.
9. D.H. Robertson, D.W. Brenner, and J.W. Mintmire, Energetics of nanoscale graphitic tubules, Phys. Rev. B, 45, p. 12592 (1992).
10. S. Sawada and N. Hamada, Energetics of carbon nano-tubes, Solid State Comm., 83, p. 917 (1992).
11. T.W. Odom, J. Huang, P. Kim, and C.M. Lieber, Structure and electronic properties of carbon nanotubes, J. Phys. Chem. B, 104, p. 2794 (2000).
12. M. Meyyappan, Carbon Nanotubes: Science and Applications, Boca Raton, CRC Press, 2004.
13. S.J. Tans, M.H. Devoret, H. Dai, A. Thess, R.E. Smalley, L.J. Geerligs, and C. Dekker, Individual single-wall carbon nanotubes as quantum wires, Nature, 386, p. 474 (1997).
14. A. Javey, J. Guo, Q. Wang, M. Lundstrom, and H. Dai, Ballistic carbon nanotube field-effect transistors, Nature, 424, p. 654 (2003).
15. D.J. Thouless, Maximum metallic resistance in thin wires, Phys. Rev. Lett., 39, p. 1167 (1977).
16. C.T. White and T.N. Todorov, Carbon nanotubes as long ballistic conductors, Nature, 393, p. 240 (1998).
17. A.G. Rinzler, et al., Large scale purification of single wall carbon nanotubes: process, product and characterization, Appl. Phys. A, 6, p. 29 (1998).
18. S. Frank, P. Poncharal, Z.L. Wang, and W.A. de Heer, Carbon nanotube quantum resistors, Science, 280, p. 1744 (1998).
19. P.G. Collins, M.S. Arnold, and P. Avouris, Engineering carbon nanotubes and nanotube circuits using electrical breakdown, Science, 292, p. 706 (2001).
20. P.J. de Pablo, E. Graugnard, B. Walsh, R.P. Andres, S. Datta, and R. Reifenberger, A simple, reliable technique for making electrical contact to multiwalled carbon nanotubes, Appl. Phys. Lett, 74, p. 323 (1999).
21. K.S. Novoselov, A.K. Geim, S.V. Morozov, D. Jiang, Y. Zhang, S.V. Dubonos, I.V. Grigorieva, and A.A. Firsov, Electric field effect in atomically thin carbon films, Science, 306, p. 666 (2004).

22. J. Lu and J. Han, Carbon nanotubes and nanotube-based nano devices, Int. J. High Speed Electron. Sys., 9, p. 101 (1998).
23. R.S. Ruoff, D. Qian, and W.K. Liu, Mechanical properties of carbon nanotubes: theoretical predictions and experimental measurements, C.R. Physique, 4, p. 993 (2003).
24. M.M.J. Treacy, et al., Exceptionally high Young's modulus observed for individual carbon nanotubes, Nature, 381, pp. 678–680 (1996).
25. D. Srivastava, M. Menon, and K. Cho, Nanoplasticity of single-walled carbon nanotubes under uniaxial compression, Phys. Rev. Lett., 83, p. 2973 (1999).
26. J.-P. Salvetat-Delmotte and A. Rubio, Mechanical properties of carbon nanotubes: a fiber digest for beginners, Carbon, 40, pp. 1729–1734 (2002).
27. D. Qian, G.J. Wagner, W.K. Liu, M.-F. Yu, and R.S. Ruoff, Mechanics of carbon nanotubes, Appl. Mech. Rev., 55, p. 495 (2002).
28. E.V. Barrera, Key methods for developing single wall nanotube composites, J. Mater., 52, p. 38 (2000).
29. J. Sandler, et al., Development of a dispersion process for carbon nanotubes in a epoxy matrix and the resulting electrical properties, Polymer, 40, p. 5967 (1999).
30. C. Park, et al., Dispersion of single wall carbon nanotubes by in situ polymerization under sonication, Chem. Phys. Lett., 364, p. 303 (2002).
31. J.-M. Benoit, et al., Transport properties of PMMA carbon nanotubes composites, Synth. Metals, 121, p. 1215 (2001).
32. R. Safdi, R. Andrews, and E.A. Grulke, Multiwalled carbon nanotube polymer composites: synthesis and characterization of thin films, J. Appl. Polymer Sci., 84, p. 2660 (2002).
33. B.E. Kilbride, et al., Experimental observation of scaling laws for alternating current and direct current conductivity in polymer-carbon nanotube composite thin films, J. Appl. Phys., 92, p. 4024 (2002).
34. M.S.P. Shaffer and A.H. Windle, Fabrication and characterization of carbon nanotube/poly(vinyl alcohol) composites, Adv. Mater., 11, p. 938 (1999).
35. G.-D. Zhan, et al., Electrical properties of nanoceramics reinforced with ropes of single-walled carbon nanotubes, Appl. Phys. Lett., 83, p. 1228 (2003).
36. R.Z. Ma, et al., Processing and properties of carbon nanotubes-nano-SiC ceramics, J. Mater. Sci., 33, p. 5243 (1998).
37. G.L. Hwang and K.C. Hwang, Carbon nanotube reinforced ceramics, J. Mater. Chem., 11, p. 1722 (2001).
38. J.-M. Ting and M.L. Lake, Vapor grown carbon fiber reinforced aluminum composites with very high thermal conductivity, J. Mater. Res., 10, p. 247 (1995).
39. S.R. Dong, et al., An investigation of the sliding wear behavior of Cu-matrix composite reinforced by carbon nanotubes, Mater. Sci. Eng. A, 313, p. 83 (2001).
40. R. Zong, et al., Fabrication of nano-Al based composites reinforced by single-walled carbon nanotubes, Carbon, 41, p. 848 (2003).
41. H. Dai, J.H. Hafner, A.G. Rinzler, D.T. Colbert, and R.E. Smalley, Nanotubes as nanoprobes in scanning probe microscopy, Nature, 384, p. 147 (1996).
42. S.S. Wong, E. Joselevich, A.T. Woolley, C.L. Cheung, and C. M Lieber, Covalently functionalized nanotubes as nanometre- sized probes in chemistry and biology, Nature, 394, p. 52 (1998).
43. C.V. Nguyen, K. Chao, R.M. Stevens, L. Delzeit, A. Cassell, J. Han, and M. Meyyappan, Carbon nanotube tip probes: stability and lateral resolution in scanning probe microscopy and application to surface science in semiconductors, Nanotechnol., 12, p. 363 (2001).
44. Q. Ye, A.M. Cassell, H. Liu, K.-J. Chao, J. Han, and M. Meyyappan, Large-scale fabrication of carbon nanotube probe tips for atomic force microscopy critical dimension imaging applications, Nano. Lett., 4, p. 1301 (2004).
45. M. Desquesnes, S.V. Rotkin, and N.R. Aluru, Calculation of pull-in voltages for carbon-nanotube-based nanoelectromechanical switches, Nanotechnol., 13, p. 120 (2002).
46. S.W. Lee, D.S. Lee, R.E. Morjan, S.H. Jhang, M. Sveningsson, O.A. Nerushev, Y.W. Park, and E.E.B. Campbell, A three-terminal carbon nanorelay, Nano. Lett., 4, p. 2027 (2004).
47. J.M. Kinaret, T. Nord, and S. Viefers, A carbon-nanotube-based nanorelay, Appl. Phys. Lett., 82, p. 1287 (2003).
48. P. Kim, L. Shi, A. Majumdar, and P.L. McCuen, Thermal transport measurements of individual multiwalled nanotubes, Phys. Rev. Lett., 87, p. 215502 (2001).
49. J. Hone, B. Batlogg, Z. Benes, A.T. Johnson, and J.E. Fischer, Quantized phonon spectrum of single-wall carbon nanotubes, Science, 289, p. 1730 (2000).
50. J. Hone, M.C. Llagun, M.J. Biercuk, A.T. Johnson, B. Batlogg, Z. Benes, and J.E. Fischer, Thermal properties of carbon nanotubes and nanotube based materials, Appl. Phys. A, 74, p. 339 (2002).

51. S.U.S. Choi, Z.G. Zhang, W. Yu, F.E. Lockwood, and E.A. Grulke, Anomalous thermal conductivity enhancement in nanotube suspensions, Appl. Phys. Lett., 79, p. 2252 (2001).
52. M.J. Biercuk, M.C. Llaguno, M. Radosavljevic, J.K. Hyun, A.T. Johnson, and J.E. Fischer, Carbon nanotube composites for thermal management, Appl. Phys. Lett., 80, p. 2767 (2002).
53. H.F. Chuang, S.M. Cooper, M. Meyyappan, and B.A. Cruden, Improvement of thermal contact resistance by carbon nanotubes and nanofibers, J. Nanosci. Nanotech., 4, p. 964 (2004).
54. Q. Ngo, B.A. Cruden, A.M. Cassell, G. Sims, M. Meyyappan, J. Li, and C.Y. Yang, Thermal interface properties of Cu-filled vertically aligned carbon nanofiber arrays, Nano. Lett., 4, p. 2403 (2004).
55. F. Entwisle, Thermal expansion of pyrolytic graphite, Phys. Lett., 2, pp. 236–238 (1962).
56. E.A. Heintz, The measurement of the coefficient of thermal expansion of graphite artefacts, Carbon, 28, p. 233 (1990).
57. H. Jiang, B. Liu, Y. Huang, and K.C. Hwang, Thermal expansion of single wall carbon nanotubes, J. Eng. Matr. And Technol., 126, p. 265 (2004).
58. S. Iijima, Helical microtubules of graphitic carbon, Nature, 354 p. 56 (1991).
59. T.W. Ebbesen and P.M. Ajayan, Large scale synthesis of carbon nanotubes, Nature, 358, p. 220 (1992).
60. A. Thess, et al., Crystalline ropes of metallic carbon nanotubes, Science, 273, p. 483 (1996).
61. R.O. Loutfy, et al., in E. Osawa, Ed., Perspectives Fullerene Nanotechnology, Kluwer, Dordrecht, 2002, p. 35.
62. J. Kong, H.T. Soh, A.M. Cassell, C.F. Quate, and H. Dai, Synthesis of individual single salled carbon nanotubes on patterned silicon wafers, Nature, 395, p. 878 (1998).
63. S.S. Fan, M.G. Chapline, N.R. Franklin, T.W. Tombler, A.M. Cassell, and H. Dai, Self-oriented regular arrays of carbon nanotubes and their field emission properties, Science, 283, p. 512 (1999).
64. A.M. Cassell, J.A. Raymakers, J. Kong, and H. Dai, Large-scale CVD synthesis of single walled carbon nanotubes, J. Phys. Chem. B, 103, p. 6484 (1999).
65. H. Dai, et al., Single-wall nanotubes produced by metal-catalyzed disproportionation of carbon monoxide, Chem. Phys. Lett., 260, p. 471 (1996).
66. B.C. Satishkumar, A. Govindraj, R. Sen, and C.N.R. Rao, Single-walled nanotubes by the pyrolysis of acetylene-organometallic mixtures, Chem. Phys. Lett., 293, p. 47 (1998).
67. P. Nikolaev, et al., Gas-phase catalytic growth of single-walled carbon nanotubes from carbon monoxide, Chem. Phys. Lett., 313, p. 91 (1999).
68. Y. Avigal and R. Kalish, Growth of aligned carbon nanotubes by biasing during growth, Appl. Phys. Lett., 78, p. 2291 (2001).
69. Y. Zhang, et al., Electric-field-directed growth of aligned single-walled carbon nanotubes, Appl. Phys. Lett., 79, p. 3155 (2001).
70. M. Meyyappan, L. Delzeit, A. Cassell, and D. Hash, Carbon nanotube growth by PECVD: a review, Plasma Sources Sci. Technol., 12, p. 205 (2003).
71. L. Delzeit, et al., Growth of multiwall carbon nanotubes in an inductively coupled plasma reactor, J. Appl. Phys., 91, p. 6027 (2002).
72. K. Matthews, B. Cruden, B. Chen, M. Meyyappan, and L. Delzeit, Plasma enhanced chemical vapor deposition of multiwalled carbon nanofibers, J. Nanosci. Nanotech., 2, p. 475 (2002).
73. G.W. Ho, A.T.S. Wee, J. Lin, and W.C. Tjiu, Synthesis of well-aligned multiwalled carbon nanotubes on Ni catalyst using radio frequency plasma-enhanced chemical vapor deposition, Thin Solid Films, 388, p. 73 (2001).
74. H. Ishida, et al., Experimental study of fullerene-family formation using radio-frequency-discharge reactive plasmas, Thin Solid Films, 407, p. 26 (2002).
75. N. Satake, et al., Production of carbon nanotubes by controlling radio-frequency glow discharge with reactive gases, Physica B, 323, p. 290 (2002).
76. Y.H. Wang, et al., Synthesis of large area aligned carbon nanotube arrays from C_2H_2–H_2 mixture by rf plasma-enhanced chemical vapor deposition, Appl. Phys. Lett., 79, p. 680 (2001).
77. L. Valentini, et al., Formation of carbon nanotubes by plasma enhanced chemical vapor deposition: role of nitrogen and catalyst layer thickness, J. Appl. Phys., 92, p. 6188 (2002).
78. B.O. Boskovic, et al., Large-area synthesis of carbon nanofibres at room temperature, Nature Mater., 1, p. 165 (2002).
79. L.C. Qin, D. Zhou, A.R. Krauss, and D.M. Gruen, Growing carbon nanotubes by microwave plasma-enhanced chemical vapor deposition, Appl. Phys. Lett., 72, p. 3437 (1998).
80. O. Kuttel, et al., Electron field emission from phase pure nanotube films grown in a methane/hydrogen plasma, Appl. Phys. Lett., 73, p. 2113 (1998).

81. S.H. Tsai, C.W. Chao, C.L. Lee, and H.C. Shin, Bias-enhanced nucleation and growth of the aligned carbon nanotubes with open ends under microwave plasma synthesis, Appl. Phys. Lett., 74, p. 3462 (1999).
82. Q. Zhang, et al., Carbon films with high density nanotubes produced using microwave plasma assisted CVD, J. Phys. Chem. Solids, 61, p. 1179 (2000).
83. Y.C. Choi, et al., Growth of carbon nanotubes by microwave plasma-enhanced chemical vapor deposition at low temperature, J. Vac. Sci. Technol. A, 18, p. 1864 (2000).
84. Y.C. Choi, et al., Effect of surface morphology of Ni thin film on the growth of aligned carbon nanotubes by microwave plasma-enhanced chemical vapor deposition, J. Appl. Phys., 88, p. 4898 (2000).
85. M. Okai, T. Muneyoshi, T. Yaguchi, and S. Sasaki, Structure of carbon nanotubes grown by microwave-plasma-enhanced chemical vapor deposition, Appl. Phys. Lett., 77, p. 3468 (2000).
86. C. Bower, W. Zhu, S. Jin, and O. Zhou, Plasma-induced alignment of carbon nanotubes, Appl. Phys. Lett., 77, p. 830 (2000).
87. H. Cui, O. Zhou, and B.R. Stoner, Deposition of aligned bamboo-like carbon nanotubes via microwave plasma enhanced chemical vapor deposition, J. Appl. Phys., 88, p. 6072 (2000).
88. Y. Chen, et al., Well-aligned graphitic nanofibers synthesized by plasma-assisted chemical vapor deposition, Chem. Phys. Lett., 272, p. 178 (1997).
89. Y. Chen, L.P. Guo, D.J. Johnson, and R.H. Prince, Plasma-induced low-temperature growth of graphitic nanofibers on nickel substrates, J. Cryst. Growth, 193, p. 342 (1998).
90. Z.F. Ren, et al., Synthesis of large arrays of well-aligned carbon nanotubes on glass, Science, 282, p. 1105 (1998).
91. J. Han, et al., Growth and emission characteristics of vertically well-aligned carbon nanotubes grown on glass substrate by hot filament plasma-enhanced chemical vapor deposition, J. Appl. Phys., 88, p. 7363 (2000).
92. Y. Hayashi, T. Negishi, and S. Nishino, Growth of well-aligned carbon nanotubes on nickel by hot-filament-assisted dc plasma chemical vapor deposition in a CH_4/H_2 plasma, J. Vac. Sci. Technol. A, 19, p. 1796 (2001).
93. Z.P. Huang, et al., Effect of nickel, iron and cobalt on growth of aligned carbon nanotubes, Appl. Phys. A, 74, p. 387 (2002).
94. B.A. Cruden, A.M. Cassell, Q. Ye, and M. Meyyappan, Reactor design considerations in the hot filament/direct current plasma synthesis of carbon nanofibers, J. Appl. Phys., 94, p. 4070 (2003).
95. V.I. Merkulov, et al., Patterned growth of individual and multiple vertically aligned carbon nanofibers, Appl. Phys. Lett., 76, p. 3555 (2000).
96. V.I. Merkulov, et al., Shaping carbon nanostructures by controlling the synthesis process, Appl. Phys. Lett., 79, p. 1178 (2001).
97. V.I. Merkulov, et al., Alignment mechanism of carbon nanofibers produced by plasma-enhanced chemical-vapor deposition, Appl. Phys. Lett., 79, p. 2970 (2001).
98. K.B.K. Teo, et al., Uniform patterned growth of carbon nanotubes without surface carbon, Appl. Phys. Lett., 79, p. 1534 (2001).
99. M. Chhowalla, et al., Growth process conditions of vertically aligned carbon nanotubes using plasma enhanced chemical vapor deposition, J. Appl. Phys., 90, p. 5308 (2001).
100. K.B.K. Teo, et al., Characterization of plasma-enhanced chemical vapor deposition carbon nanotubes by Auger electron spectroscopy, J. Vac. Sci. Technol., B, 20, p. 116 (2002).
101. M. Tanemura, et al., Growth of aligned carbon nanotubes by plasma-enhanced chemical vapor deposition: optimization of growth parameters, J. Appl. Phys., 90, p. 1529 (2001).
102. J. Han, et al., NH_3 effect on the growth of carbon nanotubes on glass substrate in plasma enhanced chemical vapor deposition, Thin Solid Films, 409, p. 120 (2002).
103. J. Han, et al., Tip growth model of carbon tubules grown on the glass substrate by plasma enhanced chemical vapor deposition, J. Appl. Phys., 91, p. 483 (2002).
104. Y.Y. Wei, et al., Effect of catalyst film thickness on carbon nanotube growth by selective area chemical vapor deposition, Appl. Phys. Lett., 78, p. 1394 (2001).
105. F. Kreupl, A.P. Graham, G.S. Duesberg, W. Steinhögl, M. Liebau, E. Unger, and W. Hönlein, Carbon nanotubes in interconnect applications, Microelectronic Engineering, 64, p. 399 (2002).
106. G.S. Duesberg et al., Growth of isolated carbon nanotubes with lithographically defined diameter and location, Nano. Lett., 3, p. 257 (2003).
107. M. Nihei, A. Kawabata, and Y. Awano, Direct diameter-controlled growth of multiwall carbon nanotubes on nickel-silicide layer, Jpn. J. Appl. Phys., 42(6B), pp. L721–L723 (2003).
108. M. Horibe, M. Nihei, D. Kondo, A. Kawabata, and Y. Awano, Mechanical polishing technique for carbon nanotube interconnects in ULSIs, Jpn. J. Appl. Phys., 43(9A), p. 6499 (2004).

109. B.Q. Wei, R. Vajtai, and P.M. Ajayan, Reliability and current carrying capacity of carbon nanotubes, Appl. Phys. Lett., 79, p. 1172 (2001).
110. J. Li, et al., Electronic properties of multiwalled carbon nanotubes in an embedded vertical array, Appl. Phys. Lett., 81, p. 910 (2002).
111. J. Li, et al., A bottom-up approach for carbon nanotube interconnects, Appl. Phys. Lett., 82, p. 2491 (2003).
112. International Technology Roadmap For Semiconductors (ITRS), Edition 2003.
113. P. Kapur, J.P. McVittie, and K.C. Saraswat, Technology and reliability constrained future copper interconnects. I. Resistance modeling, IEEE Trans. Elect. Dev., 49, p. 590 (2002).
114. P. Kapur, G. Chandra, J.P. McVittie, and K.C. Saraswat, Technology and reliability constrained future copper interconnects. II. Performance implications, IEEE Trans. Elect. Dev., 49, p. 598 (2002).
115. H. Hwang, M. Meyyappan, G.S. Mathad, and R. Ranade, Simulations and experiments of etching of silicon in HBr plasmas for high aspect ratio features, J. Vac. Sci. Technol. B, 20, p. 2199 (2002).
116. M.S. Dresselhaus, G. Dresselhaus, K. Sugihara, I.L. Spain, and H.A. Goldberg, in M. Cardona, Ed., Graphite Fibers and Filaments, Springer Series in Materials Science, Vol. 5, New York, 1988, pp. 188–202.
117. L. Zhang, D. Austin, V.I. Merkulov, A.V. Meleshko, K.L. Klein, M.A. Guillorn, D.H. Lowndes, and M.L. Simpson, Four-probe charge transport measurements on individual vertically aligned carbon nanofibers, Appl. Phys. Lett., 84, p. 3972 (2004).
118. M.P. Anantram, S. Datta, and Y. Xue, Coupling of carbon nanotubes to metallic contacts, Phys. Rev. B, 61, p. 14219 (2000).
119. N. Mingo and J. Han, Conductance of metallic carbon nanotubes dipped into metal, Phys. Rev. B, 64, p. 201401 (2001).
120. R. Rosen, W. Simendinger, C. Debbault, H. Shimoda, L. Fleming, B. Stoner, and O. Zhou, Application of carbon nanotubes as electrodes in gas discharge tubes, Appl. Phys. Lett., 76, p. 1668 (2000).
121. K.B.K. Teo, et al., Plasma enhanced chemical vapour deposition carbon nanotubes/nanofibres—how uniform do they grow?, Nanotechnology, 14, p. 204 (2003).
122. A. Rochefort, P. Avouris, F. Lesage, and D.R. Salahub, Electrical and mechanical properties of distorted carbon nanotubes, Phys. Rev. B, Condens. Matter, 60, p. 13824 (1999).
123. A. Bachtold, M. Henny, C. Terrier, C. Strunk, C. Schonenberger, J.-P. Salvetat, J.-M. Bonard, and L. Forro, Contacting carbon nanotubes selectively with low-ohmic contacts for four-probe electric measurements, Appl. Phys. Lett., 73, p. 274 (1998).
124. J. Appenszeller, R. Martel, P. Avouris, H. Stahl, and B. Lengeler, Optimized contact configuration for the study of transport phenomena in ropes of single-wall carbon nanotubes, Appl. Phys. Lett., 78, p. 3313 (2001).
125. Q. Ngo, D. Petranovic, S. Krishnan, A.M. Cassell, Q. Ye, J. Li, M. Meyyappan, and C.Y. Yang, Electron transport through metal-multiwall carbon nanotube interfaces, IEEE Trans. Nanotechnol., 3, p. 311 (2004).
126. J. Tersoff, Contact resistance of carbon nanotubes, Appl. Phys. Lett., 74, p. 2122 (1999).
127. M.P. Anantram, Which nanowire couples better electrically to a metal contact: armchair or zigzag nanotube? Appl. Phys. Lett., 78, p. 2055 (2001).
128. S.-H. Rhee, Y. Du, and P.S. Ho, Thermal stress characteristics of Cu/oxide and Cu/low-k submicron interconnect structures, J. Appl. Phys., 93, p. 3926 (2003).
129. H.J. Qi, K.B.K. Teo, K.K.S. Lau, M.C. Boyce, W.I. Milne, J. Robertson, and K.K. Gleason, Determination of mechanical properties of carbon nanotubes and vertically aligned carbon nanotube forests using nanoindentation, J. Mech. Phys. Solids, 51, p. 2213 (2003).

6

Virtual Thermo-Mechanical Prototyping of Microelectronics and Microsystems

A. Wymysłowski[a], G.Q. Zhang[b], W.D. van Driel[b], and L.J. Ernst[c]

[a]*Wroclaw University of Technology, Poland*
[b]*Delft University of Technology, and Philips Semiconductors, The Netherlands*
[c]*Delft University of Technology, The Netherlands*

6.1. INTRODUCTION

The technological trends of microelectronics and microsystems are mainly characterized by miniaturization down to nano-scale, increasing levels of technology and function integration and eco-designing, while the business trends are mainly characterized by cost reduction, short-time-to-market and outsourcing. These trends together lead to increased chances and consequences of failures, increased design complexity, dramatically decreased design margins and increased difficulty to meet quality, robustness, reliability and shorter-time-to-market requirements. Among others, it is found that thermo-mechanical (thermal, mechanical and thermo-mechanical) related failures account for about 65% of total failures in microelectronics. Thermo-mechanical reliability is becoming one of the major concerns for both current and future microelectronics technologies.

Based on the root cause analyses from observed failures of microelectronics during different life cycles, it is clear that most of the thermo-mechanical reliability problems originate from the product/process design phase. However, within electronics industry, microelectronics design and qualification are still largely depending on one's experience. Often, up to 10 cycles (material development/pre-selection, concept designing, building and testing multiple physical prototypes) are needed, with some qualitative support from numerical simulations. Quality, robustness and reliability are usually dealt with after-physical prototyping, wherein reliability qualification testing with duration of 6 months is no exception. Clearly, this experience-based design and qualification method cannot lead to competitive design with short time-to-market, optimized performance, low costs, and guaranteed quality, robustness and reliability. Therefore, there is an urgent need to develop innovative thermo-mechanical design method—virtual thermo-mechanical prototyping method.

For virtual thermo-mechanical prototyping, "accurate and efficient prediction models" and "advanced simulation-based optimization methods" are the two core building

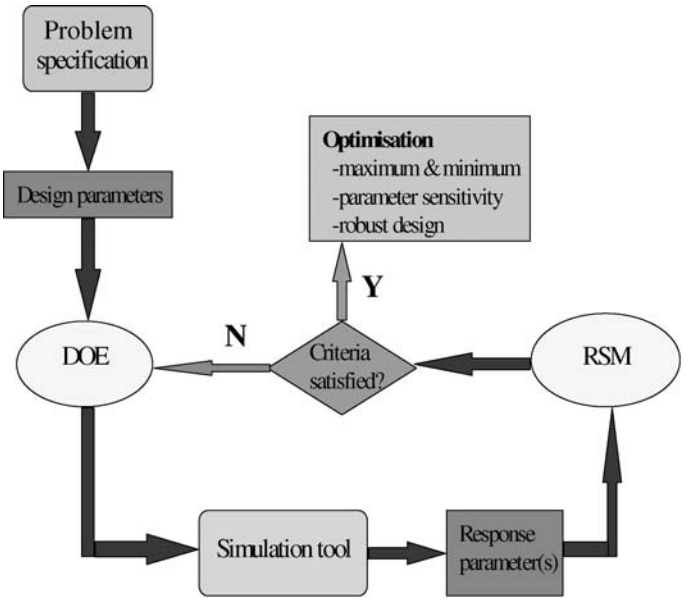


FIGURE 6.1. Typical virtual prototyping diagram.

blocks. By combining these two building blocks in a proper way, one can predict, qualify and optimize the thermo-mechanical behavior and/or trends of microelectronics against the actual requirements prior to major physical prototyping, manufacturing investments and reliability qualification tests. A typical virtual prototyping procedure would consist of the following steps (see Figures 6.1 and 6.2):

- building up a numerical model (such as FEM) of a product, which will be capable of predicting the behavior of the product under given loading conditions,
- carrying out a Design of Experiments (DoE) procedure in order to find out the correlation between the response and input factors (including interactions) and defining their significance in a sense of e.g., mean and variance,
- performing a Response Surface Modeling (RSM) that would approximate the model of the response in a form of a response surface reflecting relationship between the response and significant input factors,
- final stage should be directed towards design optimization, such as finding minimum/maximum, robust design and tolerance design [1,6,39].

6.2. PHYSICAL ASPECTS FOR NUMERICAL SIMULATIONS

In most of the engineer applications we are faced with the field theory problem of continuum domains, such as:

- Thermal fields.
- Mechanical/ structural fields.
- Electromagnetic fields.
- Fluid flow.

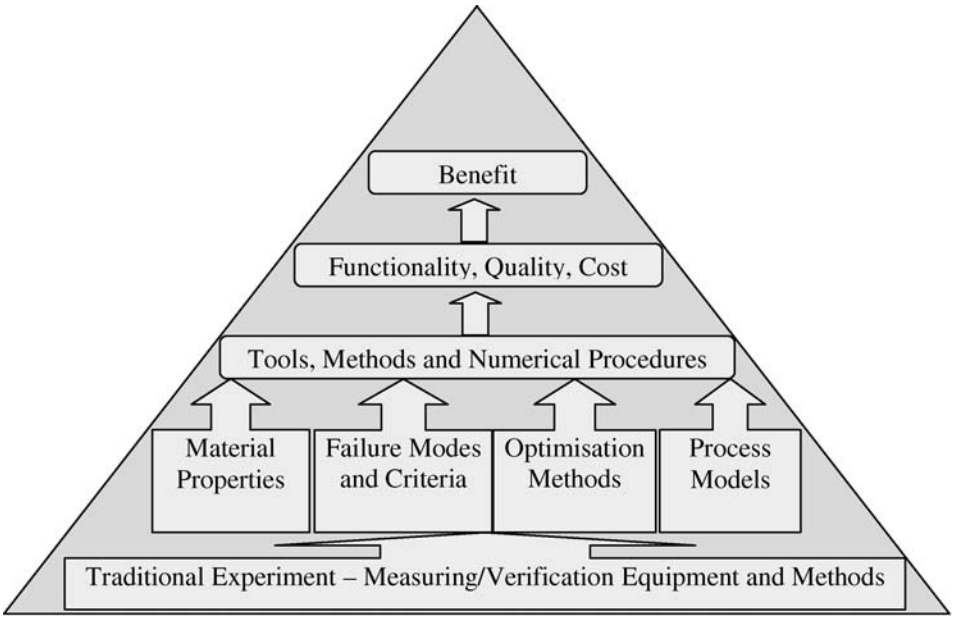


FIGURE 6.2. Advanced virtual prototyping structure.

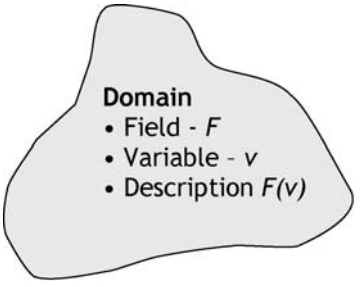


FIGURE 6.3. Continuum domain with defined physical field and its variable.

The continuum can be defined as a part of domain (e.g., solid state, fluid or gas) where the physical phenomenon is defined. The domain is defined by a selected physical field F and its variable v (e.g., thermal field and temperature T), which can receive any finite value in arbitrary point in the domain:

$$v = F(x, y, z). \tag{6.1}$$

The field variable v can receive infinite values in the domain as it is a function of field point coordinates x, y, z, see Figure 6.3.

Most of the problems in engineering practice, concerning field theory, can be described by partial differential equations (PDE). As the analytical solutions of most engineering problems seldom exit, more and more attention is directed towards numerical solutions rather than exact closed-form analytical solutions. The basic problem concerning numerical solution is the accuracy. It is a rule of thumb that numerical simulation is as

precise as the input data (GIGO—garbage in garbage out). There are a number of aspects that would influence accuracy, e.g.,

- Knowledge on physical phenomena.
- Material models and properties.
- Damage models and criteria.
- Product parameters.
- Process details.
- The best engineer practice techniques on: boundary conditions, loads, etc.

6.2.1. Numerical Modeling

Over the years, several numerical analysis methods for solving PDE equations have evolved. The most common ones are: Finite Difference Method (FDM) or Finite Element Method (FEM); the main difference is in representing complex geometrical shapes. Any numerical technique involves dividing the domain of solution into a finite number of simple sub-domains and allows obtaining approximate solutions to a wide variety of engineering problems with boundary-value problems. As an example of the most popular partial differential equations applied in engineering is Poisson's formula for evaluation of electrostatic field distribution:

$$\nabla^2 V(x, y, z) = \frac{q(x, y, z)}{\varepsilon_0}, \tag{6.2}$$

where $q(x, y, z)$ is the space charge in the analyzed region while ε_0 is the vacuum dielectric constant. If space charge is equal to 0, the above formula can be rewritten as follows

$$\nabla^2 V(x, y, z) = 0, \tag{6.3}$$

where ∇^2 stands for the Laplace's operator.

FDM method is based on the assumption that differential equations can be replaced by finite differences. According to the Taylor-series expansion, if we know the function value in point x we can evaluate the function value at point $x + h$ as long as the function is continuous and has derivatives:

$$f(x + h) = f(x) + \frac{h}{1!} f'(x) + \cdots. \tag{6.4}$$

Therefore the first derivate of the function can be replaced by the finite difference as:

$$f'(x) \cong \frac{\Delta f(x)}{h}, \tag{6.5}$$

where the error (for the central difference) will be proportional to

$$error \cong \frac{1}{27} f'''(x) \cdot h^2. \tag{6.6}$$

On the other hand FEM methods use the variational concept and approximations of the solution over the collection of finite elements. It is a numerical method for solving problems

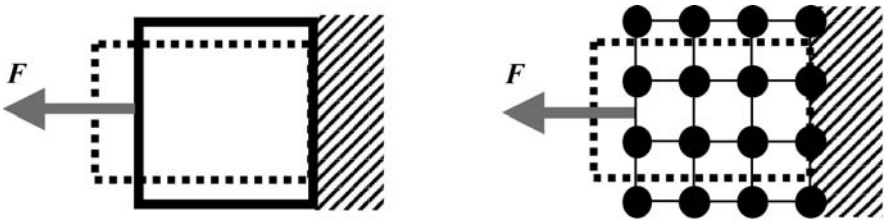


FIGURE 6.4. Problem of mechanical deformation due to force F. Example of the Finite Element approach; nodal points interconnect elements.

of engineering and mathematical physics and is especially useful for problems with complicated geometries, loadings, and material properties where analytical solutions cannot be obtained. Basically FEM method is applied in such applications as: structural analysis, fluid flow, heat transfer, electro-magnetic fields, acoustics, etc. A geometrically similar model consisting of multiple, linked, simplified representations of discrete regions, i.e., finite elements, represents the system. These units (finite elements) are interconnected at points common to two or more elements (nodes or nodal points), boundary lines and surfaces. An example problem of mechanical deformation due to the applied force F is presented in Figure 6.4.

The key problem in FEM method is proper definition of element properties, which in the above case can be defined as a spring constant. The spring constant or stiffness binds the applied force F and nodal displacement d according to the equation:

$$F = k \cdot d. \tag{6.7}$$

In fact, FEM method is based on defining the local $[\boldsymbol{k}]$ and global stiffness $[\boldsymbol{K}]$ matrix for one and number of elements respectively. The goal is to evaluate nodal displacements d_i. In case of one element the following governing equation applies:

$$\begin{aligned} F_1 &= kd_1 - kd_2, \\ F_2 &= -kd_1 + kd_2^2, \end{aligned} \tag{6.8}$$

which can be rewritten in a matrix form as:

$$\begin{bmatrix} k & -k \\ -k & k \end{bmatrix} \begin{Bmatrix} d_1 \\ d_2 \end{Bmatrix} = \begin{Bmatrix} F_1 \\ F_2 \end{Bmatrix}, \tag{6.9}$$

or

$$[\boldsymbol{k}]\{d\} = \{F\}, \tag{6.10}$$

where $[\boldsymbol{k}]$ can be referred as the local stiffness matrix.

In case of two elements the governing equation is given by (see Figure 6.5):

$$\begin{aligned} F_1 &= kd_1 - kd_2, \\ F_2 &= -kd_1 + kd_2 + kd_2 - kd_3, \\ F_3 &= -kd_2 + kd_3, \end{aligned} \tag{6.11}$$

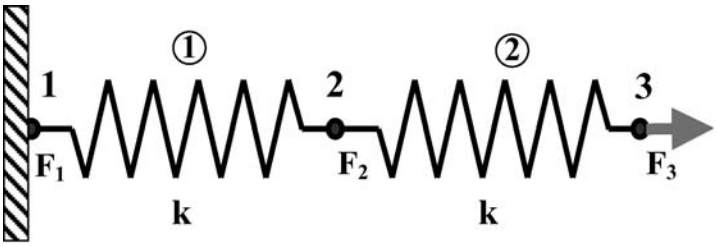


FIGURE 6.5. Global stiffness matrix.

$$\begin{Bmatrix} F_1 \\ F_2 \\ F_3 \end{Bmatrix} = \begin{bmatrix} k & -k & 0 \\ -k & k+k & -k \\ 0 & -k & k \end{bmatrix} \begin{Bmatrix} d_1 \\ d_2 \\ d_3 \end{Bmatrix}$$

FIGURE 6.6. Construction of the global stiffness matrix.

which can be rewritten in a matrix form as:

$$\begin{Bmatrix} F_1 \\ F_2 \\ F_3 \end{Bmatrix} = \begin{bmatrix} k & -k & 0 \\ -k & k+k & -k \\ 0 & -k & k \end{bmatrix} \begin{Bmatrix} d_1 \\ d_2 \\ d_3 \end{Bmatrix}, \tag{6.12}$$

or

$$[\boldsymbol{K}]\{d\} = \{F\}, \tag{6.13}$$

where $[\boldsymbol{K}]$ can be referred as the global stiffness matrix. In fact the global stiffness matrix can be constructed from local stiffness matrices according to the element interconnections. Figure 6.6 shows visualization of the stiffness matrix construction.

Solution methods in FEM are based on finding the equilibrium of minimum potential energy. The total potential energy E_p is defined as the sum of internal strain energy U_i and external U_e potential energy due to external forces $\boldsymbol{F}$:

$$E_p = U_i + U_e, \tag{6.14}$$

where

$$\begin{aligned} U_i &= \frac{1}{2}kx^2, \\ U_e &= -Fx, \end{aligned} \tag{6.15}$$

thus

$$E_p = \frac{1}{2}kx^2 - Fx. \tag{6.16}$$

Therefore the solution can be fund out by applying the variational analysis:

$$\frac{dE_p}{dx} = 0. \tag{6.17}$$

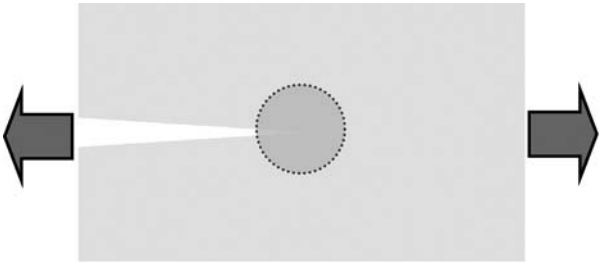


FIGURE 6.7. The singularity problem.

Basic steps involved in the numerical simulations based on FEM method can be summarized as follows [48]:

- Discretize and select element type.
- Select a displacement function.
- Define strain/displacement and stress/strain relationships.
- Derive element stiffness matrix and equations.
- Assemble equations and introduce B.C.'s (boundary conditions).
- Solve for the unknown degrees of freedom.
- Solve for element stresses and strains.
- Interpret the results.

The accuracy of numerical simulations seems to be a very important aspect, which may decide on its application. The main sources of errors are due to [8]:

- Material properties and models.
- FEM method.
- Simplification of multiphysics phenomena.
- Geometry accuracy and details.
- Mesh density and type.
- Time step for transient analysis.

The numerical modeling requires reducing the geometric complexity to the appropriate level of the problem—some features may be omitted. Sometimes the simplification can lead to singularity problems, e.g., concerning stresses at sharp edges due to infinite stress concentration, see Figure 6.7.

The evaluated value at the point depends on e.g., mesh density; however the displacements can be still good. There are a number of tasks that can be taken to avoid or minimize it:

- If the region is not important it can be ignored.
- If the region is important a detailed geometry should be included.
- The region can be modeled using sub-modeling technique.
- Extrapolating the results in the nearby region, e.g., linear.

6.2.2. Material Properties and Models

One of the basic problems of numerical modeling is connected with proper and accurate definition of material properties and models. The most common errors are due to:

- Material non-linearity.
- Thin layers.
- Temperature dependence.
- Process dependence.
- Internal discrepancies.
- Microstructure.
- Anisotropy.

The primary criterion of precise numerical modeling is knowledge on material models, characteristics and properties. Most of the engineer calculations are performed with the assumptions on elastic, linear, isotropic and isothermal behavior of materials. While this assumption can be justified in the macro scale it is not any longer applicable in the micro scale. This refers especially to microsystem and packaging of electronic devices (e.g., thin layers or grain-size). There are number of factors, which determine the description of material properties and can be described according to the following groups, as listed in Figure 6.8:

- Physical: density, melting temperature, etc.
- Mechanical: Young's modulus, shear modulus, Poisson's ratio, etc.
- Thermal: coefficient of thermal expansion, thermal conductivity, etc.
- Electromagnetic: specific resistivity, dielectric constant, etc.
- Acoustic: compression wave velocity, shear wave velocity, bar velocity, etc.

Material behavior will depend on:

- Type: metal, polymer, ceramic, etc.
- State: solid, liquids, gas.
- Form: crystalline, amorphous.
- Model: elastic, plastic, creep.

Constitutive models are very important for accurate stress/strain analysis, especially in case of damage analysis, e.g., fatigue assessment.

$$f(\sigma, \varepsilon, \dot{\varepsilon}, T) = 0, \quad d\sigma = \left(\frac{\partial \sigma}{\partial \varepsilon}\right)_{\dot{\varepsilon},T} d\varepsilon + \left(\frac{\partial \sigma}{\partial \dot{\varepsilon}}\right)_{\varepsilon,T} d\dot{\varepsilon} + \left(\frac{\partial \sigma}{\partial T}\right)_{\varepsilon,\dot{\varepsilon}} dT. \tag{6.18}$$

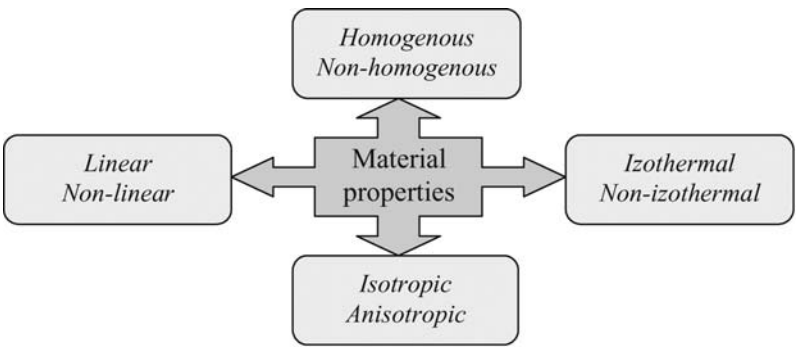


FIGURE 6.8. Material properties.

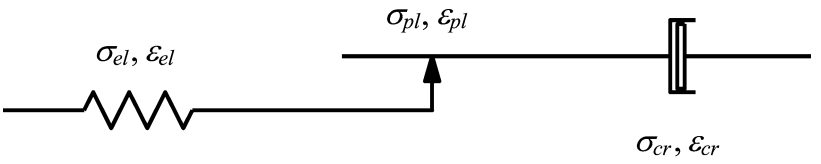


FIGURE 6.9. The combined material model.

- **Partitioned**; non-recoverable: response is modeled by separating time-dependent and time-independent behavior:

$$\varepsilon = \varepsilon_{el} + \varepsilon_{pl} + \varepsilon_{cr}. \tag{6.19}$$

- **Unified**; non-recoverable: response is modeled as a combined time-dependent manner:

$$d\sigma = E(\varepsilon, \dot{\varepsilon}, T), \tag{6.20}$$

where

$$\varepsilon = \varepsilon_{el} + \varepsilon_{inel}(\varepsilon_{pl}, \varepsilon_{cr}). \tag{6.21}$$

Actually partitioned case is applied in most of the applications and fatigue life prediction can be done by applying Miner's rule. In case of unified approach, fatigue life prediction can be achieved through the FEM analysis and is used for e.g., thermal shocks.

Any material exposed to thermo-mechanical loading conditions will deform. The total deformation can be evaluated by combining all three models: elastic, plastic and creep. In order to calculate the total strain under known loading, mathematical models of the corresponding strains should be assumed. According to the material engineering, elastic model is assumed to be linear while the plastic and creep models are nonlinear. Figure 6.9 shows the idea of this kind of partitioning:

$$\varepsilon(\sigma, t) = \varepsilon_{el}(\sigma_{el}) + \varepsilon_{pl}(\sigma_{pl}) + \varepsilon_{cr}(\sigma_{cr}, t). \tag{6.22}$$

In order to evaluate the cycle fatigue a most dominating behavior should be analyzed and the representing damage parameter taken into consideration. In fact, in case of microelectronic applications there are two dominating material behavior:

- Elasto-plasticity.
- Visco-elasticity.

6.2.2.1. Elastic Model The elastic model can be expressed by the following formula:

$$\varepsilon = \frac{1}{E}\sigma, \tag{6.23}$$

where E is a Young's modulus and ε is a strain. It is assumed that elastic model can be applied to strains $\varepsilon < 0.01$. The elastic properties are presented as an ideal string with a stiffness E. The typical characteristics of the elastic model are shown in Figure 6.10.

The main property of the elastic deformation is reversibility. This means that energy accumulated during the loading cycle can be totally recovered during the unloading. The

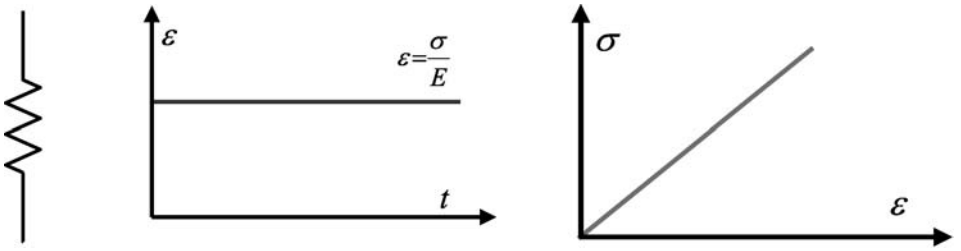


FIGURE 6.10. Characteristics of the elastic model.

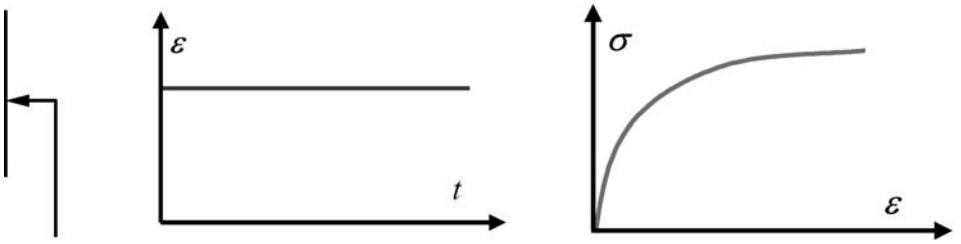


FIGURE 6.11. Characteristics of the plastic model.

same, there is no energy dissipation. The maximum stress $\sigma_{\max}$ can be applied as a representative damage parameter in case of the dominating elastic domain.

6.2.2.2. Plastic Model The plastic model can be represented by the following formula:

$$\varepsilon = \left(\frac{\sigma}{K}\right)^{\frac{1}{n}}, \tag{6.24}$$

where ε is a plastic strain, K is a strength coefficient and n is a strain hardening exponent. Plastic deformation occurs after the material is exposed to stress–strain defined as the yield point. Generally plastic domain appears for strains $\varepsilon = 0.01$–0.6. The typical characteristics of the plastic model are shown in Figure 6.11.

In comparison to elastic model, plastic deformation is not reversible. This means that during the loading cycle some of the accumulated energy is dissipated in the material in a form of a plastic deformation and it cannot be recovered during the unloading. The equivalent accumulated plastic strain $\Delta\varepsilon_{pl}$ can be applied as a representative damage parameter, in case of the dominating plastic domain.

6.2.2.3. Creep Model The creep model can be expressed by the following formula:

$$\sigma = \eta\dot{\varepsilon}, \tag{6.25}$$

where η is a viscosity and $\dot{\varepsilon} = d\varepsilon/dt$ is a rate of a strain change. From a kinematics view, creep is similar to plasticity, except that creep is a function of time. The creep rate depends strongly on temperature and stress:

$$\dot{\varepsilon} = \frac{d\varepsilon}{dt} = f(T, \sigma). \tag{6.26}$$

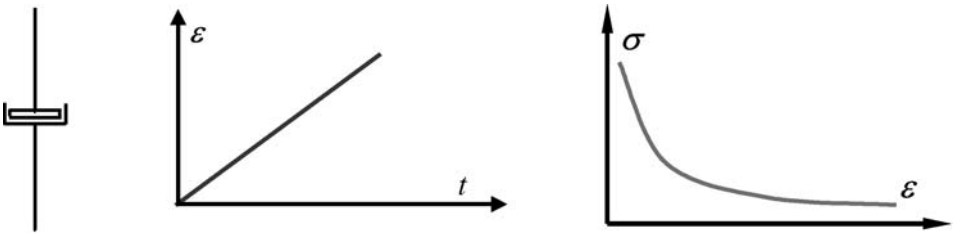


FIGURE 6.12. Characteristics of the creep model.

One of the most popular creep laws is based on time dependent inelastic behavior, described by the steady state power law:

$$\dot{\varepsilon}_{cr} = A\sigma^{\frac{1}{n}} \exp\left(\frac{-Q}{kT}\right), \tag{6.27}$$

where σ corresponds to von Mises' stress, A is a creep constant, k is a universal gas constant, Q is a creep activation energy and n is a creep strain hardening exponent. The typical characteristics of the creep model are shown in Figure 6.12.

In test at high strain, two different phenomena may be observed: cyclic creep and stress relaxation. Cyclic creep is a progressive straining under constant stress, while stress relaxation may occur under constant strain. Similarly to plastic deformation, creep deformation is not reversible as well. This means that during the loading cycle some of the accumulated energy is dissipated in the material and it cannot be recovered during the unloading. The equivalent accumulated inelastic strain energy ΔW_{in} can be applied as a representative damage parameter, in case of the dominating creep domain.

6.2.3. Thermo-Mechanical Related Failures

Typical microelectronic package consists of a number of different materials and interconnections between them. In fact, the list of possible failure modes is very long and precise analysis requires interdisciplinary knowledge including mechanics, physics, chemistry, mathematics, etc. Figure 6.13 shows the outline of a typical microelectronic package, and Figures 6.14 and 6.15 typical thermo-mechanically related failures [54].

During manufacturing and operation of microelectronic packages various thermal and mechanical loading conditions are inherited. This is the major cause of the induced failures represented in Figures 6.14 and 6.15. According to the extensive root cause analysis of failures it was concluded that:

- Thermal, mechanical and thermo-mechanical related failures may account for more than 65% of the total failure rates.
- Thermal, mechanical and thermo-mechanical related failures often originate from the product and process design phase.

Though, the list of possible failures is very long, the basic reason for thermo-mechanical failure is due to the thermal mismatch of CTE (coefficient of thermal expansion), which can be classified as, see Figure 6.16:

- Thermo-elastic.
- Plastic.

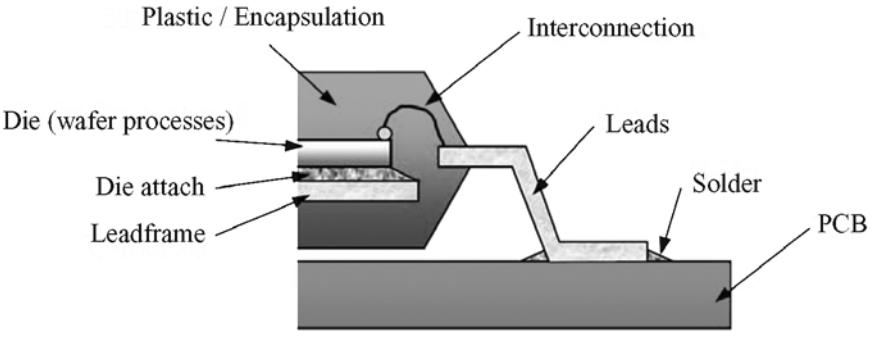


FIGURE 6.13. An example of typical microelectronic package.

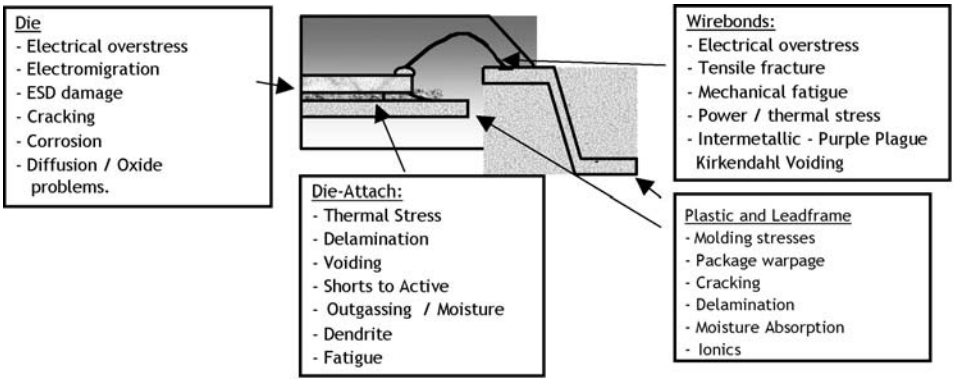


FIGURE 6.14. Failure modes for packaging level I [54].

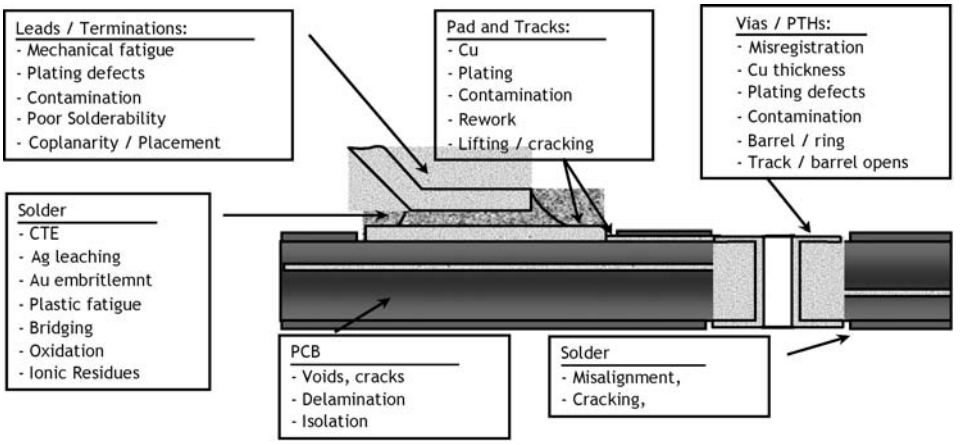


FIGURE 6.15. Failure modes for packaging level II [54].

- Creep.
- Fracture.
- Fatigue.

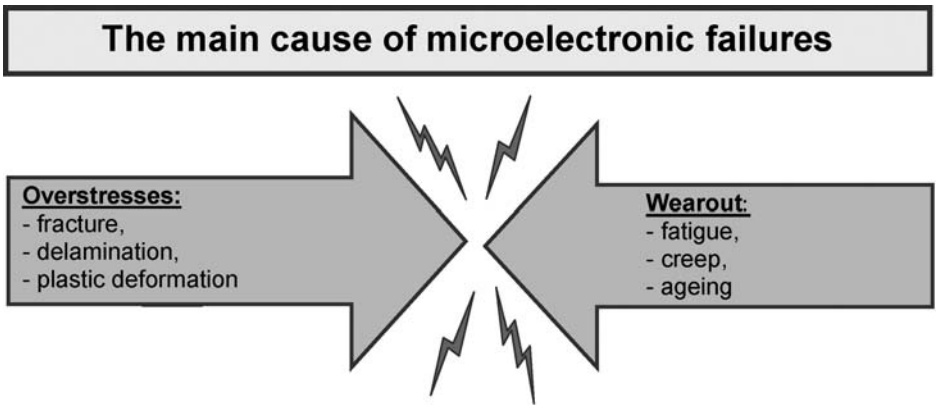


FIGURE 6.16. Thermo-mechanical related failures in microelectronics.

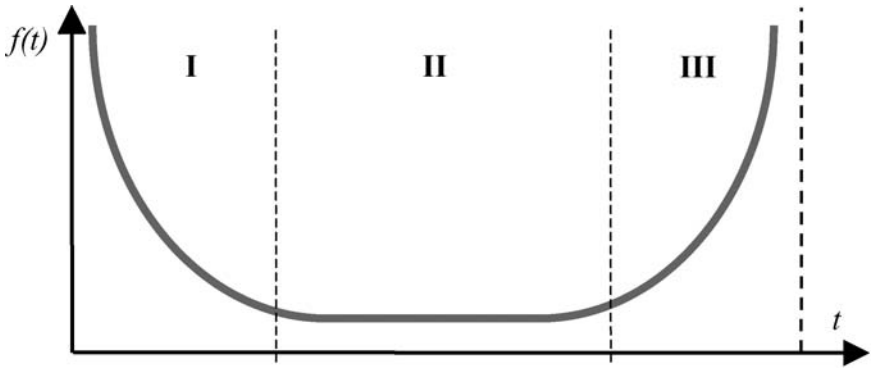


FIGURE 6.17. Failure density distribution curve.

Reliability is major concern in the microelectronic industry. Reliability is defined as *the probability that a product will survive under certain conditions during a certain period of time*. As such, failure is defined as the probability that a product is not functioning as designed. Typical measure to quantify reliability is the failure density function. The failure density function $f(t)$, or "bathtube" curve, can be divided into three parts, see Figure 6.17:

- infant mortality; this due to manufacturing defects, e.g., referred to as quality (I),
- intrinsic failure; this is due to typical user functional usage, most often caused by high-level stresses outside the design specifications, e.g., drop (II),
- wearout; gradual "wearing and tearing" in normal usage over the product life (III).

Data from tests concerning failure analysis are collected in a tabular form and most often are analyzed according to the Weibull distribution function:

$$f(t) = \frac{\beta}{\lambda} \cdot \left(\frac{t - t_0}{\lambda} \right)^{\beta - 1} \cdot e^{-(\frac{t-t_0}{\lambda})^{\beta}}, \tag{6.28}$$

where β is shape, λ is lifetime and t_0 is free lifetime. The above approach allows for reduction of the amount of data that are to be collected through experiment due to extrapolation of the approximated analytical formula:

$$F(t) = 1 - \exp\left[-\left(\frac{t}{\lambda}\right)^{\beta}\right]. \tag{6.29}$$

The product lifetime N_f is defined as:

$$f(63.2\%) \approx \lambda. \tag{6.30}$$

Figure 6.18 shows an example of the Weibull distribution.

6.2.3.1. Example Case Let's assume that we have a simple 2D package model consisting of substrate and silicon chip connected by solder bumps. The package is exposed to temperature change ΔT and the goal is to evaluate the shear stress γ. Figure 6.19 shows a schematic of the example case.

Under the assumptions on elastic material models, the strain due to temperature change is given as:

$$\varepsilon_{Si} = \alpha_{Si}\Delta T, \quad \varepsilon_{FR4} = \alpha_{FR4}\Delta T, \tag{6.31}$$

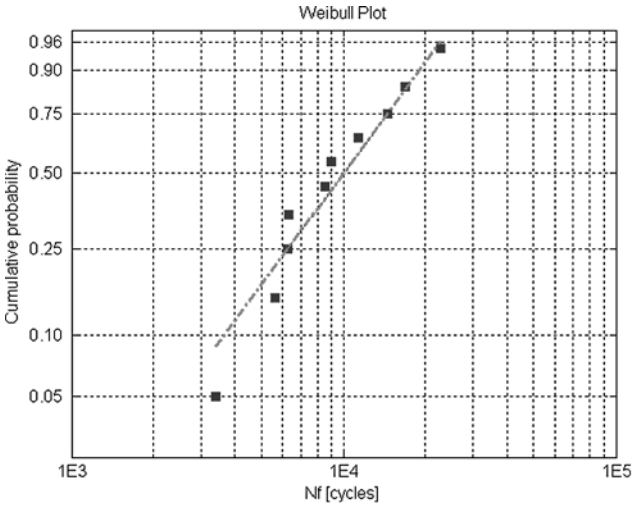


FIGURE 6.18. Weibull density distribution curve.

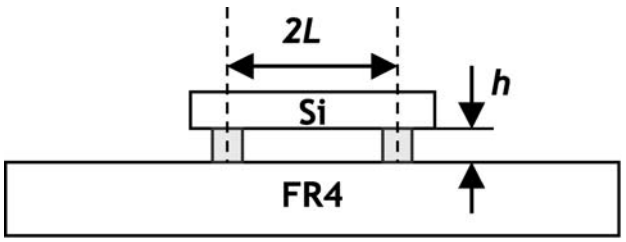


FIGURE 6.19. Thermo-mechanical related failures in microelectronics.

where the elongation at the end of the die:

$$d_{Si} = \varepsilon_{Si} L = \alpha_{Si} \Delta T L, \quad d_{FR4} = \varepsilon_{FR4} L = \alpha_{FR4} \Delta T L. \tag{6.32}$$

The shear strain in the solder can be finally evaluated as follows:

$$\gamma = \frac{d_{FR4} - d_{Si}}{h} = (\alpha_{FR4} - \alpha_{Si}) \left(\frac{L}{h} \right) \Delta T. \tag{6.33}$$

Basing on the formula it is possible to draw the final conclusions concerning the shear strain problem:

- It is proportional to:
  - CTE mismatch $(\alpha_{FR4} - \alpha_{Si})$.
  - Distance to the neutral point (DNP $- L$).
  - Temperature change ΔT.
- It is inversely proportional to height of the solder joint h.

In order to minimize the shear strain it would desirable to diminish the CTE mismatch between the interconnected materials. Unfortunately, in most of the cases it is not feasible and therefore some designing approach is required, which could improve the product reliability.

6.2.4. Designing for Reliability

The ability to assess the reliability of the product is one of the most important tasks of the design engineer. Unfortunately, reliability of the electronic packages is a complex task, which requires interdisciplinary knowledge including e.g., managing. Anyway, most often reliability problem can be presented in a form of the Ishikawa diagram by listing the causes for low reliability, see Figure 6.20.

Despite of the undisputed achievements of the reliability engineering, it is commonly known that electronic components fail. Currently there are two main approaches to reliability assessment

- Statistical: reliability is a function of time.
- Physical: reliability is a function of a component physical state.

Additionally reliability assessment can be considered as, see Figure 6.21:

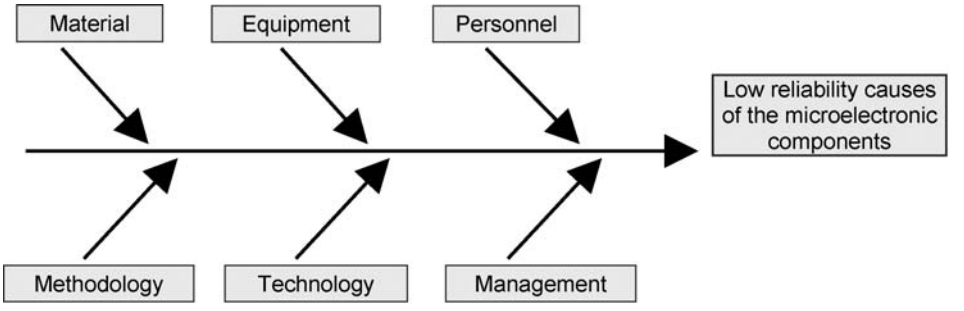


FIGURE 6.20. Ishikawa diagram for the electronic packages.

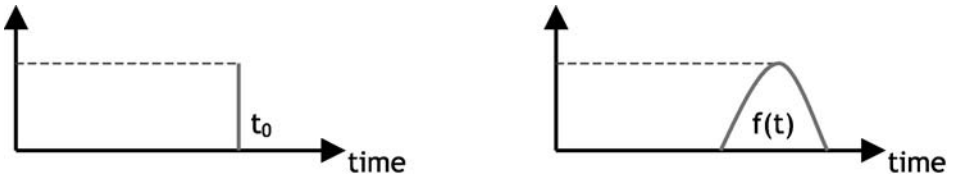


FIGURE 6.21. The difference between deterministic and probabilistic approach.

- Deterministic: input variables are precisely defined and therefore output variable would be precisely defined as well.
- Probabilistic: input variables are treated as random variables and therefore output variable has a random distribution.

Understanding the mechanisms that cause components failure is the key factor to make reliable products. The most common thermo-mechanical failure modes, which are addressed in microelectronic are due to:

- Overstress: fracture.
- Wearout: fatigue.

6.2.4.1. Fracture Fracture is probably the most dominant thermo-mechanical failure mode in microelectronic packages [47]. Fracture can cause:

- Die cracking.
- Underfill cracking.
- Solder joint failure.
- Body cracking.
- And many more.

The basic approach to fracture mechanics is based on linear elastic fracture mechanics (LEFM). The LEFM is based on the observation that near the crack tip in a brittle material, the magnitude of the singular stress field can be controlled by a single parameter. According to LEFM it is possible to evaluate the singular stress near crack tip as:

$$\sigma = \frac{K}{\sqrt{2\pi x}}, \tag{6.34}$$

where x is the distance to the crack tip, K is the stress intensity factor, which is an indication of the stress magnitude near the crack tip. Fracture will occur when:

$$K \geq K_c, \tag{6.35}$$

where K_c is called fracture toughness and is an intrinsic material property that can be measured using a number of techniques. Two distinct fracture mechanisms can be discriminated:

- Brittle fracture: separation of materials along crystallographic planes due to atomic bonds breaking.
- Ductile fracture: initiation, growth and coalescence of micro-voids.

In fact, ductile fracture is considered to be dominant mechanism in microelectronics. Figure 6.22 shows the stress–strain relation for both mechanisms.

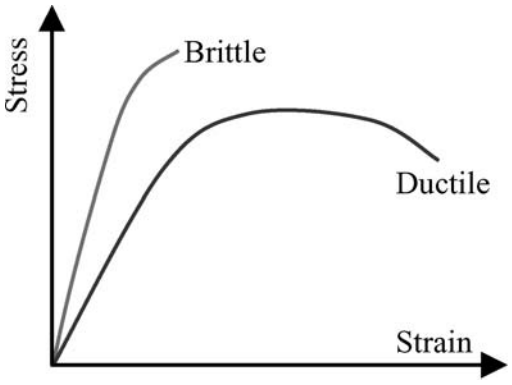


FIGURE 6.22. The difference between deterministic and probabilistic approach.

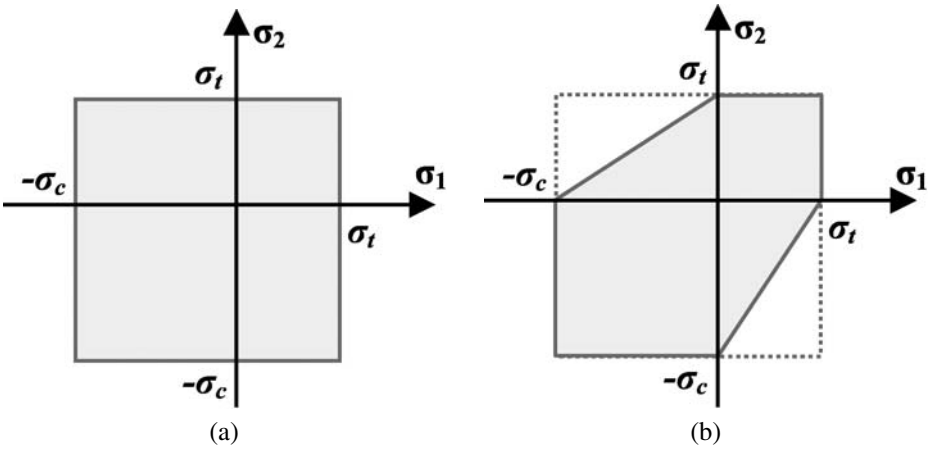


FIGURE 6.23. The fracture criteria for brittle materials: Coulomb (a) and Mohr (b).

One of the basic problems during fracture analysis is due to failure criteria. In fact, it is meant that fracture occurs when the stress level exhibits the material strength. Due to specific behavior of the both materials, failure criterion should be evaluated differently.

Brittle materials (Figure 6.23):

- The maximum stress criterion, also known as the normal stress, Coulomb, or Rankine criterion:

$$\sigma_c < \{\sigma_1, \sigma_2\} < \sigma_t, \tag{6.36}$$

  where σ_t is uniaxial tension strength and σ_c is uniaxial compression strength.
- The Mohr Theory of Failure, also known as the Coulomb-Mohr criterion or internal-friction theory, is based on the Mohr's Circle.

Ductile materials (Figure 6.24):

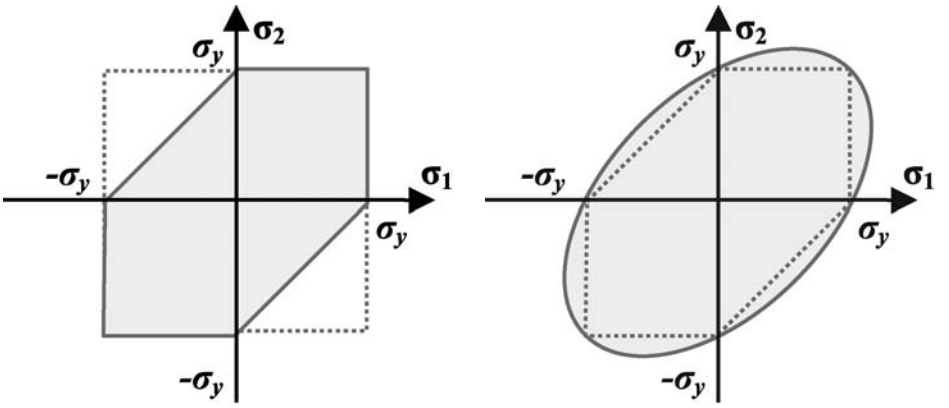


FIGURE 6.24. The fracture criteria for ductile materials: Tresca (a), von Mises (b).

- The maximum shear stress criterion, also known as Tresca's or Guest's criterion:

$$|\sigma_1| \leq \sigma_y,\ |\sigma_2| \leq \sigma_y,\ \text{and}\ |\sigma_1 - \sigma_2| \leq \sigma_y, \tag{6.37}$$

  where σ_y stands for the yield point.
- von Mises criterion, also known as the maximum distortion energy criterion:

$$\frac{1}{2}[(\sigma_1 - \sigma_2)^2 + (\sigma_2 - \sigma_3)^2 + (\sigma_3 - \sigma_1)^2] \leq \sigma_y. \tag{6.38}$$

6.2.4.2. Fatigue In practical applications, fracture can occur even when the load amplitude is below the static strength. This is due to fatigue. Fatiguestress is a failure mode caused by the cyclic load, e.g., mechanical, thermal, etc. A commonly method for characterizing the fatigue behavior is the S/N curve, where S stands for load amplitude and N is the number of cycles to failure for the given load amplitude (Figure 6.25).

Thermal cyclic fatigue occurs under repeated applications of thermal stress, resulting from the temperature changes as well as from component materials with different thermal expansion coefficients. The primary reason of the thermal cyclic fatigue failure is fracture. Therefore studying the physics of cracks can bring some light on predicting the fatigue failure of electronic packages (Figure 6.26). Fatigue failure is mainly due to energy (damage) accumulation over cycles, which can be assessed according to the response hysteresis. There are different approaches to fatigue analysis:

- Steady-state crack growth da/dN_f: based on LEFM.
- Stress based σ: based on elastic deformation.
- Strain based ε: based on elastic and plastic deformation.
- Inelastic energy W: based on inelastic energy dissipation.

According to the material model the cyclic fatigue depends on damage parameter and can be considered as LCF (low cycle fatigue) and HCF (high cycle fatigue) domain. The LCF ($N_f < 10^5$) failure is assumed to be the inelastic strain method while the HCF ($N_f > 10^6$) is assumed to be the elastic strain method, where:

- high cycle fatigue for stress $S \ll \sigma_y$ (yield point)
- low-cycle fatigue for stress $S \gg \sigma_y$ (yield point)

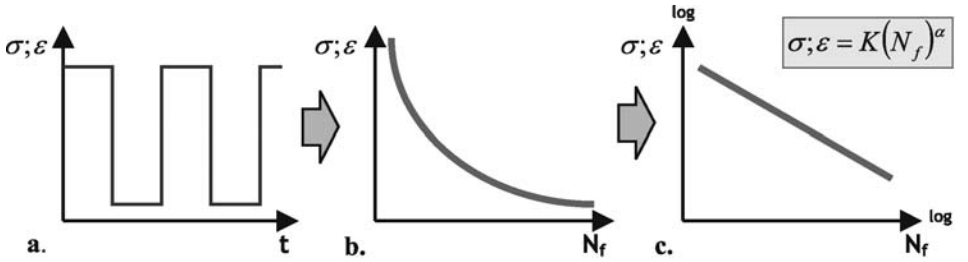


FIGURE 6.25. Failure analysis due to fatigue; cyclic load (a), S/N curve (b), S/N curve in log–log scale (c).

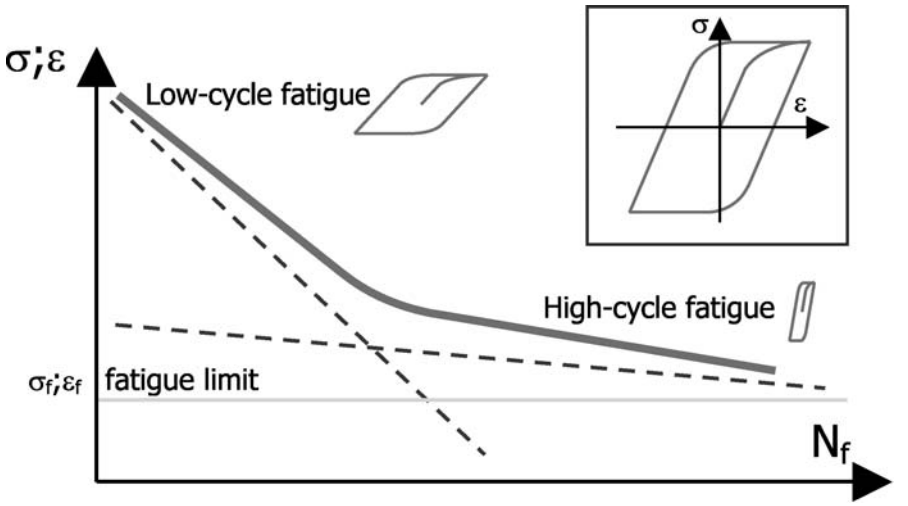


FIGURE 6.26. Cycle fatigue.

6.2.4.2.1. Steady-State Crack Growth. The crack propagation law was introduced basing on the experimental results. According to that, a crack growth rate da/dN shows a linear correlation in the region II, which could be expressed by the following formula, see Figure 6.27:

$$\frac{da}{dN} = C(\Delta K)^n, \tag{6.39}$$

where C is a crack growth rate factor, ΔK is a stress intensity factor, n is an exponent taken from experiments. The stress intensity factor ΔK can be expressed by the formula:

$$\Delta K = \alpha(\sigma_{\max} - \sigma_{\min})\sqrt{\pi a}, \tag{6.40}$$

where α is a proportional coefficient (e.g., $2/\pi$), $\sigma_{\max}$ and $\sigma_{\min}$ are the maximum and minimum stress and a is the crack length.

6.2.4.2.2. Strain Based Approach. In order to determine the thermal fatigue life in case of temperature-dependent elasto-plastic analysis an accumulated equivalent plastic strain

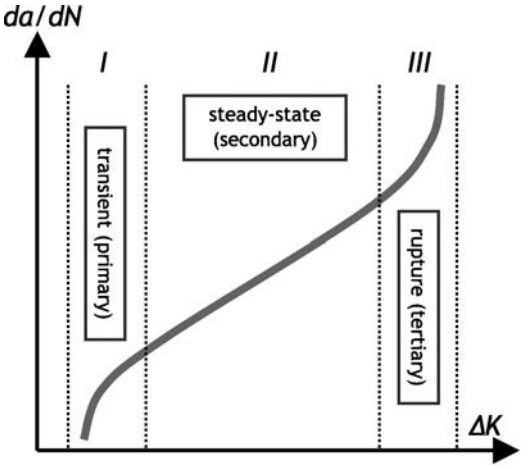


FIGURE 6.27. Dependence of crack growth rate vs. stress intensity.

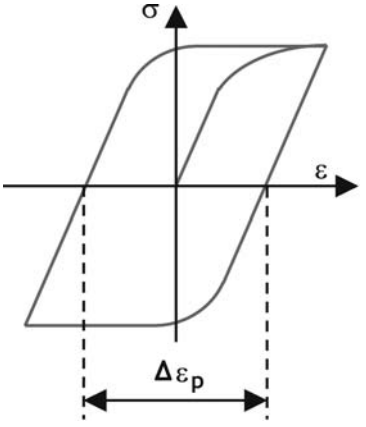


FIGURE 6.28. Equivalent plastic deformation.

range $\Delta\varepsilon_p$ is a good solution (Figure 6.28). The thermal fatigue life can be then estimated using the Coffin–Manson relation:

$$\Delta\varepsilon = \Delta\varepsilon_e + \Delta\varepsilon_p = \frac{\sigma_f}{E}(2N_f)^b + \varepsilon_f(2N_f)^c, \tag{6.41}$$

where σ_f is a fatigue limit and material b is a material constant while c and ε_f are ductility coefficient and exponent. The modified Coffin-Manson model, excluding the elastic component can be rewritten as follows:

$$N_f = \frac{1}{2}\left(\frac{\Delta\varepsilon_p}{2\varepsilon_f}\right)^{\frac{1}{c}}. \tag{6.42}$$

6.2.4.2.3. Inelastic Energy Approach. (Figure 6.29). The fatigue damage begins with the accumulation of damage at a localized region due to cycling loads, which usually leads to

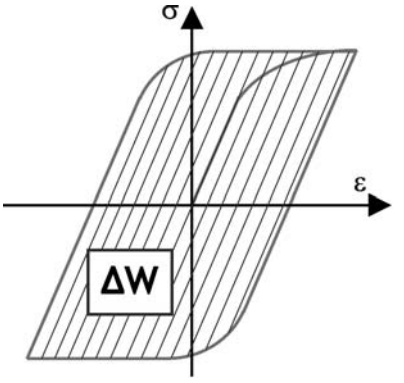


FIGURE 6.29. Inelastic energy dissipation.

cracking. The cracking process is governed by two key phenomena: crack initiation and crack propagation. The cracks usually initiate at the surface where the highly stress regions exist, e.g., notches. Once the crack is initiated it may propagate as a result of further cyclic deformation. So, the expected number of cycles to fatigue failure can be a sum of: number of cycles to crack initiation N_{fi} and number of cycles to complete fracture N_{fp}:

$$N_f = N_{fi} + N_{fp}. \tag{6.43}$$

In time-dependent material models (e.g., visco-elastic) both: the number of cycles to crack initiation N_{fi} and a number of cycles to complete fracture N_{fp} could be estimated from the following equations:

$$N_{fi} = \frac{B}{\Delta W},$$
$$N_{fp} = \frac{1}{C}\int_{a_i}^{a_f} \frac{1}{(\Delta W)^n} da, \tag{6.44}$$

where ΔW is the strain energy per cycle, a is the crack length of the solder joint, a_i is initial crack size and a_f is the crack length at failure, B, C and n are material constants. The inelastic strain energy per cycle ΔW, seems to be an appropriate damage parameter, as it correlates well with the experimental data.

6.3. MATHEMATICAL ASPECTS OF OPTIMIZATION

The first step of virtual prototyping is to create model of the process or product. Most often the procedure for model recognition would be based on controlling the input variable values and observing the output. Unfortunately, for most of the projects the above procedure seems to be very expensive and long lasting. But thanks to mathematical tools the above procedure may become efficient by reducing the required number of experiments. The most popular mathematical tools are:

- DOE (design of experiments): DOE method would be responsible for selecting an optimal set of experiments for recognizing the model of a product or process, which is selecting the most significant factors.

- RSM (response surface modeling): RSM method would be responsible for fitting surface model of the response.

The experiment is understood as planned series of tests. There are a number of problems connected with this task:

- Number of experiments, which are required for model recognition—the goal is to run as less as possible.
- Proper selection of input variables—the goal is to assess the potential power of each of them.
- Induced random distribution of input and noise variables—the goal is to assume their basic random parameters: density distribution, average, variance, etc.
- If the model is linear or non-linear—most often for simplification reasons the model is assumed to be linear, while it is highly non-linear.
- If the model requires multi-response analysis range.
- If variables are continues or step-wise.
- Influence of the noise factors—the goal is to make the final design as less sensitive as possible to noise variables, e.g., temperature.
- Multi-criteria design—the goal is to select the compromised solution in reference to a number of output variables, e.g., quality and price.
- Optimization—the goal is to find out the optimal solution in reference to required level of the response: minimal, maximal or nominal.
- Parameter and tolerance design—the goal is to select the most appropriate values of input variables to achieve optimal and at the same time the most robust design.

6.3.1. Design of Experiments

The DOE method is to answer the question how to design the experiment so it can be done in a minimum number of tests and at the same time it can deliver enough information (Figure 6.30). Unfortunately, the number of tests that are to be done in the experiment grow exponentially with a number of input variables. Therefore, the basic challenge is to design the so-called proper experiment, which could save a lot of time and money. The experiment should include only the "good" tests, which provide appropriate information on the model and to skip the tests that are "overlapped" or not required for model definition [21,32,43].

The main idea behind the design of experiments is controlling the input variables and recording the output signal of a product or process. Most often the input variables are referred to as factors while the output variable as a response. The factors could be continuous or step variables having at least two values. The factor values are called levels and can be divided into:

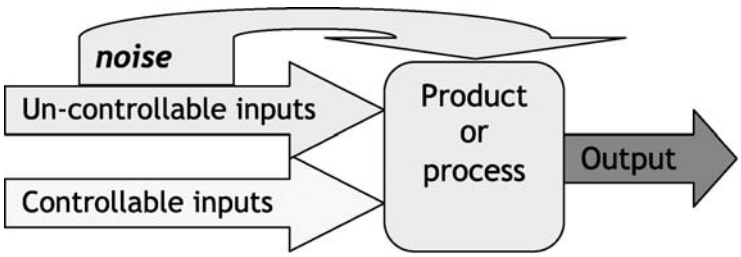


FIGURE 6.30. Selection and description of input and output factors in DOE method.

- Controllable factors, which are easy to identify and control.
- Uncontrollable factors, which are difficult to identify and control, e.g., temperature; the uncontrollable factors can be additionally divided into: internal and external; both of them are referenced as noise factors.

When searching for the adequate experiment some tests are run and the response observed. The most often approach and straightforward would be to evaluate the effect of one input parameter on the response while keeping the values of other factors at constant level. The worst case would be to evaluate the effect of a number of input parameters all at the same time, which is sometimes a common approach. It is possible to change one or more process variables (or factors) in order to observe the effect of changes on one or more response variables. The statistical design of experiments is an efficient procedure for planning experiments so that the data obtained can be analyzed to yield valid and objective conclusions [28,49] (Figure 6.31).

Well-chosen experimental designs maximize the amount of "information" that can be obtained for a given amount of experimental tests. The possible outputs of the experiment can answer the question whether selected factor influences the response mean and/or variance, see Figure 6.32.

A problem arises when response has some statistical distribution due to e.g., noise. Then the basic question refers to the problem of selecting a proper number of tests in order to get the statistical significance of the experiment and appropriate comparison of sample averages. From statistical point of view in such a case in order to have more confidence it is required to increase degrees of freedom for that factor. This can be achieved by increasing the number of tests but possibly getting more degrees of freedom for an error rather than for an analyzed factor. Additionally in most practical cases there would be always some interactions between selected input factors, which should be evaluated as well. Therefore, it seems, that the systematic approach to design of an experiment would be profitable as in cost as in time. There is a number of applications with different objectives where the experimental design can be used:

- *Screening objective*; the primary purpose of the experiment is to select or screen out the few important main effects from the many less important ones; these screening designs are also termed main effects designs.
- *Optimal fitting of a regression model objective*; if the goal is to model a response as a mathematical function of a few continuous factors and a "good" model parameter estimates are required.

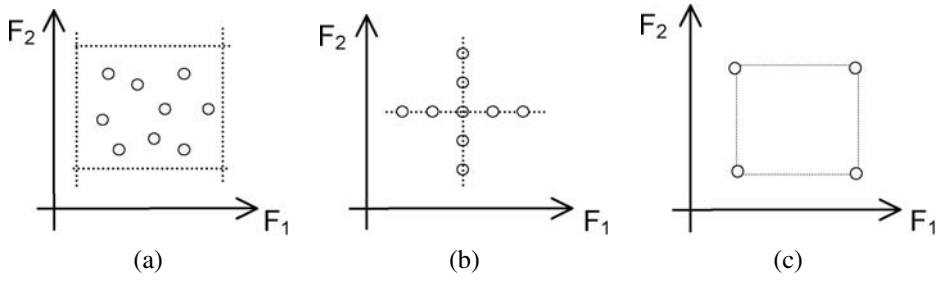


FIGURE 6.31. Different types of experiment designs: best guess (a), one-factor at time (b), statistically designed (c).

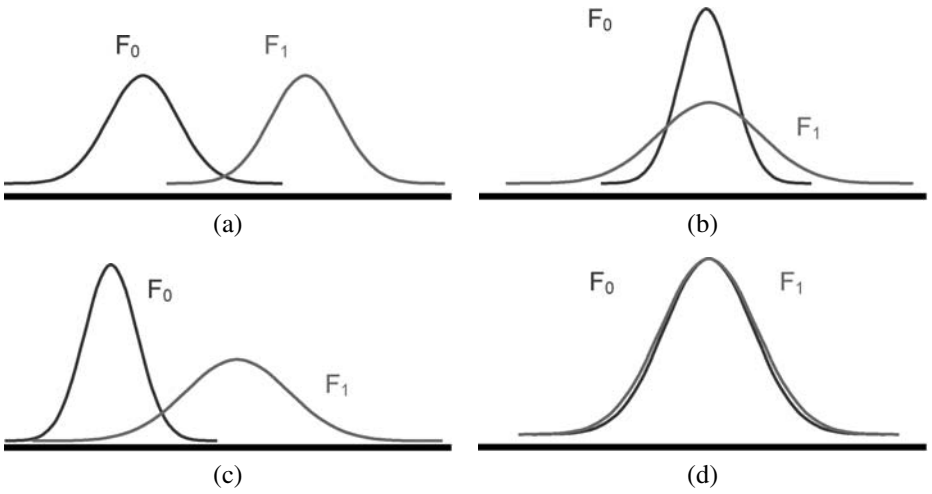


FIGURE 6.32. Possible experiment outputs: factors that shift the mean (a), factors that shift the variability (b), factors that shift both the mean and variability (c), factors that have no affect (d).

- *Robust design objective*; the goal is to perform an experiment, which would allow adjusting the essential factors such that the response would be insensitive to noise and induced variability of the factors.

The choice of an experiment design depends on: number of factors to be investigated and whether an experiment is physical or numerical. In reference to the last condition, the choice of an experiment can be made according to:

- "Corner DOE" experiments: Factorial, CCD.
- Space filling DOE: Latin hypercube, Monte Carlo.

The full-factorial design produces a uniform grid with user-specified density covering the input parameter space. There are other conventional designs, including fractional-factorial. Finally, Latin hypercube sampling (LHS) provides an orthogonal array that randomly samples the entire design space. LHS can be looked upon as a stratified Monte Carlo sampling where the pair-wise correlations can be minimized to a small value (which is essential for uncorrelated parameter estimates) or else set to a desired value. LHS is especially useful in exploring the interior of the parameter space, and for limiting the experiment to a fixed (user specified) number of runs [16,18].

There is a difference between physical and numerical experiment and it seems to be quite important. The basic characteristic of the both can be summarized as follows:

- in computer experimentation noise does not play a role, since running a computer simulation twice yields exactly the same results. Therefore, no information is gained from the repeated simulation of the same design, which is often a rule in classical DOE,
- in physical experiments it is often optimal to have design points lie on the borders of the design region. In computer experiments other parts of the design region are often equally interesting.

A basic drawback of most classical experiments is that they are only applicable for rectangular design regions. Therefore, in case of computer experiments it is almost a rule of thumb to use experiments that are:

- Space filling (test points are evenly spread out throughout the design region).
- Non-collapsing (every test gives bears information on about the influence of the other design parameters on the response).
- Sequential (minimizing the required number of experiments by selecting the minimal initial scheme and then caring out an additional experiments in order to improve the specified criterion on RSM model accuracy).
- Able to handle non-box design regions (in most cases the feasible design region is non-box as points outside this region may have no physical interpretation).

6.3.1.1. Full Factorial Design The most popular scheme of experimental design would be based on orthogonal arrays (OA). Orthogonality means that:

- Factors can be evaluated independently of one another; the effect of one factor does not bother the estimation of the effect of another factor,
- Orthogonal experiments are balanced, which means equal number of samples under various levels of analyzed factors in the selected domain (Figure 6.33).

Using the above experiment, both factor and interaction effect can be estimated. The experiment based on OA can avoid the unnecessary experiments and to keep only the necessary ones. In case of two-level full factorial experiment, the number of required experiment can be evaluated by the following formula:

$$N = 2^f, \tag{6.45}$$

where f is the number of factors (each at two levels). Unfortunately, the number of experiment grows exponentially with the number of factors. The worst situation appears in case of evaluating the factor effects at three levels of each factor. In this case, the number of required experiments can be evaluated as follows:

$$N = 3^f. \tag{6.46}$$

In fact, the full factorial experiment based on OA can be accepted only under certain conditions. The two-level experiment is performed with a few factors, e.g., 4. Otherwise the

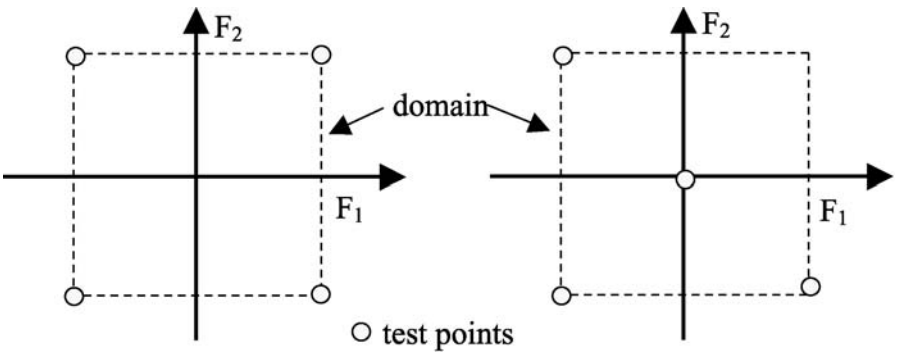


FIGURE 6.33. The orthogonal and non-orthogonal experiment.

TABLE 6.1.
A comparison of a required number of experiments.

Number of factors	Two level experiment	Three level experiment
1	2	3
2	4	9
3	8	27
4	16	81
5	32	243
6	64	729

number of test, included in the experiment may be too large, see Table 6.1. Unfortunately, in most of the engineer tasks the number of factors involved would be greater than 5 and additionally because of the high non-linear effects they should be selected at three levels for each factor.

Nevertheless, the full factorial experiment includes all the main factor effects as well as all possible interactions that are orthogonal to one another. Results of the experiment are most often interpreted according to the analysis of variance (ANOVA).

6.3.1.1.1. ANOVA Analysis. The ANOVA analysis comes from one of the fundamental theorems of statistics, the Central Limit Theorem (CLT). The CLT refers to the population despite of its distribution. The population average and its variance is the only thing that is taken into account. According to the CLT theorem if we select samples from the population than the following tenets are true [30,31]:

- Sample averages tend to be normally distributed regardless of the distribution of the individuals,
- The average of the distribution of sample averages will approach the average of the distribution of the individuals,
- The variance of the sample averages is less than the variance of the distribution of the individuals.

The CLT can be summarized in the sentence (Figure 6.34):
"if the population is sampled than the sample averages have the normal distribution:

$$\overline{E}_{yi} = N(E, \sigma) \tag{6.47}$$

with the following parameters:

$$E = E_y, \quad \sigma = \frac{\sigma_y}{\sqrt{n}}, \tag{6.48}$$

where n is a sample size."

In ANOVA analysis, the most important tenet is that the variance of sample averages σ_y^2 will be equal to the variance of the individuals σ_y^2 divided by the sample size n used to obtain the sample averages. The same is true for the variance estimates:

$$\sigma_{\frac{2}{y}} = \frac{\sigma_y^2}{n} \quad \Rightarrow \quad S_{\frac{2}{y}} = \frac{S_y^2}{n}. \tag{6.49}$$

The above formula can be used to make an estimate of the variance of sample averages by taking individuals variance and dividing it by the sample size. Therefore by comparing the

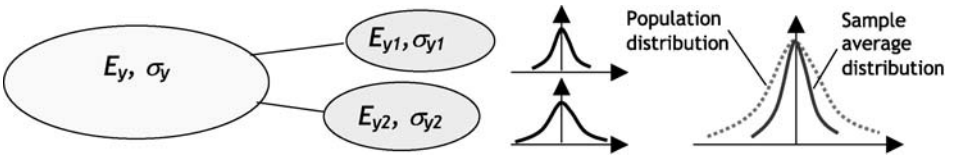


FIGURE 6.34. The concept of CLT.

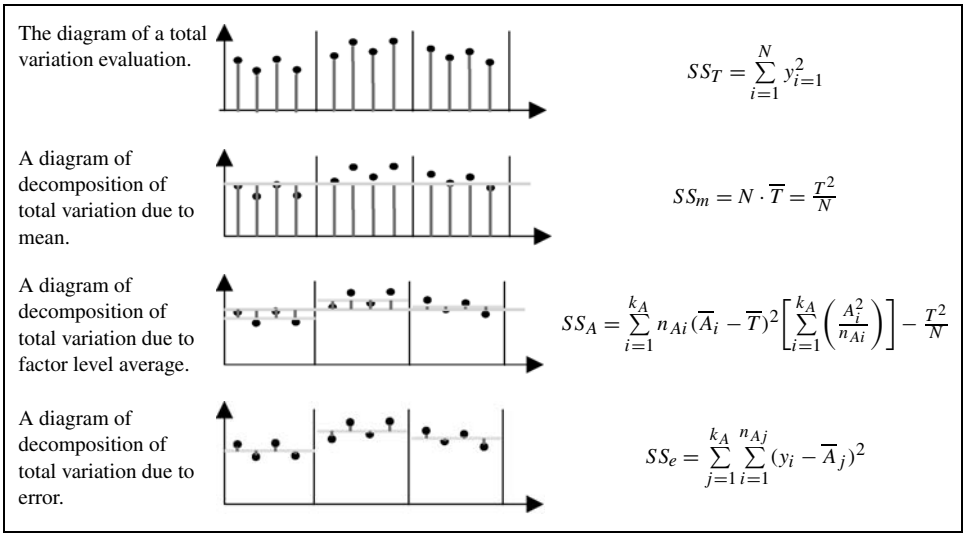


FIGURE 6.35. The concept of decomposition procedure.

variance determined from variation of sample averages and variance of an error determined from the variation of individuals can give a clue whether samples come from the same population. In practice it means that during the experiment we have to change the value of a chosen factor and then read the response and repeat the procedure for every factor depending on the selected design scheme. If the factor affects the response then through the ANOVA analysis it will be clear whether samples come from the same or different populations. The first step in the ANOVA analysis is a separation of different sources of variation in reference to individual tests through the decomposition procedure.

Generally, in case of one factor the decomposition can be presented as in Figure 6.35 and the total variation will sum up to:

$$SS_T = SS_m + SS_A + SS_e. \tag{6.50}$$

The variation due to the mean most often does not affect the overall calculations; therefore it can be excluded from ANOVA analysis:

$$SS_T = \sum_{i=1}^{N} y_{i=1}^2 - SS_m = \sum_{i=1}^{N} y_{i=1}^2 - \frac{T^2}{N}. \tag{6.51}$$

In case of more factors, additionally the variation due to factors interactions should be included, e.g., SS_{A*B}:

$$SS_T = SS_m + SS_A + SS_B + SS_{A*B} + SS_e. \tag{6.52}$$

It is easy to calculate the variance S due to the known variation SS; it is to be divided by degrees of freedom v corresponding to the chosen variation:

$$S = \frac{SS}{V}. \tag{6.53}$$

The main goal is to answer the question whether the change of a selected factor affects the response or not. That can be answered by comparing the factor variance S_A with the error variance S_e:

$$SS_A \leftrightarrow SS_e. \tag{6.54}$$

The comparison is done by analysis of variance (ANOVA). ANOVA analysis is based on F test (ratio of sample variances). This provides a decision at some confidence level whether two samples come from the same population:

$$F = \frac{S_1^2}{S_2^2}. \tag{6.55}$$

6.3.1.2. Fractional Factorial Experiment Design There is a possibility of reducing the required number of experiments but with some cost to pay, which is the lost concerning some interaction effects. Statisticians have developed more effective plans in comparison to the full factorial based on orthogonal array, which are referred to as fractional factorial experiments (FFEs). FFEs use only a portion of the total possible combinations to estimate mainly the main factor effects and some of the interactions, which seem to play a vital role (Figure 6.36). The FFEs are divided according to the part of the full factorial experiment: one-second: ½ FFE, one-fourth: ¼ FFE, etc.

Reduction of the full factorial experiment, e.g., by eighth, it is very tempting, especially from the cost and time perspective. There is a family of FFE matrices, called orthogonal arrays (OA), which can be utilized in various situations. They are referred to as: L8, L9, L16, etc. depending on number of test and number of levels for each factor (Table 6.2).

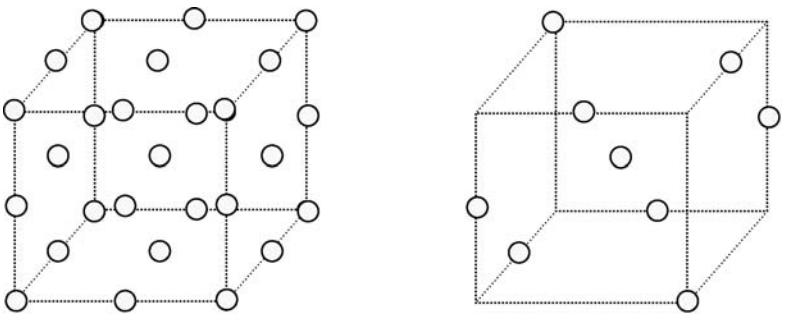


FIGURE 6.36. Example of orthogonal array for full and fractional factorial experiment.

TABLE 6.2.
The most often used orthogonal arrays.

Two-level experiment	Three-level experiment
L8	L9
L12	L18
L16	L27

Due to the fact, that orthogonal arrays based on FFEs have reduced number of tests in comparison to the full factorial experiment, there is a basic question about the resolution of the experiment and how it depends on the number of factors, levels and tests associated to the array. It is up to the expert to decide whether to perform high resolution experiment or low resolution experiment, which is mainly directed towards main factor effects, excluding the interactions. Nevertheless, the real power in using an OA is the ability to evaluate several factors in a minimum number of tests. This is considered an efficient experiment since much can be obtained from a few trials.

In fact the whole OA experiment can be divided into two steps, while the first would be focused on finding out the main factor effects, the second one to recognizing the interaction effects. The first experiment is sometimes refereed to as *screening experiment*. It should include many factors but at few levels of which two are recommended in order to minimize the size of the beginning experiment. In result of the initial experiment many factors will be eliminated from father evaluation and the remaining ones can be investigated with multiple levels without increasing the size of an experiment. The same, both interactions and non-linear effects can be studied more deeply. The above procedure of running two experiments seems very efficient and allows for time and cost reduction.

6.3.1.3. Central Composite Design (CCD) Sometimes, the preliminary experiment would be done with the assumption on linearity of the response. On the other hand, it is almost the rule of thumb that response is non-linear. In order to determine whether response in a promising region is non-linear, e.g., higher order polynomial, the Central Composite Design (CCD) scheme may be planned.

The CCD design contains an embedded factorial or fractional factorial design with center points that are augmented with a group of "star points" that allow estimation of curvature. If the distance from the center of the design space to a factorial point is ± 1 unit for each factor, the distance from the center of the design space to a star point is $\pm\alpha$ with $|\alpha| > 1$. The precise value of α depends on certain properties desired for the design and on the number of factors involved. In case of the CCD, an experiment design would include additionally test points that are within a selected domain region in spite of the orthogonal array experiment scheme (Figure 6.37).

CCD design always contains twice as many star points as factors. The star points represent new extreme values (low and high) for each factor in the design. There is variety of the central composite designs:

- Circumscribed (CCC).
- Inscribed (CCI).
- Face Centered (CCF).

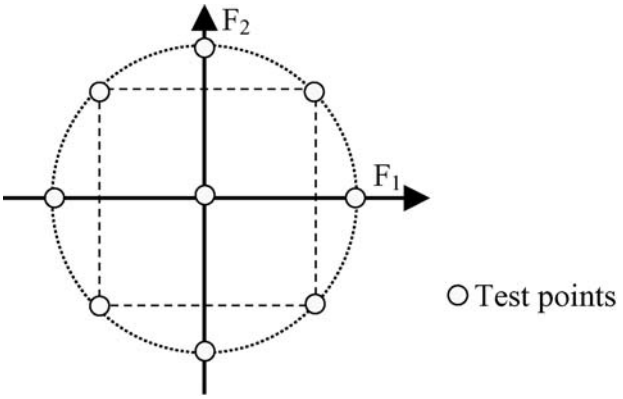


FIGURE 6.37. An idea of the Central Composite Design (CCD).

The value of parameter α depends on so-called rotatability of an experiment. To maintain rotatability, the value of α depends on the number of experimental runs in the factorial portion of the central composite design:

$$\alpha = [\textit{number or factorial runs}]^{1/4}. \tag{6.56}$$

The value of α also depends on whether or not the design is orthogonally blocked. That is whether or not the design is divided into blocks such that the block effects do not affect the estimates of the coefficients in the 2nd order model. Under some circumstances, the value of α allows simultaneously for rotatability and orthogonality of the experiment.

6.3.1.4. D-Optimal Design D-optimal designs are most often provided by a computer algorithm. These types of designs are particularly useful when classical designs do not apply. Unlike standard classical designs such as factorials and fractional factorials, D-optimal design matrices are usually not orthogonal and effect estimates are correlated. These types of designs are always an option regardless of the type of model the experimenter wishes to fit (for example, first order, first order plus some interactions, full quadratic, cubic, etc.) or the objective specified for the experiment (for example, screening, response surface, etc.).

The design is said to be D-optimal if $|\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X}|/n^p$ is maximized where $\boldsymbol{X}$ is the expanded design matrix which has n rows (one for each design setting) and p columns (one column for each coefficient to be estimated plus one column for the overall mean). The D-efficiency statistic for comparing designs is as follows:

$$\textit{D-efficiency} = \left(\frac{|\boldsymbol{X}^T\boldsymbol{X}|_{\textit{design}}}{|\boldsymbol{X}^T\boldsymbol{X}|_{\textit{D-optimum}}}\right)^{1/p}. \tag{6.57}$$

Therefore, it compares a design against a D-optimal design—normalized by the size of the matrix in order to compare designs of different sizes. D-optimal designs are straight optimizations based on a chosen optimality criterion and the model that will be fit. The optimality criterion used in generating D-optimal designs is one of maximizing $|\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X}|$—the determinant of the information matrix $\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X}$.

This optimality criterion results in minimizing the generalized variance of the parameter estimates for a pre-specified model. As a result, the "optimality" of a given D-optimal

design is model dependent. That is, the experimenter must specify a model for the design before a computer can generate the specific treatment combinations. Given the total number of treatment runs for an experiment and a specified model, the computer algorithm chooses the optimal set of design runs from a candidate set of possible design treatment runs. This candidate set of treatment runs usually consists of all possible combinations of various factor levels that one wishes to use in the experiment.

6.3.1.5. Latin Hypercube Design The Latin Hypercube (LH) design scheme of an experiment is mostly applied in case of computer experiments. Latin hypercube sampling (LHS) provides an array that randomly samples the entire design space broken down into r^n equal-probability regions (where r is the number of runs, and n is the number of input variables). LHS can be looked upon as a stratified Monte Carlo sampling where the pairwise correlations can be minimized to a small value (which is essential for uncorrelated parameter estimates) or else set to a desired value. Additionally in LHS design experiment in comparison to Monte Carlo approach, the test points in design region do not cluster. LHS is especially useful in exploring the interior of the parameter space, and for limiting the experiment to a fixed (user specified) number of runs. The LH cube experiment can be constructed as follows:

- Selecting the number of tests n that are to be simulated.
- Dividing each factor dimension into n equidistant levels.
- Sampling for each factor n random permutation of the levels.
- Combining permutation of the factors' levels into a simulation scheme.

In case of non-box region, more levels than test points are selected and then randomly generated an LH design (LHD) on the finer level grid. If the LHD is infeasible, the process is repeated while increasing the number of levels. The LHD experiment for a non-box region is referred to as constrained LHD (Figure 6.38).

In practice different LHD schemes exist mainly because there are number of possibilities to assign levels to factor dimensions. It can be done, for instance, uniformly or randomly. Much of the attention is directed towards so called maximum distance simulation scheme for which the minimal distance between test points is maximal. The minimal distance is a measure of the space region fillingness.

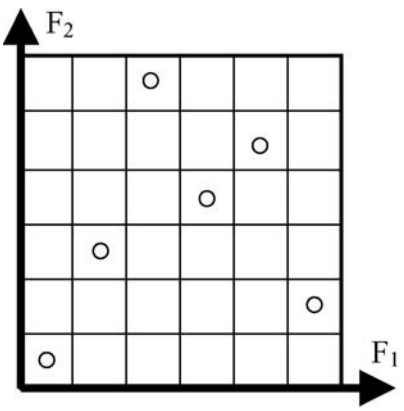


FIGURE 6.38. An idea of the Latin Hypercube experiment.

6.3.2. *Response Surface Modeling*

Thanks to the field theory, material engineering and numerical computing methods, any phenomena, process or a product can be, with some accuracy, described by a mathematical model. The model allows for predicting the behavior of a model under different conditions and can be described by a general formula, which combines the known and unknown variables influencing its behavior:

$$\boldsymbol{Y} = f(\boldsymbol{X}, \boldsymbol{Z}), \tag{6.58}$$

where matrix $\boldsymbol{Y}$ means the output variables, matrix $\boldsymbol{X}$ means the input known and controllable variables and finally matrix $\boldsymbol{Z}$ means the input unknown or uncontrollable variables, which could be referred to as noise variables.

The knowledge on model of a product or process is crucial in virtual prototyping. The model is assumed to be a black box or may require some correction according to the most current knowledge. The first step in model recognition is experiment, which in fact is a planned or unplanned series of tests. The main idea behind that is to elaborate a model, even without a complete understanding of the hidden phenomena, which would allow predicting the behavior of a process or a product within assumed input variable domain. As the response an output is selected, which is supposed to be dependent on input variables, which can be controllable or uncontrollable [22,25].

Most of the contemporary engineer tasks are directed towards proper model recognition. The traditional method is mostly based on physical experiments while the contemporary ones are more directed towards numerical experimentation. Nevertheless both methods are based on an experiment, no mater if it is physical or numerical. The method that allows for response model fitting is referenced as Response Surface Modeling (RSM) (Figure 6.39). It has been almost a rule that in many publications, the RSM is interchangeably referenced as RSA, which stands for Response Surface Analysis. The RSM method allows elaborating the mathematical model, which describes behavior of a product or process due to changes of selected essential factors. Having the model it is possible to design a contour plot of the response as a function of selected factors and selecting their most appropriate values. In the simplest cases RSM method does not require using even a computer but most often it does, especially in cases of non-linear interpolation or multi-domain problems. The benefit of RSM method is that it allows for direct application of the following advanced designing tools:

- Optimization.
- Robust design.
- Tolerance design.

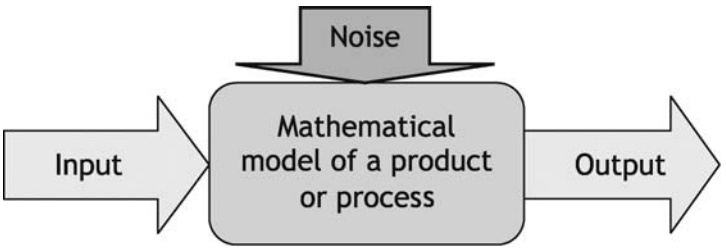


FIGURE 6.39. The idea of RSM method.

Therefore, RSM method is meant to save a lot of simulation runs in numerical prototyping or experiments in traditional prototyping. The method in most cases is refereed to as especially devoted DOE scheme for the response design. Once one knows the primary variables (factors) that affect the responses of interest, a number of additional objectives of the response design may be pursued. These include [11,19,50]:

- Hitting a target: This is a frequently encountered goal for an experiment. One might try out different settings until the desired target is "hit" consistently. Rather than experimenting in an ad hoc manner until we happen to find a setup that hits the target, one can fit a model estimated from a small experiment and use this model to determine the necessary adjustments to hit the target.
- Maximizing or minimizing a response: Many processes are being run at sub-optimal settings, even though each factor has been optimized individually over time. Finding settings that increase yield or decrease the amount of scrap and rework represent opportunities for substantial financial gain.
- Reducing variation: A product may be affected by high internal variation. Excessive variation can result from many causes: lack of having or following standards or due to certain hard-to-control inputs that affect the critical output characteristics. When this latter situation is the case, one may experiment with these hard-to-control factors, looking for a region where the surface is flatter and the process is easier to manage.
- Making a process robust: An item designed and made under controlled conditions will be later tested in the hands of the customer and may prove susceptible to failure modes not seen in the lab or thought of by design, e.g., operation under extremes of external temperature. Designing an item so that it is robust for a special experimental effort. It is possible in the lab to determine the critical components affecting its performance.
- Seeking multiple goals: A product or process seldom has just one desirable output characteristic. There are usually several, and they are often interrelated so that improving one will cause a deterioration of another. Any product is a trade-off between these various desirable final characteristics. This is done by either constructing some weighted objective function (desirability function) and optimizing it, or examining contour plots of evaluated responses.

The choice of an experiment design scheme in RSM method would depend on the behavior of a response function (model), which reflects the real life phenomena and other requirements:

- Model linearity.
- Experiment rotability.

In case of a model linearity, the possible behaviors of responses as functions of factor settings can include linear, quadratic or cubic. If a response behaves as linear, the design matrix to quantify that behavior need only contain factors with two levels and can be referenced by factorial and fractional factorial designs. If a response behaves as quadratic, the minimum number of levels required for a factor to quantify that behavior equals to three. It may be assumed that adding centre points to a two-level design would satisfy that requirement, but the arrangement of the treatments in such a matrix confounds all quadratic effects with each other. While a two-level design with centre points cannot estimate individual pure quadratic effects, it can detect them effectively. A solution to creating a design

matrix that permits the estimation of simple curvature would be to use a three-level factorial design. Finally, in more complex cases including cubic function, the design matrix must contain at least four levels of each factor to characterize the behavior of a response adequately [9,15,23].

Basic goal of RSM method is evaluation of a process or product model with an output being a good-fitting mathematical function with high predictive power, and to have good estimates of coefficients in that function with maximal accuracy. The output from process modeling is a fitted mathematical function whether based on approximation or interpolation of the data points. For a given data set the most common response would be given by a formula:

$$y = (\mathbf{X}), \tag{6.59}$$

or simply by a matrix in case of numerical evaluation. In fact, there are a number of mathematical methods, which allow for response approximation or interpolation. Additionally, in case of advanced prototyping methods as e.g., sequential procedure, the ability of approximation or interpolation error estimation would be the additional benefit and key factor for reducing number of "needed" experiments [2,37,40].

There is a distinct difference between the traditional laboratory experiments and numerical experiments, which is noise. The numerical experiments tend to give the same results while the laboratory experiments are error/noise biased. The above requires different approach to RSM model of the response:

- Laboratory experiments: approximation.
- Numerical experiments: interpolation.

The application of advanced prototyping procedures is more difficult in case of traditional experiments due to approximation method. The both approaches require different RSM models and the same different evaluation method of the model error estimation (Figure 6.40).

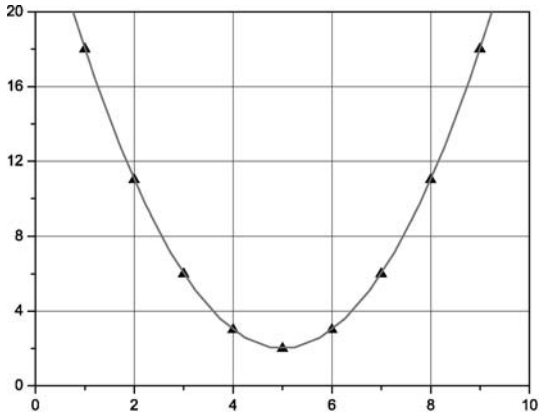
6.3.2.1. Polynomial Model For the most of response surfaces, the approximation functions are polynomials mainly because of simplicity. The most widely used are the low-order polynomials, e.g., first or second. For low curvature, a first-order polynomial can be used while for significant curvature, a second-order polynomial including all two-factor interactions.

In case of the first-order polynomial and low curvature the response surface is described as follows:

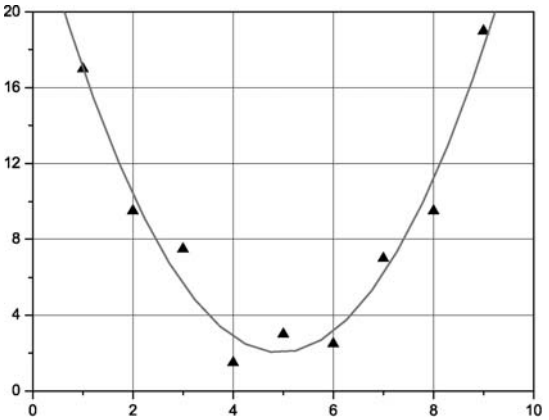
$$\hat{y} = \beta_0 + \sum_{i=1}^{k} \beta_i x_i, \tag{6.60}$$

where β's are coefficients. The above equation can be rewritten in the matrix form as:

$$\mathbf{Y} = \mathbf{BX} + \mathbf{E}, \tag{6.61}$$



(a)



(b)

FIGURE 6.40. The difference between interpolation (a) and approximation (b).

where

$$\mathbf{Y} = \begin{Bmatrix} y_1 \\ y_2 \\ \vdots \\ y_k \end{Bmatrix}, \quad \mathbf{B} = \begin{Bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{Bmatrix}, \quad \mathbf{X} = [x_1 \quad x_2 \quad \cdots \quad x_k], \quad \mathbf{E} = \begin{Bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_k \end{Bmatrix}. \tag{6.62}$$

The unbiased estimator $\boldsymbol{b}$ of the coefficient vector $\boldsymbol{B}$ is obtained using the least square error method:

$$\mathbf{b} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}. \tag{6.63}$$

The variance-covariance matrix of the $\boldsymbol{b}$ is obtained as follows:

$$\operatorname{cov}(b_i, b_j) = C_{ij} = \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}, \tag{6.64}$$

where the σ is an error of $\boldsymbol{Y}$. The estimated value of σ is obtained as follows:

$$\sigma^2 = \frac{\mathbf{Y}^T\mathbf{Y} - b^T\mathbf{X}^T\mathbf{Y}}{k-1}. \tag{6.65}$$

For the cases of quadratic polynomials and significant curvature, the response surface is described as follows:

$$\hat{y} = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \beta_{ii} x_i^2 + \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \beta_{ij} x_i x_j. \tag{6.66}$$

The coefficients β of the above polynomials are usually determined, as described above, by least squares regression analysis and fitting the response surface approximations to existing data. Higher order polynomials are used seldom because in case of more advanced response surfaces rather different approximation functions and methods are used.

6.3.2.2. Spline Model Polynomials are the approximating functions of choice when a smooth function is to be approximated locally otherwise the degree n of the approximating polynomial may have to be chosen unacceptably large. The alternative is to subdivide the interval $[a \ldots b]$ of approximation into sufficiently small intervals $[\xi_j \ldots \xi_{j+1}]$ where:

$$a = \xi_1 < \cdots < \xi_{l+1} = b, \tag{6.67}$$

so that, on each such interval, a polynomial p_j of relatively low degree can provide a good approximation of the function f. This can even be done in such a way that the polynomial pieces blend smoothly, e.g., so that the resulting patched or composite function $s(x)$ that equals $p_j(x)$ for a chosen interval $x \in [\xi_j \ldots \xi_{j+1}]$, for all j, has several continuous derivatives. Any such smooth piecewise polynomial functions are called a spline.

There are two commonly used ways to represent a polynomial spline, the pp-form and the B-form. While a spline in pp-form is often referred to as a piecewise polynomial, then B-form is often referred to as a spline. This reflects in the fact that piecewise polynomials and (polynomial) splines are just two different views of the same thing.

6.3.2.3. Stochastical Model The stochastical modeling approach is based on considering the deterministic response $y(\boldsymbol{X})$ as a realization of a stochastic process, which means that an error B is replaced by another term $Z(\boldsymbol{X})$ representing a random process. For example, computer analysis is deterministic and not subjected to a measurement error therefore the usual uncertainty derived from least-squares residuals have no meaning, therefore the response model can be treated as a combination of a polynomial model and additional factor refereeing to the deviation from the assumed model [20,26]:

$$\hat{y}(x) = \sum_{i=1}^{k} \beta_i f_i(x) + \varepsilon(x), \tag{6.68}$$

where $\varepsilon(x)$ is the systematic deviation from the assumed model, see Figure 6.41.

In fact function $\varepsilon(x)$ representing the realization of a stochastic process is assumed to have zero mean and covariance V between two inputs u and v given by:

$$V(u, v) = \sigma^2 R(u, v) \tag{6.69}$$

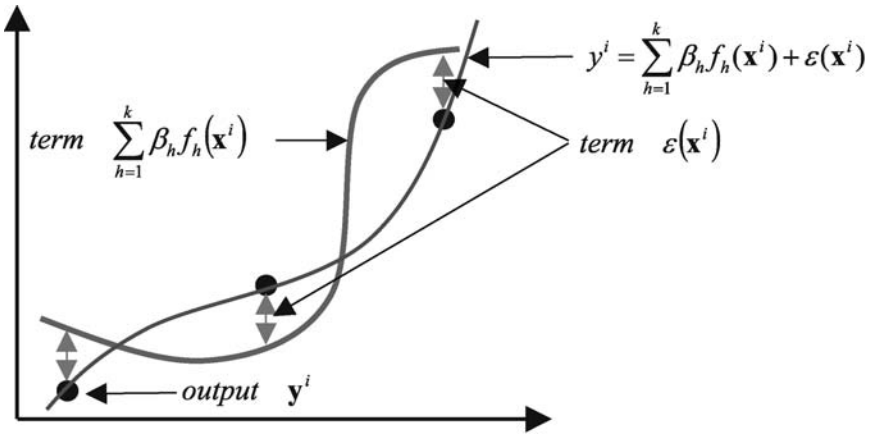


FIGURE 6.41. The concept of the stochastical model.

between $\varepsilon(u)$ and $Z(v)$, where σ^2 is the process variance and $R(u, v)$ is a correlation. The covariance structure of Z relates to the smoothness of the fitting surface. For a smooth response, a covariance function with some derivatives might be adequate, whereas an irregular response might call for a function with no derivatives. The fitting procedure can be viewed as two stage problem:

- Calculation of the generalized least-squares predictor.
- Interpolation of the residuals at the design points as if there were no regression.

One of the most popular methods for such a stochastic model interpolation is *kriging*. Kriging is extremely flexible due to the wide range of correlation functions *R(u,v)*, which may be chosen. Depending on the choice of a correlation function, kriging can either result in exact interpolation of the data points or smooth interpolation, providing an inexact interpolation. It is worth noticing that kriging is different than fitting splines and in fact it is believed even better than splines. One of the most crucial aspects of Kriging method is a problem of the interpolation error assessment. Once the interpolation error is estimated it is possible to locate a position of an additional experiment point, as required by the iterative approach procedure, which may improve the RSM model accuracy.

6.3.2.4. Radial Basis Function Model The name Radial Basis Function (RBF) comes from the function properties. Radial function is defined as function which value depends only on a distance of its argument from the center. There is a number of different radial functions:

- Multiquadratic.
- Gaussa.
- etc.

One of the most popular RBF method is based on multiquadratic functions of the form:

$$\hat{y}(x) = \sum_{j=1}^{N} C_j \varphi_j(x), \tag{6.70}$$

where matrix C_j is evaluated as:

$$A_{ij}C_j = y_i, \tag{6.71}$$

and

$$A_{ij} = \varphi_j(x_i). \tag{6.72}$$

In a standard method function $\varphi_j(x)$ is given by:

$$\varphi_j(x) = \sqrt{\|x - x_j\|^2 + h^2}. \tag{6.73}$$

In fact, this type of radial function was primarily used to approximate geographical surfaces and only recently has been adopted for optimization. In fact, approximation fitting based on RBF functions is nowadays believed to be as good as the interpolation fitting based on stochastical method.

6.3.3. Advanced Approach to Virtual Prototyping

The virtual prototyping is based on numerical whether uncorrelated (traditional) or correlated (advanced), in order to achieve usually optimal or sub-optimal designs. Numerical experiments are usually selected according to the knowledge and experience of an expert or defined according to selected experimentally/statistically orientated methodologies. The traditional virtual prototyping is based on a sequence of uncorrelated sequence of tests and procedures (Figure 6.42). The basic steps applied in a traditional virtual experiment would be based on [5]:

- Capturing the simulation sequence necessary to be able to simulate the desired design attributes, examples here are stress simulation, fatigue life simulation, thermal simulation.
- DOE (design of experiments) in order to scatter the simulations in the region of interest, according to the selected scheme, e.g., orthogonal, random and so forth.
- RSM (response surface modeling) in order to interpolate/approximate the model of the response by a mathematical model.
- Optimization, in order to find out the required response: minimum, maximum or nominal value.

Apart from the advantages well documented in the literature, there are some recognized drawbacks of the traditional virtual prototyping. The most essential ones are:

- The inability to automate the capturing and federation of the necessary simulation sequences.

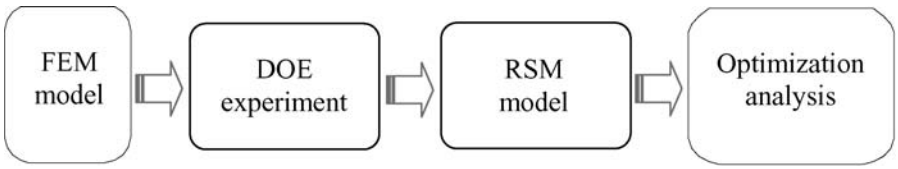


FIGURE 6.42. A simplified schematic diagram of the traditional virtual prototyping procedure.

- Efficiency, due to the required number of simulations, which grow exponentially with the number of input factors (variables).
- Quality, due to the reliability/credibility of the response model, which would be in turn used for an optimization and robust design.

The first is a limiting factor since the engineer will have to perform a lot of error prune manual work. Beyond, most often the engineer would have to decide on the compromise between the both as it seems unrealistic to e.g., improve the quality of the response model and at the same time reducing the number of tests in the chosen experiment [10,36].

In contrast to traditional methods the advanced virtual prototyping methods are based on a sequence of correlated tests and procedures. The advanced virtual prototyping methods provide the ability to capture simulation processes and further automate the running of the simulation programs. Secondly it allows for saving the total number of required experiments in order to achieve the reliable model of the response by improving the quality of the response model at a fraction of the number of experiments compared to the traditional methodologies. Figure 6.43 shows a schematic diagram of the advanced prototyping method.

First, a screening experiment is performed in order to deduct the main essential factors. Secondly, the post-screening experiment is used in order to fit the model of the response, which will be used for optimization.

The advanced scheme can be summarized by the following steps [44,45]:

- Building up numerical FEM model of a product or process capturing the physics of the problem and simulated all design critical attributes.
- Capturing and automating the design process.
- Carrying out a screening experiment based on orthogonal DOE scheme procedure in order to find out the correlation between the response and input factors (including interactions) and defining their significance in a sense of e.g., mean and variance.
- Selecting the most essential/significant input factors and adding additional experiment tests according to the elaborated modified LH design scheme.
- Interpolating/approximating the initial RSM model of the response in a form of a response surface reflecting relationship between the response and the most significant factors.
- Implementing the iterative approach in order to improve the final model of the response by sequential adding additional experiment points basing on the estimation of the interpolation error.
- Running up an optimization and sensitivity analysis to find out the best levels of the essential factors due to the expected response in the evaluated region.

Nowadays more and more companies are applying the advanced prototyping approach in order to design more reliable and high quality products. Though the current

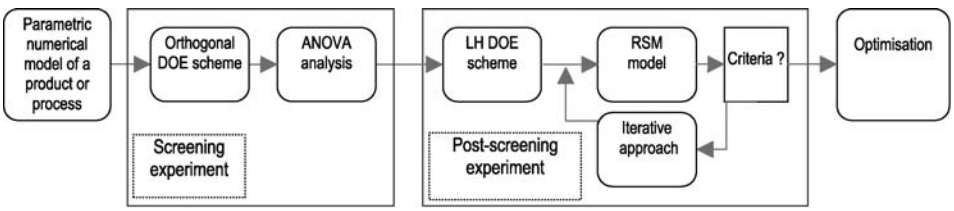


FIGURE 6.43. The advanced prototyping procedure.

procedures of DOE and RSM are broadly used worldwide, is has been a quite of an effort to adopt them for numerical prototyping in the area of electronic packages and assemblies. This is mainly due to increasing level of packaging and higher functionality of products and partly due to scale change (micro to nano). The above is due to: high non-linearity, multi-domain, multi-response, multi-factor interactions and high number of input variables. A few solutions were suggested to solve those problems: smart and sequential DOE and RSM methods, knowledge base support, etc. Nevertheless, the main goal is to reduce the number of required experiments and to improve reliability of the applied procedures.

6.3.3.1. Sequential RSM The sequential RSM (SRSM) optimization method was introduced mainly in reference to computer-aided design of experiments but in fact can be used with some modifications to traditional experiments as well [12,13,33]. The main idea of sequential method is to find a good fitting curve to the optimum of an unknown function of several variables in a minimum number of function evaluations–experiments. This can be achieved by sequential exploring of the domain of interest. The idea behind the sequential approach is to perform an iterative procedure to improve the RSM model of the response by adding additional one at a time points of experiment (Figure 6.44). The method is based on a two-step approach:

- Initial experiment based on space filing DOE scheme, e.g., LH.
- Iterative experiment based on RSM model and interpolation error estimation, e.g., Kriging model.

At the first stage an initial number of experiments is selected using e.g., Latin Hypercube method and then if required additional experiments are performed, which would improve the model accuracy. At each stage an interpolating function, derived from e.g., stochastic RSM model (Kriging) of the objective function, is set up, and this is used to determine the location of the next function evaluation and afterwards the model improvement. This process continues until agreement is reached between the optimum interpolating function value and the true value of the objective function. A balance between exploring unknown regions and optimizing the function in known regions is struck by means of a weighting factor, which varies as new data are accumulated.

The main problem is a choice of initial tests N to explore the region of interest. If N is too large, then more function evaluations will be carried out than are required for a good appreciation of the general form of the objective function. If N is too small, then the exploration of the region of interest will be insufficient and possible optimum locations may be overlooked. Obviously, the number of data points needed to "explore" the region effectively depends on a structure of the function, which is not known a priori. Data points are inserted into the region of interest one at a time and the parameters of the interpolating

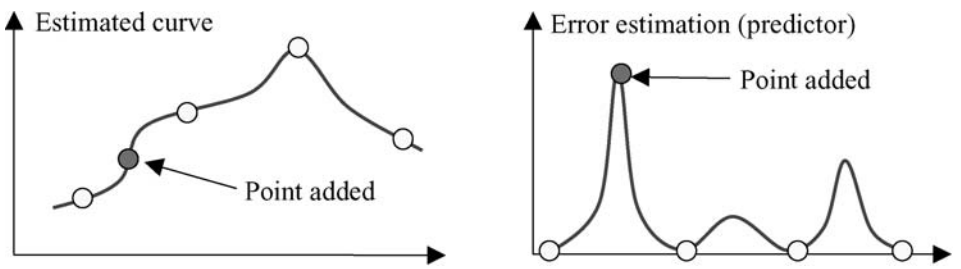


FIGURE 6.44. The advanced prototyping procedure.

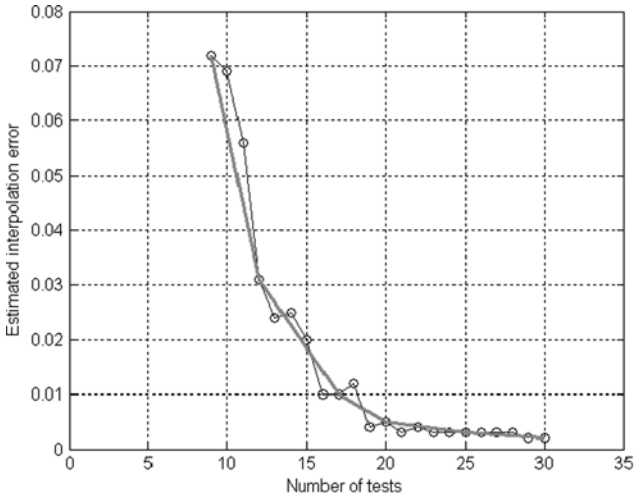


FIGURE 6.45. Example of the estimated interpolation error as a function of tests.

function are constantly updated. To define a position of a new data point, a balance is struck between the twin criteria of "exploration" and "optimization." In the early stage ("exploration") a new point is positioned as far as possible away from all existing points. In the later stages ("optimization") new points tend to concentrate around optimum values of the interpolating function. The balance between these twin criteria is defined by means of a weighting factor, which depends not only on the current number of data points but also on the apparent structure of the objective function (Figure 6.45).

The sequential approach to exploring unknown functions for optimization has the advantage of not depending on an estimate of the number of initial data points required to explore fully the region of interest. The shape of the function itself determines, to a certain extent, the point at which the "exploration" gives way to "optimization."

6.3.3.1.1. The Kriging Error Estimation. Application of the sequential approach requires one of the most crucial aspects of Kriging method, which is an assessment of the interpolation error. The idea of the sequential approach algorithm is to use the interpolation curve (predictor) together with the knowledge about the accuracy of the interpolation (standard error) in order to iteratively add points in the design space in those locations where the *expected improvement* of the objective function is the highest, as mentioned earlier, the function $\varepsilon(\mathbf{x}^i)$ representing the realization of a stochastic process is assumed to have zero mean and covariance V between two inputs u and v are given by [51–53]:

$$V(u, v) = \sigma^2 R(u, v) \tag{6.74}$$

between $\varepsilon(u)$ and $\varepsilon(v)$, where σ^2 is the process variance and $R(u, v)$ is a correlation. The covariance structure of ε relates to the smoothness of the approximating surface. For a smooth response, a covariance function with some derivatives might be adequate, whereas an irregular response might call for a function with no derivatives. The fitting procedure can be viewed as two stage problem:

- calculation of the generalized least-squares predictor,
- interpolation of the residuals at the design points as if there were no regression.

As in computer simulation, being deterministic by its nature, the error is totally due to modeling error and not to e.g., measurement error or inner or outer noise, then it is justified to treat the error $\boldsymbol{\varepsilon}^i$ as a continuous function of $\mathbf{x}^i$:

$$\boldsymbol{\varepsilon}^i = \boldsymbol{\varepsilon}^i(\mathbf{x}^i). \tag{6.75}$$

As the error is the continuous function then the errors could be considered as correlated by the distance function between the points. If points are close, then the relevant errors should also be comparable. This means high correlation. Therefore it can be assumed that the correlation between errors would be related to the distance between the corresponding points. As the distance function can be used a special weighted distance formula, which in comparison to the Euclidean distance does not weights all the variables equally:

$$d(\mathbf{x}^i, \mathbf{x}^j) = \sum_{h=1}^{k} \Theta_h |x_h^i - x_h^j|^{p_h}, \tag{6.76}$$

where $\Theta \geq 0$ and $p_h \in [1, 2]$. Using this distance function, the correlation between the errors can be defined as follows:

$$corr[\varepsilon(\mathbf{x}^i), \varepsilon(\mathbf{x}^j)] = \frac{1}{e^{d(\mathbf{x}^i, \mathbf{x}^j)}}. \tag{6.77}$$

The so defined correlation function has obvious properties, which means that in case of small distance the correlation is high while in case of large distance the correlation will approach zero. The values of the correlation function define the correlation matrix $\mathbf{R}$ of the order $n \times n$, which has practical meaning in the final response model definition:

$$r_{i,j} = corr[\varepsilon(\mathbf{x}^i), \varepsilon(\mathbf{x}^j)], \tag{6.78}$$

$$\mathbf{R} = \begin{pmatrix} r_{1,1} & \cdots & r_{1,n} \\ \vdots & & \vdots \\ r_{n,1} & \cdots & r_{n,n} \end{pmatrix}, \tag{6.79}$$

where the values of matrix $\mathbf{R}$ depend on parameters (θ_h, p_h). Thanks to the so defined correlation function and the correlation matrix $\mathbf{R}$ it is possible to get a simple linear regression model and avoid a quite complicated functional form of the response. The evaluation of the so defined stochastic model has a very important virtue, which allows replacing the regression terms by the constant value μ:

$$\mu = \sum_{h=1}^{k} \beta_h f_h(\mathbf{x}^i), \quad i = 1, \ldots, n, \tag{6.80}$$

and the same the stochastic model of the response can be rewritten as follows:

$$y(\mathbf{x}^i) = \mu + \varepsilon(\mathbf{x}^i), \quad \varepsilon(\mathbf{x}^i) \to N(0, \delta^2). \tag{6.81}$$

Therefore in order to define the stochastic model of the response it is required to estimate $2k + 2$ parameters: μ, δ^2, $\theta_1, \ldots \theta_2$, $p_1, \ldots p_k$. This task can be achieved by maximizing the likelihood function F of the sample, which is defined as follows:

$$F = \frac{1}{(2\pi)^{n/2}(\delta^2)^{n/2}|\mathbf{R}|^{1/2}} \frac{1}{e^{\frac{(\mathbf{y}-\mathbf{1}\mu)'\mathbf{R}^{-1}(\mathbf{y}-\mathbf{1}\mu)}{2\delta^2}}}, \tag{6.82}$$

where $\mathbf{1}$ denotes the n-vector of ones and $\mathbf{y}$ denotes the n-vector of observed function values:

$$\mathbf{y} = \left(y^1, y^2, \ldots, y^n\right)'. \tag{6.83}$$

The estimators of parameters μ and δ^2 that maximize the likelihood function are given in a closed form by:

$$\hat{\mu} = \frac{\mathbf{1}'\mathbf{R}^{-1}\mathbf{y}}{\mathbf{1}'\mathbf{R}^{-1}\mathbf{1}}, \tag{6.84}$$

$$\hat{\delta}^2 = \frac{(\mathbf{y} - \mathbf{1}\hat{\mu})'\mathbf{R}^{-1}(\mathbf{y} - \mathbf{1}\hat{\mu})}{n}. \tag{6.85}$$

By substituting the above in the likelihood function ones gets the so-called "concentrated likelihood function," which depends only on parameters (θ_h, p_h):

$$L = L(\Theta_h, p_h). \tag{6.86}$$

Optimization of this function gives finally the estimates of parameters (θ_h, p_h) and hence the estimate of the correlation matrix $\mathbf{R}$. Finally it is possible to evaluate the estimates of μ and δ^2:

$$(\hat{\Theta}_h, \hat{p}_h) = \max_{(\Theta_h, p_h)} L(\Theta_h, p_h). \tag{6.87}$$

The best linear unbiased estimator of the response value y at point $\mathbf{x}^*$ is defined as:

$$y(\mathbf{x}^*) = \hat{\mu} + \mathbf{r}'\mathbf{R}^{-1}(\mathbf{y} - \mathbf{1}\hat{\mu}), \tag{6.88}$$

where the $\mathbf{r}$ is the n-vector matrix given as follows:

$$r_i(\mathbf{x}^*) = corr[\varepsilon(\mathbf{x}^*), \varepsilon(\mathbf{x}^i)]. \tag{6.89}$$

It is very important to assess the estimation of the prediction accuracy at point $\mathbf{x}^*$, which can be evaluated as the mean squared error $s^2(\mathbf{x}^*)$ as follows:

$$s^2(\mathbf{x}^*) = E\left\{[\hat{y}(\mathbf{x}^*) - y(x^*)]^2\right\} = \hat{\delta}^2\left[1 - \mathbf{r}'\mathbf{R}\mathbf{r} + \frac{(1 - \mathbf{1}'\mathbf{R}^{-1}\mathbf{r}')}{\mathbf{1}'\mathbf{R}^{-1}\mathbf{1}}\right]. \tag{6.90}$$

Most often it would most convenient to work with the square root of the mean squared error $s(\mathbf{x})$ instead:

$$s(\mathbf{x}) = \sqrt{s^2(\mathbf{x})}. \tag{6.91}$$

This provides a standard error or estimated standard deviation for measuring uncertainty in the response prediction.

The parameter Θ can be treated as a measuring factor of the importance of the variable x_h, which can be related to statement that even small values of:

$$d_h^{i,j} = \left| x_h^i - x_h^j \right| \tag{6.92}$$

may lead to large differences in the function values at $\mathbf{x}^i$ and $\mathbf{x}^j$. Therefore from the statistical sense it can be stated:

- If parameter Θ_h is large then small values of $d_h^{i,j}$ would lead to high distance d value and hence low correlation *corr*.
- If parameter Θ_h is small then small values of $d_h^{i,j}$ would lead to small distance d value and hence high correlation *corr*.

Anyway, one of the most crucial factors of Kriging method is estimation of Θ parameter. It can be done according to the calculus of variations and finding the extreme of the defined parametric functional $F(\Theta, p)$ as a total error of the interpolation, which can be defined as:

$$F(\Theta, p) = \int_{\mathbf{x}'}^{\mathbf{x}''} s(\mathbf{x}, \Theta, p) d\mathbf{x}, \tag{6.93}$$

or in case of sampled data it can be evaluated numerically as:

$$F(\Theta, p) = \sum_{\mathbf{x}'}^{\mathbf{x}''} s(\mathbf{x}, \Theta, p) \Delta\mathbf{x}, \tag{6.94}$$

where $\mathbf{x}$ is a vector of the prediction points, the $s(\mathbf{x})$ is the square root of the mean squared error of the interpolation:

$$s(\mathbf{x}, \Theta, p) = \sqrt{\hat{\delta}^2 \left[1 - \mathbf{r}'\mathbf{R}\mathbf{r} + \frac{(1 - \mathbf{1}'\mathbf{R}^{-1}\mathbf{r}')}{\mathbf{1}'\mathbf{R}^{-1}\mathbf{1}} \right]}. \tag{6.95}$$

According to the previous considerations it can be noticed that both matrix $\mathbf{R}$ and vector $\mathbf{r}$ are the functions of θ and p while the estimator of δ^2 is a function of the correlation matrix $\mathbf{R}$. As the parameter p is most often selected at value 2, in fact $p \in [1..2]$, therefore the best estimator of θ can be found as the value that minimizes the functional $F(\Theta, p = 2)$:

$$\hat{\Theta} = \min_{\Theta} F(\Theta, p = 2). \tag{6.96}$$

6.3.3.2. Multiresponse Analysis In many experimental situations, a number of responses are measured for each setting of a group of design variables. The goal of the multiresponse analysis would be to find out the optimal solution due to a few responses whether a real one or most often the compromised [4,7,14].

There are some works which stress the importance of analyzing multiresponse by means of multivariate techniques that take into account interrelations among the responses [24]. For example, the models that represent the responses may have several parameters in common. It would, therefore, make sense to combine information from all

the responses to estimate these parameters. Some problems connected with multiresponse analysis are:

- Multiresponse estimation.
- Multiresponse design of experiments.
- Multiresponse optimization.

An example of multiresponse estimation may be one proposed by Box and Draper, who developed a method of estimating the parameter vector B in a general multiresponse model. The responses are assumed to be normally distributed and to have a variance-covariance matrix Σ considered constant for the various runs within the experimental region. Box and Draper used a Bayesian argument by considering a non-informative prior distribution for B and Σ. Estimates of the elements of B are obtained by maximizing the marginal posterior density of B. This method is known as the Box-Draper estimation criterion. It applies to linear as well non-linear models. There are some drawbacks of the above method and one of them is that in case of exact linear relationships among the responses, the Box-Draper estimation criterion can lead to meaningless results [41,42].

Multiresponse optimization applies to, in fact, most of the engineer or research problems. This means, that conditions that are optimal for one response may be far from optimal or even physically impractical for the other responses. RSM methods introduce some solution for such multiresponse optimization:

- Graphical, based on superimposing response contours and visual searching for a common region where the responses achieve near optimal values or to find a location of a "compromised" optimum. Unfortunately, this procedure is difficult and almost impossible to apply when the umber of responses is greater than three.
- Another approach is based on an assumption that each response function undergoes a certain transformation into a desirability function ϕ such that $0 \leq \phi \leq 1$. The choice of transformation depends on a subjective judgment concerning the importance or desirability of the corresponding response values. A measure of the overall desirability of the responses is obtained by combining the individual desirability functions through the use of geometric mean.
- Advanced approach is based on a procedure for the simultaneous optimization of responses that are represented by linear multiresponse model. A distance function is chosen that measures the overall closeness of the response functions to achieving their respective optimal values at the same set of operating conditions. Optimum operating conditions are then derived by minimizing this distance function over the experimental region.

Unfortunately, multiresponse analysis is not as developed as its single-response counterpart. It is still relatively new, and its utility has yet to be fully appreciated. This is mainly attributed to the fact that it requires advanced numerical procedures. A lot depends as well on an expert knowledge, which can be helpful in defining so-called objective function for a number of responses.

6.3.4. Designing for Quality

Design for quality is the final stage of virtual prototyping. In the simplest case, it can be only devoted to optimization as finding the optimal factor/parameter values in a sense of the expected output, whether maximum, minimum or a nominal value. In general, it should

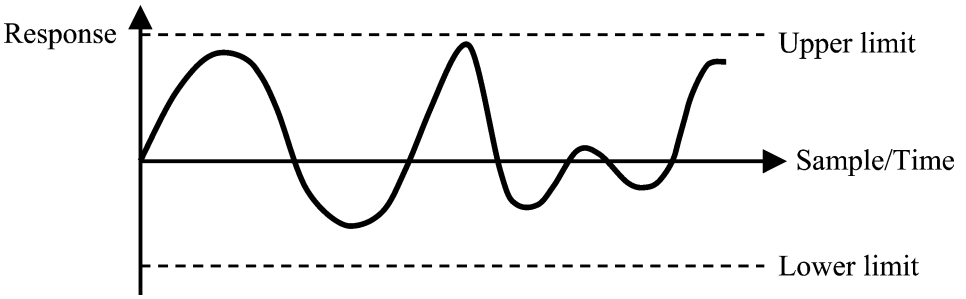


FIGURE 6.46. The scatter of the response due to time, noises, operation conditions or production phase.

include additionally such stages as robust and/or tolerance design [3,29,35]. If we have a product or a process and a selected output variable referred to as a response, then its value will differ from product to product or along timescale due to:

- Noises.
- Operation conditions.
- Production phase caused by induced random distribution of geometrical dimensions or material properties.
- Material or process parameters.

Taking into account all the above, the selected product sample could be described by random distribution with an average and variance response (Figure 6.46). As long as its average is close to the expected one and variance is low, the loss function would be small but otherwise the loss function will be high. In fact the quality of a product can be measured by such parameters as: nominal, minimal or maximal value. Additional to that, tolerances of selected factors to acceptable levels should be defined accordingly (Table 6.3). The more sample responses are within the stated limits the higher quality of a product and lower loss [17,27].

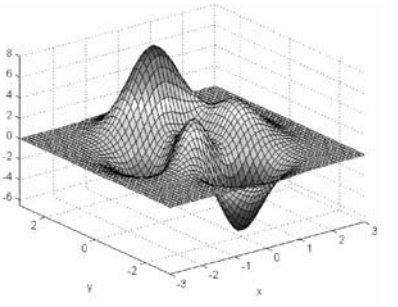
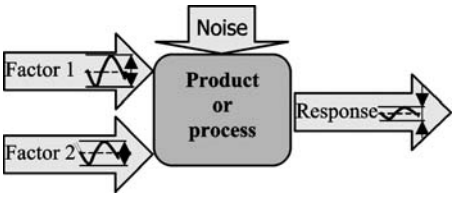
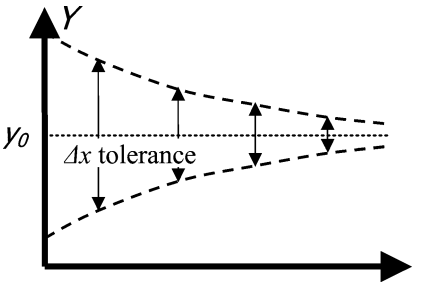
In the last few decades the quality design has been revolutionized by the innovative approach elaborated by Genichi Taguchi. At first he wrote a book devoted to experiment design and a few years later a following book on the signal to noise rate. Nevertheless his main idea was based on introducing the loss function in the experiment design. The goal was to improve the quality of the process or products so as the lost caused by the need of having them mended or improved was as low as possible. The loss does not only refer to the company profit but primarily to the society by e.g., environment pollution, noise, client complaint and so on. Therefore the higher quality then the so defined loss function value is lower. In comparison with other theoreticians, Taguchi prefers to refer to quality lost rather than the quality itself. The typical loss function L would be defined as [34]:

$$L = k\lfloor S^2 + (\overline{y} - m)^2 \rfloor, \tag{6.97}$$

where k is a constant, S is a variance of the output signal sample, m is a designed average while y is an average value of the output signal sample. Figure 6.47 shows a typical sketch of the loss function.

Though Taguchi method is very handy and does not require high knowledge on statistics, in fact it is directed towards engineers. It allows for designing such a product or process that would be satisfying to the client and at the same time reduce the costs of the

TABLE 6.3.
Design for quality characteristic.

Optimization	• The objective is to find the operating conditions or factor levels X that would optimize the system response y. • The common problem of the optimization is how to distinguish the local optimum from the global one.	
Robust design	• The aim is to make a product or process less sensitive (more robust) in the face of variation over which we have little or no control. • The robust design can be based on the Monte Carlo approach.	
Tolerance design	• The tolerance design can be performed if the robust design is not enough. • The goal is to tighten up the tolerances so as the response can be set up in the acceptable range and e.g., balanced quality vs. cost.	

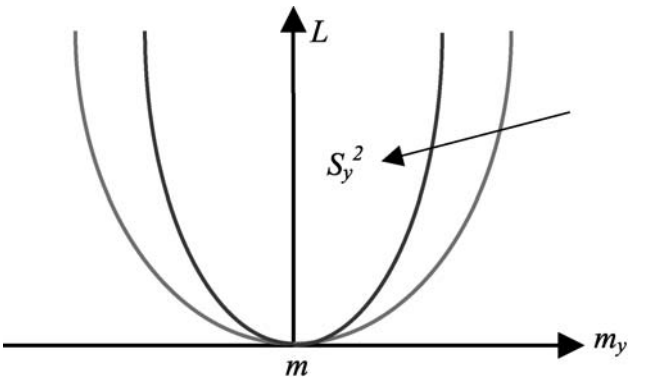


FIGURE 6.47. A sketch of a typical loss function.

company. Primarily it is based on orthogonal DOE scheme and ANOVA analysis. Though, no questionable simplicity of Taguchi approach to quality design there are some drawbacks

of which the ability of analyzing highly non-linear responses and multiple responses are the most crucial ones. In such a case, a more precise analysis would be required including advanced approach, especially in case of virtual prototyping. The method could be based on sequential DOE and RSM, which lead to more accurate product or process model in the whole domain, and can be used for optimization, robust and tolerance analysis.

As mentioned already the main reason for product or process scatter is due to controllable and uncontrollable factors. The primary task would be the recognition of controllable and uncontrollable or difficult to control factors and defining their power. In fact the lower number of uncontrollable factors the better quality of a product can be achieved, which is depicted by lower loss function value.

Most often the question would be not only how to improve the product quality but additionally how to do it with the lower cost e.g., by changing construction or adjust process parameters. This task can be performed during the designing and implementation phase. For that a product design would be required and implementation of the Quality Function Deployment QFD method. The QFD method refers to proper adjusting of the product functionality and quality so as the expectation of the clients could be met. The next phase would be designing and testing for the best input variable values and defining their quality. The goal is to achieve such a design that would be the least sensitive to outer and inter noises. If this goal is achieved than we can go to the cost reduction. At least we can be able to improve the quality and keeping the cost at the same level, which is desirable as well.

6.4. APPLICATION CASE

Numerical prototyping method can be used to predict, evaluate, optimize, and eventually qualify the thermo-mechanical behavior of electronic packages against the actual package requirements prior to major physical prototyping and manufacturing investments. Any electronic packaging is strongly non-linear, including: material non-linearity's, such as visco-plasticity, creep and/or elasto-plastic behavior, geometric non-linearity's, such as large deformation, boundary non-linearity's as edge and contacting effects. Reliable and efficient FEM-based thermo-mechanical prediction models can only be obtained if such non-linearity is taken into account.

In order to present the basics of virtual prototyping an example case is demonstrated. As could be expected there are a number of possible approaches to solve a certain problem, primarily depending on a final goal but additionally on a possibility of experimental verification, available software tools, time of evaluation, expected improvement, experience of a person and design cost.

6.4.1. Problem Description

One of the latest developments in packaging technology is the so-called exposed pad family, for instance QFN (Quad Flat Non-lead) package. An exposed pad package is a package composed of an Integrated Circuit (IC) attached to an exposed pad and in a later stage encapsulated with an epoxy moulding compound. It has been introduced into the semi-conductor market as a thin, cost effective, thermal and high frequency package solution. The exposed pad is a metal plate that is located on the bottom of the package. Exposed pads on the top of the package are less common but they exist. Many variations exist; exposed pads are found on many packages types. Mature package types with gull

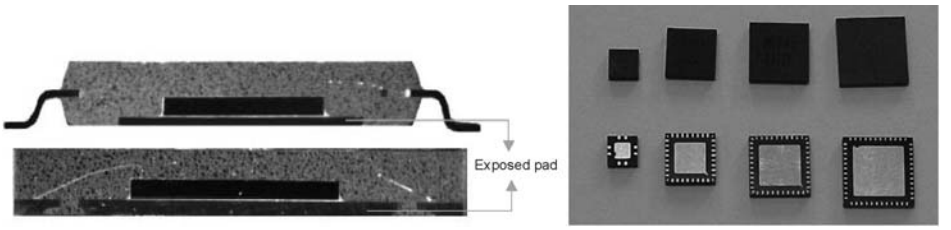


FIGURE 6.48. Examples of exposed pad packages; cross-section (left) and 3D view (right).

wing leads, such as TSSOP, offer exposed pads as an optional configuration. The exposed pad is a standard feature for QFN packages. For the leaded packages with a gull wing lead, exposed pad products are made using leadframes with a 'deep downset' paddle which is exposed to the outside of the package after the mold process. Figure 6.48 shows examples of exposed pad packages. Exposed pad packages are now on the market and exhibit a number of advantageous over other packages. Because of their size, price and performance the packages typically can be found in mobile phones and laptop computers. The presented prototyping method is applied to a typical application of a microelectronic exposed pad package. For these packages, it is vital to optimize the thickness of the leadframe toward the following restrictions [38,46]:

- Costs: minimal material (copper) will be beneficial.
- Reliability: during processing a vertical die-crack phenomenon may occur which is strongly related to the thickness of the leadframe.

6.4.2. Numerical Approach to QFN Package Design

Exposed pad packages have a very simple construction, see Figure 6.49. There is a flat leadframe, which consist of a square or rectangular diepad. The diepad in surrounded by small island of copper (which could be called leads), which are used to create to contact from the die to these islands by means of a gold wire. The die is glued on the diepad. And moulding compound in the box form shape covers the complete construction.

In order to design the parametric numerical model of the selected case, following assumptions were applied: 2D model with axi-symmetric elements were found to match well with the reality, the displacement in the x direction was fixed along the symmetry axis, the node at the left bottom corner was fixed in both the x and y directions, the single-crystal silicon die was modeled as a linear elastic material, the leadframe material was modeled as an ideally elasto-plastic material, Young's modulus and the yield stress were temperature dependent, molding compound was modeled as linear elastic material where both the Young's modulus E and the coefficient of thermal expansion were temperature dependent, the Poisson's ratio for the compound was estimated as 0.33. As the output two stresses at the top and bottom of the silicon-die surface were selected, which seemed to induce the failure of vertical die cracking after soldering and moulding process. Therefore it was necessary to include the process dependent numerical model, which is presented in Figure 6.50. The geometrical design values and the design space are listed in Table 6.4.

The phenomenon of vertical die-crack is related to the stress levels in the chip. Allowable silicon stress levels are provided by tensile and compression tests:

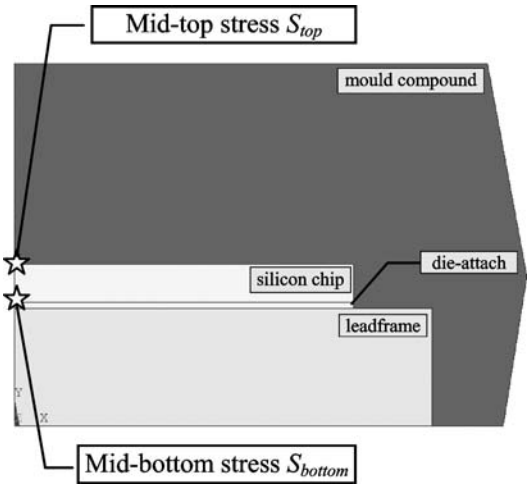


FIGURE 6.49. The parametric numerical model of a QFN product.

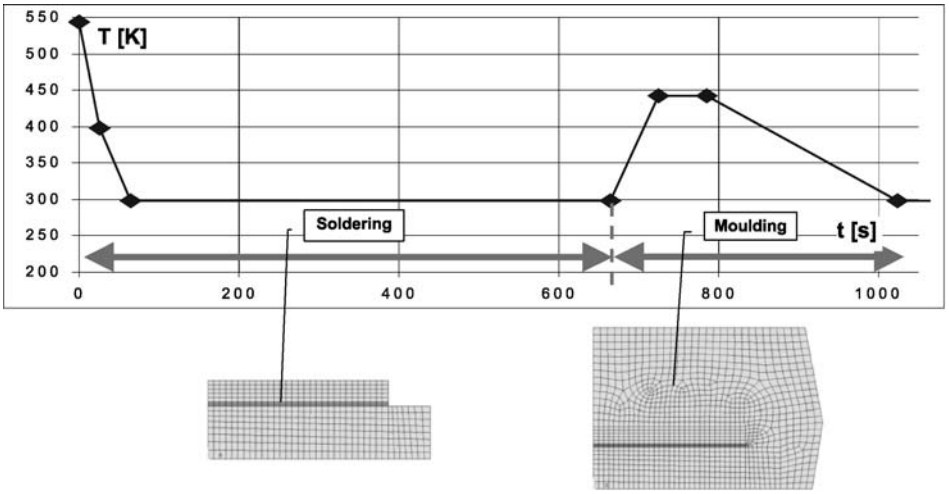


FIGURE 6.50. The major process specification, including: soldering and moulding.

TABLE 6.4.
The design materials, corresponding models and parameter values.

Material	Model	Parameters and symbols	Nominal values (mm)	Screening experiment	Post-screening experiment (mm)
Compound	linear-elastic, with:	length	6.5	None	None
	$E = f(T), a = f(T)$	thickness	2.3	None	None
Silicon die	linear-elastic	length L_{chip}	4.3	3.87–4.73	3.3–5.3
		thickness h_{chip}	0.240	0.216–0.264	None
Leadframe	elasto-plastic, with:	length $L_{leadframe}$	5.3	4.77–5.83	None
	$E = f(T), Ys = f(T)$	thickness $h_{leadframe}$	0.6	0.54–0.66	0.2–1.0
Solder	visco-plastic, with:	length	4.3	None	None
	$E = f(T), Ys = f(T)$	thickness h_{solder}	0.05	0.045–0.055	None

- In tensile direction the maximum stress is approximately +150 MPa, but strongly depends on the surface treatment.
- In compressive direction the maximum stress is approximately −500 MPa.

In contrast to traditional methods the presented approach is based on an advanced sequential approach. First, a screening experiment is performed in order to deduct the main parameters responsible for the maximum stress levels at the top and bottom side of the chip. Secondly, the post-screening experiment is used to reduce these stress levels while changing the significant geometrical parameters.

6.4.2.1. Screening Experiment Table 6.5 shows the results of the screening experiment. The whole experiment was based on a two level fractional factorial orthogonal experiment according to the L16 scheme (Figure 6.51).

According to the presented results and corresponding ANOVA (analysis of variance) analysis it was concluded that there are three dominating parameters: length of the chip L_{chip}, thickness of the leadframe $h_{leadframe}$ and thickness of the chip h_{chip}. Finally, it was decided that for the next stage only two: chip length L_{chip} and leadframe thickness

TABLE 6.5.
The screening experiment and results.

No	L_{chip}	h_{chip}	h_{solder}	$h_{leadframe}$	$L_{leadframe}$	S_{top} (MPa)	S_{bottom} (MPa)
1	1	1	1	1	1	−32.54	−3.85
2	1	1	1	2	2	−28.13	−14.66
3	1	1	2	1	2	−31.47	−5.16
4	1	1	2	2	2	−26.89	−15.34
5	1	2	1	1	2	−32.05	8.53
6	1	2	1	2	1	−27.12	−2.49
7	1	2	2	1	1	−31.57	9.02
8	1	2	2	2	2	−25.55	−4.07
9	2	1	1	1	2	−40.26	−5.82
10	2	1	1	2	1	−35.69	−17.36
11	2	1	2	1	1	−39.15	−4.15
12	2	1	2	2	2	−33.30	−19.01
13	2	2	1	1	1	−41.05	13.31
14	2	2	1	2	2	−34.42	−3.43
15	2	2	2	1	2	−39.03	11.25
16	2	2	2	2	1	−32.96	−2.49

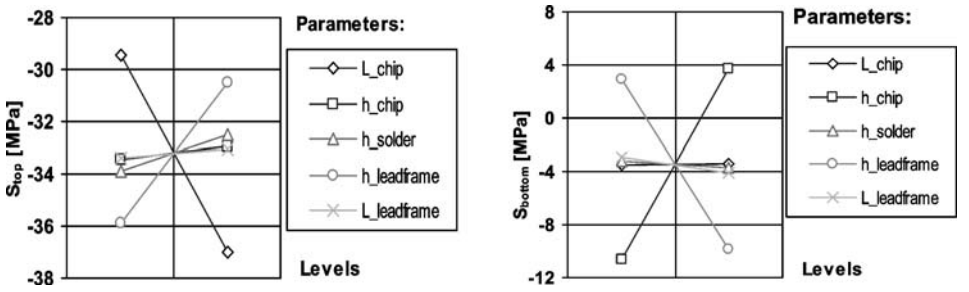


FIGURE 6.51. Main factor effects for S_{top} and S_{bottom} stress.

$h_{leadframe}$ out of three essential parameters were selected why the third parameter: chip thickness h_{chip} was set at the nominal value.

6.4.2.2. Post-Screening Experiment Table 6.6 and Figure 6.52 show the generated LH scheme and results of the sequential approach for the selected design parameters. The ex-

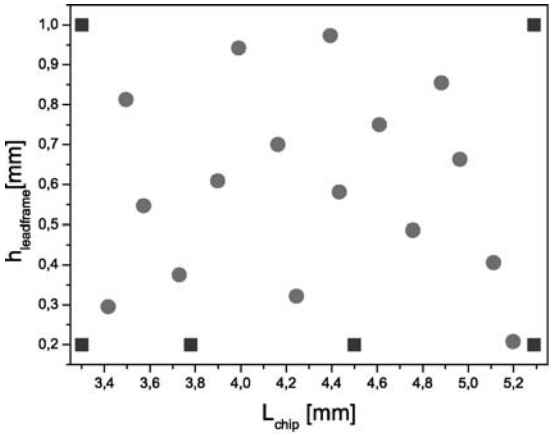
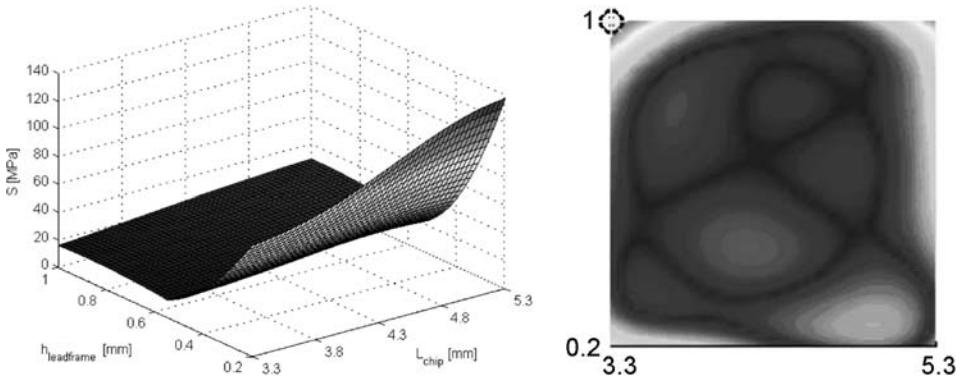


FIGURE 6.52. Selected LH scheme, where: ● initial experiments and ■ additional experiments.

TABLE 6.6.
The sequential experiment and results.

Test No	L_{chip} (mm)	$h_{leadframe}$ (mm)	S_{top} (MPa)	S_{bottom} (MPa)
Initial tests				
1	3.417	0.295	−54.03	56.74
2	3.495	0.813	−20.21	−15.11
3	3.573	0.547	−28.40	1.02
4	3.730	0.375	−45.00	33.99
5	3.898	0.609	−28.67	−4.70
6	3.991	0.942	−22.64	−20.50
7	4.163	0.700	−28.03	−11.73
8	4.245	0.322	−60.97	60.10
9	4.394	0.973	−25.34	−23.22
10	4.433	0.581	−34.58	−1.74
11	4.609	0.750	−30.33	−16.00
12	4.757	0.486	−43.98	14.17
13	4.882	0.855	−30.44	−22.22
14	4.964	0.663	−35.60	−10.76
15	5.112	0.406	−57.24	38.22
16	5.198	0.208	−112.92	149.96
Sequential tests				
17	3.30	1.00	−16.72	−18.39
18	3.30	0.20	−72.89	95.01
19	3.78	0.20	−83.76	109.95
20	5.29	1.00	−31.51	−29.54
21	5.29	0.20	−117.68	158.24
22	4.50	0.20	−99.68	132.43



(a) (b)

FIGURE 6.53. The combined stress (a) and the evaluated interpolation error (b).

periment was based on LH scheme with 16 initial tests. Basing on the interpolation error estimation additional 6 tests were done, which were to improve the overall accuracy of the final RSM model of the response.

The post-screening experiment was extended in order to include the multi-response, which seems to be important in daily engineering practice. For the presented case, in order to minimize the overall stress, two selected stresses S_{top} and S_{bottom} were combined by the following formula and then analyzed, which is presented in Figure 6.53.

$$S = 0.5(|S_{top}| + |S_{bottom}|). \tag{6.98}$$

That in turn allowed for optimization of the selected design parameters: chip length L_{chip} and leadframe thickness $h_{leadframe}$.

In order to verify the accuracy of the evaluated RSM model there was a comparison made between the predicted results and the simulation results. In order to do that, a few additional numerical experiments were done and compared with the prediction given by the RSM model. The experiment points were selected randomly within the domains of L_{chip} and $h_{leadframe}$. The verification results for the combined stress are presented in Figure 6.54. The performed verification experiment confirmed high accuracy of the RSM model, which is lower than 2% that allows for the next step of prototyping, which is optimization.

6.4.2.3. Optimization The final step of the prototyping is always devoted to optimization. In the simplest case optimization is only directed towards finding the optimal factor/parameter values in a sense of the expected output, whether maximum, minimum or a nominal value. Figure 6.55 shows the results of this optimization step.

According to the results presented in figure the expected solution for the analyzed case can be found by minimizing the objective function, which is achieved for the following factor/parameter values:

- F1: $L_{chip} = 3.3$ (mm),
- F2: $h_{leadframe} = 0.55$ (mm).

In more general case optimization should include as well sensitivity analysis due to the induced scatter of the factor/parameter values. In fact, the scatter is quite difficult to be defined a priori but it could be described by the normal density distribution. Sensitivity analysis can be achieved by Monte Carlo analysis. In case of physical experiments,

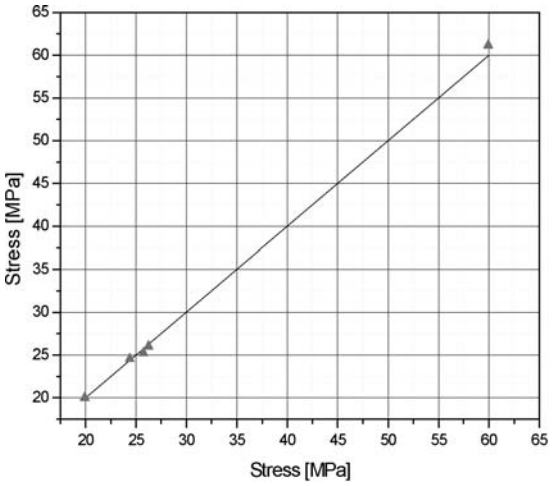


FIGURE 6.54. The results of the verification experiment.

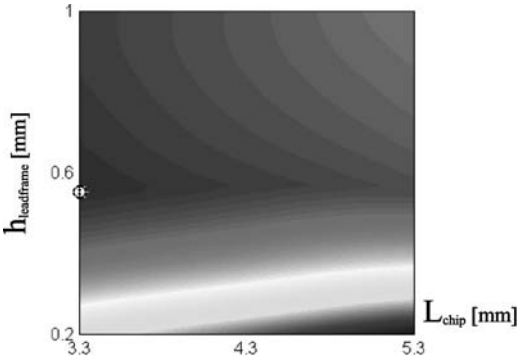


FIGURE 6.55. The optimal stress value S search for the combined stress.

Monte Carlo analysis is not feasible as it requires a large number of experiments, e.g., more than 1000. Nevertheless, in case of numerical experiments, Monte Carlo analysis can be reduced to RSM model of the defined response. Unfortunately, the whole optimization procedure requires precise RSM model, which is difficult to be precisely defined unless the whole prototyping procedure is supported by the expert knowledge.

For the current case, the Monte Carlo analysis of the defined response is given in Figure 6.56. The optimal solution can found by minimizing the both: objective and sensitivity function. The optimal solution is then located at the following parameter values:

- F1: $L_{chip} = 3.3$ (mm),
- F2: $h_{leadframe} = 0.60$ (mm).

The main difference between the optimal solution without and with sensitivity analysis is due to leadframe thickness $h_{leadframe}$ (factor F2). The optimization results with the sensitivity analysis can be interpreted as a need of changing the value of leadframe thickness $h_{leadframe}$ by 0.05 (mm) from 0.55 (mm) to 0.60 (mm). The above change is required

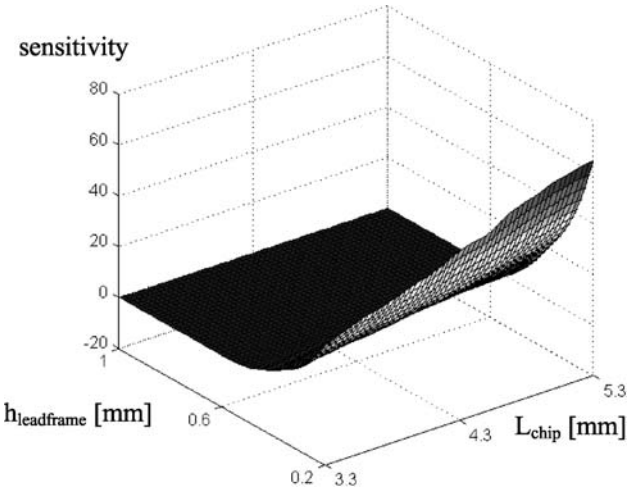


FIGURE 6.56. The sensitivity graph for the combined stress (Gaussian distribution with 1000 Monte Carlo samples).

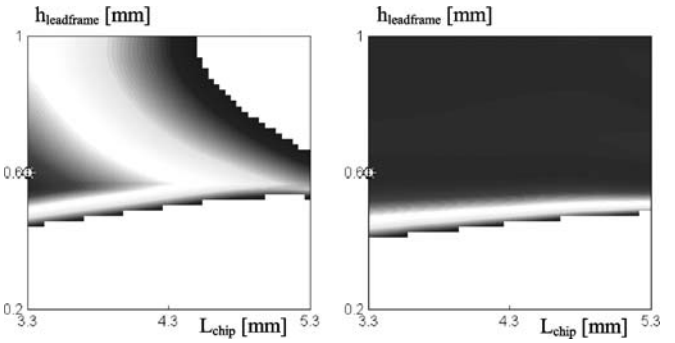


FIGURE 6.57. The optimal values search for the combined stress and sensitivity analysis with 10% tolerances.

when assuming that the input factors/parameters have the Gaussian distribution with the defined nominal value (Figure 6.57).

6.5. CONCLUSION AND CHALLENGES

This chapter highlights our major research and development results and the state-of-the-art methodology of virtual prototyping of microelectronics. Focus is on the method of virtual thermo-mechanical (thermal, mechanical and thermo-mechanical) prototyping. The results of virtual thermo-mechanical prototyping can be used to predict, qualify and optimize the thermo-mechanical behavior and/or trends of microelectronics against the actual requirements prior to major physical prototyping, manufacturing investments and reliability qualification tests.

One should also notice, that traditional experiments and tests will continue to play an important role in the content of virtual prototyping. First, they are needed in providing inputs for modeling, such as characterizing material and their interface behavior (mater-

ial properties, damage initiation, evolution and failure criteria). Secondly, very often, the correctness and accuracy of the developed simulation models and virtual prototyping results need to be verified via experiments for the whole range of the design spaces, and by covering all the critical processes.

There are mainly three important success factors for virtual thermo-mechanical prototyping. The first is to develop accurate and efficient simulation models to predict the thermo-mechanical behavior and/or trends of microelectronics. Accuracy is essential to capture the response of microelectronics correctly under manufacturing, testing and usage conditions. It can be either qualitative or quantitative, depending on the application requirements. Efficiency is needed due to the fact that it is expensive (if it is not impossible) to predict the complicated responses of microelectronics covering the whole design space and whole life cycle, and to conduct global design optimization, such as finding the maximum/minimum, robust designing, parameter sensitivity. In order to do that, one should pay attention to the following aspects:

Product/process inputs
Without reliable product/process inputs, the simulation models cannot be reliable. Two types of product/process inputs are required. One is the design parameter, another is the deviation parameter. The feasible design parameters and the associated design spaces are the starting point for modeling. The process design parameters, for example, determine the actual loading, the boundary conditions, damage initiation and evolution, partially the geometry and the material properties. The deviation parameters are fixed design parameters without given design space. However, they do have inherited scatters compared with the desired normal design values. Knowing the probabilistic distribution of these deviation parameters is important for designing for product/process robustness. Acquisition of reliable design and deviation parameters is not a trivial task, due to the facts that the real inputs, especially the deviation parameters, have strong probabilistic character and it is difficult, time and money consuming to extract the real data via in situ measurements and observations.

Tests and experiments
There are mainly two types of tests used for microelectronics, namely, the functionality test and the reliability qualification test. For functionality test, due to increased design complexity and technology and function integration, new test strategy and methods are needed to achieve maximum test coverage with minimum costs. For reliability qualification test, the essential is the correlation between the accelerated reliability tests with reliability qualification specifications, and between the real application conditions with reliability qualification specifications. The vital issue here is to achieve qualitative and quantitative matching for failure mechanisms and failure criteria.

Three types of experiments are widely used in supporting thermo-mechanical simulation, namely, experiments for material and interface characterization; for damage and failure criteria extraction; and for simulation model verification. Developing experimental methods and tools with sufficiently accurate resolution, correlating the experimental conditions with the real loading history and constraints, and designing samples representing the real product configurations, are the obvious difficulties in microelectronics.

Multiscale mechanics
The major challenge for the fundamental mechanics knowledge (both theoretic and experimental) in microelectronics is the multiscale nature of microelectronics in both geometric (from nano to millimeters) and time (from nanosecond to years) domains.

First, much effort should be spent on developing non-continuum (or enriched continuum) mechanics capable of simulating the behavior of microelectronics with micron, deep-submicron and nano dimensions, in order to capture the strong size effects inherited in microelectronics. These size effects are often related with microstructures and their evolution, various gradient effects (chemical, electrical, thermal and mechanical) and surface effect. For example, metals with a grain size of around 10 nanometers can be as much as seven times harder and tougher than their ordinary counterparts with grain sizes in the micrometer range. Tremendous size effects like this are known to play a significant role in miniaturization, and the implementation of these effects in future design processes is a necessary prerequisite to make optimal use of materials and structures at the nano scale. Presently the required knowledge in this area is substantially under-developed. The product/process behavior at nano scales cannot be predicted by simply applying the conventional macro-scale based approaches, such as continuum mechanics and thermal management, because they do not include any peculiarities of the small-scale structure of materials, but merely represent an average behavior. For this reason, they are not directly applicable for current and future product/process.

Secondly, a bridge should be developed to close the gap between non-continuum (or enriched continuum) theories, simulation tools and results of atomistic scale with the continuum theories, simulation tools and results of macro-scale. So that it will be possible to conduct multi-scale modeling, such as integrated process modeling starting from wafer processing, packaging to systems levels.

Advanced simulation tools

FEM is a well-established technique for predicting thermo-mechanical behavior of product/process. It has made significant progress especially during the last 20 years due to the rapid development of computer hardware and software. However, the commercially available FEM tools are not specifically developed for applications and needs of microelectronics. To make the FEM tools suitable for applications in microelectronics, the following issues deserve special attentions:

- Developing efficient and robust algorithms and solvers. In many applications, in order to obtain accurate results, the complicated geometric effects with high aspect ratio, nonlinear, time and temperature dependent material behavior and the complicated process history should all be considered. Despite the rapid development of computer hardware, days are still needed for the results of a single run. Therefore, virtual prototyping and qualification cannot be efficiently conducted without more efficient and robust algorithms and solvers.
- Developing efficient and reliable stochastic simulation methods. Since the design and response parameters are all probabilistic in nature, deterministic modeling and results alone cannot lead to the optimal thermo-mechanical solutions for business. Very often, robust designs are targeted, and failure probability is required. Therefore, efficient and reliable stochastic simulation methods and tools should be further developed.

Multi-physics experiments and modeling

Multi-physics modeling is another challenge of microelectronics. Microelectronics is strongly multi-discipline and multi-process. Most of the time, it is not possible to predict the behavior of microelectronics correctly by covering only one single discipline and single process. The design and qualification of microelectronics should base on integrated

understanding and solutions covering all the major involved disciplines (electric, mechanics, physics, chemistry, metallurgy, etc.) and processes (IC, packaging, assembly, testing, etc.). Taking CMP process as one example, both mechanical and chemical simulations are needed. Many important failure modes and mechanisms, such as humidity and moisture related failures, electrical overstress, stress corrosion, various cracking and fractures, MEMS striction, void formation, combined diffusion, undesired intermetallic growth, material aging, electro/thermal/stress/chemical migrations, etc. are results of strong multidisciplinary interactions. From process modeling aspect, it is well known that stress/strain induced from IC process will have impact on packaging, and packaging processes might have significant impact on the design and reliability of board level assembly as well. As the ever-increasing application of SiP, wherein different technologies (such as IC, packaging and assembly technologies), different functionalities (such as electrical, optimal, mechanical, etc.), different scales (from deep-submicron to mm) and different discipline (electric, mechanical, optimal, chemical, etc.) are strongly interact with each other, multi-physics simulation and experimental capabilities is essential part of the virtual prototyping.

Material and interface behavior

Reliable models to describe the process dependent behavior (properties, damage initiation, evolution and failure criteria) of materials and their interfaces are essential for not only predictive modeling, but also material development, pre-selection and process optimization. The ultimate aim for characterization and modeling of material and interface behavior is to develop chemical/metallurgy/physics based material design rules to tailor and manipulate material properties according to specific application needs. Several issues requires special research attention:

- For the "macro-scale" application in microelectronics, damage models and failure criteria are vital for reliable failure predictions. Knowing stress/strain distributions alone is not sufficient. The materials used for microelectronics are usually size and processes (time, stress, temperature, constraint, etc.) dependent. However, it is not easy to obtain quantitatively reliable damage models describing the size and process dependent damage initiation, evolution and failures, using the available theories and experimental techniques. Yet another difficulty is the characterization and modeling multi-damage problems, where different failure modes occur simultaneously or consequentially.
- Presently, the characterization and modeling of material behavior are usually based on the partitioning of constitutive models that describe the material properties with damage models that describe the damage and failures of materials. This practice cannot meet the need of microelectronics with nano-scale and strong multi-disciplinary interaction. Strictly speaking, material properties are always linked with damage initiation, evolution and failures. In the scale of microelectronics, the conventional constitutive models should be integrated with damage models, to form a law that describes and governs the total behavior of materials. From computational point of view, the macroscopic equations of physics of failures and the kinetic equations of the microstructural transformation (including micro-damages) should be solved simultaneously. From experimental point of view, practical identification techniques for evolution of microstructures and failure should be further developed.
- Interface strengths and interfacial failures are probably the most prevalent and pervasive issues in the electronic industry. In particular, as more organic materials being used in various types of microelectronic and Microsystems, and the ongoing trends

of miniaturization towards nano-scale, the interfaces between polymer and metal and between polymer and other adjacent materials are becoming more critical. From Integrated Circuits (ICs) to their packages and to the electronic systems made of these packages, numerous material interfaces exist. The quality, robustness, and reliability of these electronic devices depend, to a large extent, on the adhesion and durability of these interfaces. Debonding or delamination of these interfaces often results in the malfunction or failures of these electronic devices. Although interfacial adhesion had been studied for decades by numerous researchers, a few studies dealt with the adhesion issue from a multi-disciplinary viewpoint. The vast majority of studies focused on either the chemical, or the physical or the mechanical aspect alone. Because of such compartmentalized approaches, no effective methodologies, models and tools are available currently for the prediction of interfacial strength in microelectronics and Microsystems, and industry is still heavily depending on trial-error method for determining the interfacial strength. This situation is becoming even more critical due to mainly the ongoing trends of miniaturization towards nano-scale, which adds the size effect as an extra dimension to the existing scientific challenges. Thus, it is important that a generic framework for prediction of interface strengths incorporating the combined effects of and interactions among physical, chemical and mechanical bonding be developed.

The second important success factor for virtual prototyping is to develop advanced simulation based optimization method. Despite many progresses, the currently available simulation-based optimization methods are still not always reliable and rather expensive to deal with design optimization with requirements of

- Strong nonlinear responses.
- Multi-objective targets.
- Multi-level constraints.
- Large numbers of design parameters.
- Combination of continuous with discrete design parameters.

This chapter presents some development results of smart DOE algorithms in conjunction with advanced RSM methods and software. The focus is to obtain more accurate RSM with less DOE. However, many questions, such as, infill sampling criteria for multi-objective optimization, efficiency and accuracy of cross validation scheme, multiple constraints, convergence properties, Gaussian distribution assumption, etc. should be further investigated.

The third important success factor for virtual prototyping is the way to integrate the simulation models with optimization method, wherein accuracy correlation is essential. There are two types of accuracy specifications for outputs. First is the accuracy of the developed simulation models, second is the accuracy of the developed RMS. Beside that, due to the fact that all the design parameters, in principal, are statistic in nature, the accuracy of these design parameters will have important impact on the accuracy of all the outputs. Successful implementation of virtual prototyping needs reliable correlation method to interlink the accuracy specifications between the modeling results and RSM results, with predefined error criteria. Research is ongoing to make hybrid- description of the correlations of different types of errors, i.e., both mathematically and experience-based, and to make hybrid description of the error specification and control for different types of errors in both FEM models and RSM. That is to say, either from given modeling error to the resulting RSM accuracy specification, or from given RSM error to the modeling accuracy specification.

6.6. LIST OF ACRONYMS

ANOVA	Analysis of Variance
CCD	Central Composite Design
CLT	Central Limit Theorem
CTE	Coefficient of Thermal Expansion
DNP	Distance from Neutral Point
DOE	Design of Experiments
FDM	Finite Difference Method
FEM	Finite Element Method
FFE	Fractional Factorial Experiment
GIGO	Garbage In Garbage Out
HCF	High Cycle Fatigue
IC	Integrated Circuits
LCF	Low Cycle Fatigue
LEFM	Linear Elastic Fracture Mechanics
LH	Latin Hypercube
LHD	Latin Hypercube Design
LHS	Latin Hypercube Sampling
MEMS	Micro Electro-Mechanical Systems
MLH	Modified Latin Hypercube
OA	Orthogonal Arrays
PCB	Printed Circuit Board
PCP	Product Creation Process
PDE	Partial Differential Equation
QFD	Quality Function Deployment
QFN	Quad Flat Non-Lead
RBF	Radial Basis Function
RSA	Response Surface Analysis
RSM	Response Surface Model
SRSM	Sequential Response Surface Modeling
VP	Virtual Prototyping

ACKNOWLEDGMENTS

The work presented in this book chapter is part of the project results of MEVIPRO, financed by the EC in the 5th European Research Program.

REFERENCES

1. G.Q. Zhang, A. Tay, and L.J. Ernst, Virtual thermo-mechanical prototyping of electronic packaging—Bottlenecks and solutions of damaging modeling, 3rd Electronic Packaging Technology Conference (EPTC), Singapore, 2000.
2. A.R. Conn and Ph.L. Toint, An algorithm using quadratic interpolation for unconstrained derivative free optimization, in G. di Pillo and F. Giannes, Eds., Nonlinear Optimization and Applications, Plenum Publishing, 1996, pp. 27–47.
3. M.J.D. Powell, A direct search optimization method that models the objective and constraint functions by linear interpolation, Presentation at the SIAM Conference, Virginia, 1996.

4. V.V. Toropov, Multipoint approximation method in optimization problems with expensive function values, in A. Sydow, Ed., Computational System Analysis, Elsevier, 1992, pp. 207–212.
5. D. Belavic, K.P. Friedel, A. Wymyslowski, and M. Santo-Zarnik, Virtual prototyping of the ceramic pressure sensor, Proceedings of the 3nd International Conference on Benefiting from Thermal and Mechanical Simulation in (Micro)- Electronics, Paris, 2002, pp. 38–44.
6. G.Q. Zhang, J. Bisschop, and P. Maessen, Virtual thermo-mechanical prototyping of microelectronics products—towards optimised designing in reliability, Advancing of Microelectronics, 28(1) (2001).
7. V.V. Toropov, A.A. Filatov, and A.A. Polynkine, Multiparameter structural optimization using FEM and multipoint explicit approximations, Structural Optimization, 6, pp. 7–14 (1993).
8. V. Adams and A. Askenazi, Building Better Products with Finite Element Analysis, OnWord Press, 1999.
9. R.H. Myers, Response surface methodology—current status and future directions, Journal of Quality Technology, 31(1), pp. 30–44 (1999).
10. P. Stehouwer and D. Hertog, Simulation-based design optimisation: methodology and applications, Engineering Design Optimization Proceedings of 1st ASMO UK, 1999.
11. I.M. Edwards and A. Jutan, Optimization and control using response surface methods, Computers Chem. Engng, 21(4), pp. 441–453 (1997).
12. G.G. Wang and Z. Dong, Design optimization of a complex mechanical system using adaptive response surface method, Transaction of the CSME, 24(1B), pp. 295–306 (2000).
13. I.P. Schagen, Sequential exploration of unknown multi-dimensional functions as an aid to optimization, IMA Journal of Numerical Analysis, 4, pp. 337–347 (1984).
14. V.V. Toropov, A.A. Filatov, and A.A. Polynkin, Multiparameter structural optimization using FEM and multipoint explicit approximations, Structural Optimization, 6, pp. 7–14 (1993).
15. R.H. Myers, A.I. Khuri, and W.H. Carter, Response surface methodology: 1996–1988, Technometrics, 31(2), pp. 137–157 (1989).
16. T.W. Simpson, J. Peplinski, P.N. Koch, and J.K. Allen, On the use of statistics in design and the implications for deterministic computer experiments, Proceedings of DETC'97, 1997 ASME Design Engineering Technical Conferences, Sacramento, California, 1997.
17. M.C. Li and J. Chen, Determining process mean for machining while unbalanced tolerance design occurs, Journal of Industrial Technology, 17(1), pp. 2–6 (2001).
18. A.E. Raferty and D. Madigan, Bayesian model for linear regression models, Journal of the American Statistical Association, pp. 179–191 (1998).
19. M.D. Profirescu, G. Dima, B. Govoreanu, and O. Mitrea, HEMT Optimization by Advanced RSM Models, University Politehnica of Bucharest, Romania.
20. D.M. Ghiocel, Stochastic field models for approximating highly nonlinear random responses, Proceedings of the 15th ASCE Engineering Mechanics, Columbia University, New York, 2002.
21. L. Trocine and L.C. Malone, Finding important independent variables through screening designs: a comparison of methods, Proceedings of 2000 Winter Simulation Conference, 2000, pp. 749–754.
22. A. Todoroki, Teach Yourself response Surface Methodology, Tokyo Institute of Technology, Tokyo, Japan.
23. A. Wu, K.Y. Wu, R.M.M. Chen, and Y. Shen, Parallel optimal statistical design method with response surface modelling using genetic algorithms, IEE Proc. Circuit Devices Syst., 145(1), pp. 7–12 (1998).
24. T.H. Smith, B.E. Goodlin, and D.S. Boning, A statistical analysis of single and multiple response surface modeling, IEEE Transactions on Semiconductor Manufacturing, 12(4), pp. 419–430 (1999).
25. A.G. Greenwood, L.P. Rees, and F.C. Siochi, An investigation of the behavior of simulation response surfaces, European Journal of Operational Research, 110, pp. 282–313 (1998).
26. J. Sacks, S.B. Schiller, and W.J. Welch, Design for computer experiments, Technometrics, 31(1), pp. 41–47 (1989).
27. P.C. Benjamin, M. Erraguntla, and R.J. Mayer, Using simulation for robust system design, Simulation, 65(2), pp. 116–128 (1995).
28. T.J. Green and R.G. Launsby, Using DOE to reduce costs and improve the quality of microelectronic manufacturing processes, The International Journal of Microcircuits and Electronic Packaging, 18(3), pp. 290–296 (1995).
29. R. Khattree, Robust parameter design: a response surface approach, Journal of Quality Technology, 28(2), pp. 187–198 (1996).
30. P.J. Ross, Taguchi Techniques for Quality Engineering, McGraw-Hill Book Company, 1988.
31. W.W. Hines and D.C. Montgomery, Probability and Statistics in Engineering and Management Science, John Wiley & Sons.
32. M.R. Beauregard, R.J. Mikulak, and B.A. Olson, Experimenting for Breakthrough Improvement, Resource Engineering, 1989.

33. D.R. Jones, M. Schonlau, and W.J. Welch, Efficient global optimization of expensive black-box functions, Journal of Global Optimization, 13, pp. 455–492 (1998).
34. G. Taguchi, System of Experimental Design: Engineering Methods to Optimize Quality and Minimise Costs, UNIPUB/Kraus International Publications, 1987.
35. J.R. Koshel, Enhancement of the downhill simplex method of optimisation, International Optical Design Conference, 2002.
36. R.O. Bowden and J.D. Hall, Simulation optimisation research and development, Simulation Conference Proceedings, Vol. 2, 1998, pp. 693–1698.
37. M.J.D. Powell, A direct search optimisation method that models the objective and constraint functions by linear interpolation, Proceedings SIAM Conference, 1996, pp. xx-yy.
38. W.D. Van Driel, G.Q. Zhang, J.H.J. Janssen, and L.J. Ernst, Response surface modelling for non-linear packaging stresses, Journal of Electronic Packaging, 125(4), pp. 490–497 (2003).
39. G.Q. Zhang, The challenges of virtual prototyping and qualification for future microelectronics, J. Microelectronics Reliability, 43, pp. 1777–1785 (2003).
40. A.R. Conn and Ph.L. Toint, An algorithm using quadratic interpolation for unconstrained derivative free optimisation, in G. di Pillo and F. Giannes, Eds., Nonlinear Optimization and Applications, Plenum Publishing, 1996, pp. 27–47.
41. V.V. Toropov, Multipoint approximation method in optimisation problems with expensive function values, in A. Sydow, Ed., Computational System Analysis, Elsevier, 1992, pp. 207–212.
42. V.V. Toropov, A.A. Filatov, and A.A. Polynkine, Multiparameter structural optimisation using FEM and multipoint explicit approximations, Structural Optimisation, 6, pp. 7–14 (1993).
43. D.C. Montgomery, Design and Analysis of Experiments, John Wiley & Sons Inc., 2005.
44. W.D. Van Driel, J. Van de Peer, N. Tzannetakis, A. Wymyslowski, and G.Q. Zhang, Advanced numerical prototyping methods in modern engineering applications, Proceedings of the 5th International EuroSimE Conference, 2004, pp. 211–218.
45. A. Wymyslowski, W.D. Van Driel, G.Q. Zhang, J. Van de Peer, and N. Tzannetakis, Smart and sequential approach to numerical prototyping in micro-electronic application, JMEP, 2(1), pp. 1–7 (2005).
46. R. Van den Boomen and M.C. Seegers, Leadframe materials, Internal Philips Report, 1996.
47. J.D. Wu, C.Y. Huang, and C.C. Liao, Fracture strength characterization and failure analysis of silicon dies, Microelectronics Reliability, 43, pp. 269–277 (2003).
48. O.C. Zienkiewicz and R.L. Taylor, The Finite Element Method, Volumes 1–3, Butterworth-Heinemann, London, 2000.
49. D.C. Montgomery, Design and Analysis of Experiments, John Wiley & Sons Inc., 2005.
50. R.H. Myers and D.C. Montgomery, Response Surface Methodology: Process and Product Optimisation Using Designed Experiments, John Wiley & Sons Inc.
51. M. Sasena, M. Parkinson, P. Goovaerts, P. Papalambros, and M. Reed, Adaptive experimental design applied to an ergonomics testing procedure, Proceedings of DETC'02, Montreal, Canada, October 2002.
52. D.R. Jones, M. Schonlau, and W.J. Welch, Efficient Global Optimization of Expensive Black-Box Functions, Kluwer Academic Publishers, Netherlands, 1998.
53. J. Sacks, W.J. Welch, W.J. Mitchell, and H.P. Wynn, Design and analysis of computer experiments, Statistical Science, 4(4), pp. 409–435 (1989).
54. O.F. Slattery, G. Kelly, and J. Greer, Benefiting from thermal and mechanical simulation in microelectronics, in G.Q. Zhang, L.J. Ernst, and O. de Saint Leger, Eds., Thermal and Mechanical Problems in Microelectronics, Kluwer Academic Publishers, 2000, pp. 17–26.

MATERIALS MECHANICS

7

Fiber Optics Structural Mechanics and Nano-Technology Based New Generation of Fiber Coatings: Review and Extension

E. Suhir

University of California, Santa Cruz, CA, University of Maryland, College Park, MD, and ERS/Siloptix Co., Los Altos, CA, USA

7.1. INTRODUCTION

This chapter consists of two parts—review and extension.

The review part deals with typical fiber optics structures (bare, single- and dual-coated fibers; fibers experiencing low temperature micro-bending; fibers soldered into ferrules or adhesively bonded into capillaries; role of the non-linear stress–strain relationship, etc.) subjected to thermally induced and/or mechanical loading in bending, tension, compression, or to various combinations of such loadings. The emphasis is on the state-of-the-art in the area of optical fiber coatings and the functional (optical), mechanical and environmental problems that occur in polymer-coated or metalized fibers. The solutions to the examined problems are mostly obtained using analytical methods (predictive models) of structural mechanics. The review is based primarily on the author's research conducted at Bell Laboratories, Murray Hill, NJ, during his eighteen years tenure with this company.

The extension part addresses a new generation of optical fiber coatings and deals with the application of a newly developed (by the ERS/Siloptix Co.) nano-particle material (NPM) that is used as an attractive substitute for the existing optical fiber coatings. This NPM-based coating has all the merits of polymer and metal coatings, but is free of many of their shortcomings. The developed material is an unconventional inhomogeneous "smart" composite material, which is equivalent to a homogeneous material with the following major properties: low Young's modulus, immunity to corrosion, good-to-excellent adhesion to adjacent material(s), non-volatile, stable properties at temperature extremes (from $-220°C$ to $+350°C$), very long (practically infinite) lifetime, "active" hydrophobicity—the material provides a moisture barrier (to both water and water vapor), and, if necessary, can even "wick" moisture away from the contact surface; ability for "self-healing" and "healing:" the NPM is able to restore its own dimensions, when damaged, and is able to fill existing or developed defects (cracks and other "imperfections") in contacted surfaces; very low

(near unity) effective refractive index (if needed). NPM can be designed, depending on the application, to enhance those properties most important. NPM properties have been confirmed through testing. The tests have demonstrated the outstanding mechanical reliability, extraordinary environmental durability and, in particular applications, improved optical performance of the light guide.

7.2. FIBER OPTICS STRUCTURAL MECHANICS

Fiber-Optics Structural Mechanics (FOSM) deals with the application of methods and approaches of Structural Engineering (Structural Analysis), as well as of Engineering and Applied Mechanics, to the stress/strain evaluations, and physical design for reliability of fiber-optics structures and systems. FOSM considers the specifics, associated with the properties of the materials used, typical structures employed, as well as the nature, magnitude and variability of the applied loads. FOSM treats an optical fiber system as a structure. In other words, it examines hardware systems, in which the materials' interaction, their size and configuration, and the loads, whether thermally induced or "mechanical," are as important as the characteristics of the employed materials. The main objectives of FOSM have to do with physical design for reliability of fiber optic systems and could be defined as follows:

- determine the loading conditions (which could be due to the thermal expansion mismatch of materials, lateral and angular misalignments, test loads, dynamic loads due to shocks and vibrations, etc.),
- evaluate stresses, strains, and fracture characteristics of the photonics structure, and
- ensure that the chosen strength and reliability criteria will remain, during the lifetime of the structure, within the limits acceptable from the standpoint of structural integrity, elastic stability, dependability, and normal operation, both mechanical (structural) and functional (optical), of the system.

The application of the methods and approaches of FOSM can be very helpful in creating a viable and reliable fiber optics products and networks. In this review we examine a number of practically important problems of the mechanical behavior and structural ("physical") design of bare or coated optical fibers, experiencing thermal, mechanical or dynamic loading. The following major topics are addressed: (1) bending of bare silica fibers, (2) fibers under the combined action of bending and tension, (3) large deflections of fibers, (4) effect of material's nonlinearity, (5) mechanical behavior of polymer coated or metalized fibers, (6) elastic stability and low temperature microbending of optical fibers, (7) solder materials and joints employed in fiber optics, and (8) dynamic response of optical fibers to shocks and vibrations.

7.2.1. Review

The state-of-the-art in the application of methods of materials, mechanical and reliability engineering to photonics structures, including optical fibers, with an emphasis on analytical modeling, design for reliability, and application of probabilistic methods, can be found in the [1–15]. A brief review of the major directions in Fiber Optics Structural Mechanics is given in [4]. Thermal loading is responsible for many failures in photonics engineering. The state of the art in this area is outlined in [5,9,11]. Methods and approaches

in evaluating structural reliability of fiber optics systems, including stress/strain analyses and accelerated life testing were addressed in [3,12,14].

Bending of bare fibers, idealized as a single span beams clamped at the ends and subjected to both lateral and/or angular misalignment, was examined, based on the elementary beam theory (see, for instance, [1]), primarily in application to the mechanical behavior of optical fiber interconnects [16–31]. In optical fiber interconnects, angular misalignments and ends offsets are typically due to the inability of the given technology to ensure good alignment of the interconnect ends and/or end cross-sections. In other cases, misalignments are essential, and quite often even desirable, features of a particular package design. In many cases, elevated optical fiber curvatures, caused by various misalignments, can be responsible for both its functional (optical) performance and mechanical (structural) reliability. These curvatures and the resulting bending stresses can be predicted [17–19] and, if necessary, successfully minimized [18,21–24] for lower curvatures. These are responsible for both added transmission losses and mechanical reliability.

Three- or four-point bending is often used to experimentally evaluate the Young modulus of the silica material and its ultimate flexural strength. If such testing is conducted, it is important to make sure that the test specimen is long enough so that the effect of shear would not have to be considered [16].

Elevated lateral gradients of the coefficients of thermal expansion and Young's moduli (in direction of the fiber diameter) can be responsible for the fiber "curling" during drawing of optical silica fibers [19]. The analysis, reported in [19], was carried out on the basis of both analytical ("mathematical") and numerical (finite element) modeling.

Bare fibers under the combined action of bending and tension were examined in connection with proof testing of optical fibers soldered (epoxy bonded) into ferrules (capillaries) [25]. It has been shown that the fiber under testing should be long enough so that the inevitable end misalignment would not result in appreciable bending stresses. The effect of these stresses should be accounted for, if the test specimen cannot be made sufficiently long and/or if a sufficiently good alignment could be achieved. If the lateral misalignment of an optical fiber interconnect is not very small, and the interconnect ends cannot move closer when it experiences substantial bending deformations, then reactive tension occurs. The resulting tensile stresses can be analyzed on the basis of a linear theory. This could be done, if bending deformations are large enough to produce appreciable reactive tensile stresses, but still small enough not to necessitate the application of a nonlinear approach [27,28]. Tensile stresses are highly undesirable in silica fibers. In combination with surface cracks and moisture, such stresses can lead to a rapid rupture of a silica fiber. On the other hand, elevated compressive stresses can lead to fiber buckling, thereby producing tensile stresses due to bending. In some cases, the thermal contraction mismatch between the silica fiber enclosure and the fiber itself can be effectively used to minimize the tensile stress in an optical fiber interconnect subjected to both end misalignment and thermally induced compression [28,29].

Consideration of the structural ("geometric") and materials ("physical") nonlinearity might be necessary, if the fiber experiences large bending and/or axial deformations [32–44]. The effect of the structural nonlinearity, which is due to the significant bending deformations of optical fibers, and the materials nonlinearity, caused by the nonlinear stress-strain relationship in silica materials subjected to tension, has been analyzed by many investigators (Cowap and Brown [36], France et al. [35], Krause et al. [34], McMullin and Freeman [37], Murgatroyd [32], Sinclair [33], Suhir [38–40], Muraoka [44]). The com-

bined effect of the materials nonlinearity and the nonprismaticity of a fused biconical taper (FBT) optical coupler on the induced thermally induced stresses was analyzed in [41].

Optical fiber "pigtails" often experience three-dimensional bending deformations. Substantial relief in the induced bending stresses and curvatures can be achieved by optimizing the "pigtail's" configuration [42,43].

Coated fibers, whether polymerically coated (see, for instance, Devadoss [48], Gebizioglu and Plitz [47], King and Aloisio [54]) or metalized (see, for instance, papers published in [1]), are widely employed for better short- and long-term reliability of the silica material, which is both brittle and moisture-sensitive. Problems encountered during design, manufacturing, testing, and reliability assessments for dual-coated glass fibers include: evaluation of the effect of coating on the bending stresses [45,46], understanding the delamination mechanisms and improving strippability [46,49,52], prediction of the magnitude and distribution of stresses occurring during proof (pull-out) testing [49–62], etc.

Understanding and optimizing the interaction of "global" and "local" thermally induced stresses in optical glass fibers adhesively bonded or soldered at the ends into capillaries or ferrules is important for the "physical" design of, and reliability assessments for, these structures. This can be done on the basis of a simplified analytical stress model developed for a cylindrical b-material assembly adhesively bonded at the ends [56].

Elastic stability and microbending of optical fibers is important primarily in connection with the added transmission losses associated with these phenomena (Cocchini [71], Ostojic [76], Shiue [66–68,70,72–75,79,80], Suhir [63–65,69,77,78,81,82]). In metalized fibers, however, the high Young modulus of metalization can cause significant strength problems for both the metalization and the fiber [62]. Therefore the application of "hard" metalization should be carried out with caution, unless no appreciable thermal deformations of the interconnect are expected.

It has been noticed [77] that external periodic loading with the period of about 100 nm can cause appreciable microbending losses in dual-coated fibers, and therefore should be avoided in actual designs. It has been noticed also [65] that the threshold of elevated low temperature transmission losses in polymer-coated fibers corresponds to the temperature at which the stresses at the coating/glass interface start to increase rapidly. This circumstance enables one to predict this threshold by stress calculation; instead of much more complicated optical calculations or measurements.

Voids in the solder joints and adhesives in soldered or adhesively bonded fiber optic assemblies cause natural concerns of fiber optics designers. Effect of voids in epoxy-bonded fibers was analyzed in [78] for different void size, configurations, locations, etc.

Solder materials and joints are as important in photonics and, particularly, in fiber optics, as they are in microelectronics. There is, however, a number of specific requirements for the solder materials and joints used in photonics: ability to achieve high alignment, requirement for a very low creep, etc. [84]. Thermally induced stresses in optical fibers soldered into various ferrules were addressed in [83]. It has been shown that low expansion enclosures is not always the right choice from the standpoint of the thermally induced stresses in optical fibers soldered into ferrules, as well as the stresses in the solder material itself.

Dynamic response of fiber optic structures to shocks and vibrations was examined in [85–90]. The ability to predict and possibly minimize the dynamic stresses in fiber-optic systems is of obvious practical importance. It has been shown, particularly, that the

application of the maximum acceleration as the criterion of strength of a structural element in a microelectronic or photonic design can be misleading [87].

7.3. NEW NANO-PARTICLE MATERIAL (NPM) FOR MICRO- AND OPTO-ELECTRONIC APPLICATIONS

7.3.1. New Nano-Particle Material (NPM)

A new advanced nano-particle material (NPM) has been developed by the ERS/ Siloptix Co. [91–95]. This material has many attractive applications in micro- and opto-electronic packaging. Particularly, an effective practical technology for making NPM-based optical silica fiber coatings has been developed under grants from DARPA/Navy. The developed technology enables one to create ultra-thin, highly cost-effective, highly mechanically reliable, and highly environmentally durable coatings for silica light-guides. The obtained results have demonstrated the performance superiority of the developed technology over polymer-coated and metalized fibers, as well as a potential that the NPM has for various commercial and military applications in micro- and opto-electronics packaging and related areas. It can have many attractive applications also well beyond the "high-tech" field. NPM is an unconventional inhomogeneous "smart" composite material. It is equivalent to a "hypothetical" homogeneous material with the following major properties:

- Low Young's modulus
- Immunity to corrosion
- Good adhesion to the adjacent material
- Non-volatile
- Stable properties at temperature extremes (from $+350°C$ to as low as $-220°C$)
- Very long (practically infinite) lifetime
- Strong hydrophobicity (against both water and water vapor)
- Ability for "self-healing:" ability to restore its dimensions and initial structure when damaged
- Ability for "healing" the surfaces of the adjacent materials, i.e., to fill in the existing and/or the developed defects (surface cracks, flaws, and other imperfections) and to slow down their propagation and/or even to "arrest" them completely
- Very low effective refractive index (if needed)
- High dielectric constant (if needed).

The NPM can be designed, depending on the particular application, in such a way that its most important particular properties are enhanced. The conducted tests have confirmed these properties. In general, it is desirable to provide application-specific modifications of the NPM to master/optimize its properties and performance. Because it is a nano-material, its surface chemistry and its performance depend a lot upon the contact materials and surfaces. With this in mind, the following applications are viewed as the most attractive ones.

- NPM is able to hermetically seal packages, components and devices, such as laser packages, MEMS, displays and plastic LEDs.
- NPM can be used as an effective protective coating for various metal and non-metal surfaces, well beyond the area of micro- and opto-electronic packaging: in cars, aerospace structures, offshore and ocean structures, marine vehicles, civil engineering structures (bridges, towers, etc.), tubes, pipes and pipe-lines, etc. These

applications benefit because the material is actively hydrophobic, does not induce additional stresses (owing to its low modulus), is inexpensive, is easy-to-apply, has practically infinite lifetime, and is self-healing. Application of this material can result in a significant resistance of a metal surface to corrosion, and, in addition, in substantial increase in the fracture toughness of the material, both initially and during the system's operation (use).

- The NPM can be added in the formulation of various coatings such as paints, thereby providing protective benefits without changing the application techniques.
- Because of a low refractive index, the NPM can be used, if necessary, as an effective cladding of optical silica fibers. The use of the NPM cladding eliminates the need to dope silica for obtaining light-guide cores. The new perform will consist of a single (undoped and, hence, less expensive) silica material.
- A derivative application is flexible light-guides. Multicore flexible fiber cables employing NPM are able to provide high spatial image resolution. As such, they might find important applications, when there is a need to provide direct high-resolution image transmission from secluded areas. Possible applications can be found in biomedicine, nondestructive evaluations, oil and other geological explorations, in ocean engineering, or in other situations when an image needs to be obtained and transmitted from relatively inaccessible locations. In such applications, the plane ("butt") end of the fiber bundle (cable) will play the role of a small size pixel array. The transmitted image can be concurrently or subsequently enlarged to a desirable size, as needed.
- Another derivative application is a multicore fiber cable. Ultra-small diameter glass fibers with an NPM-based cladding/coating can be placed in large quantities within a NPM medium ("multiple cores in a single cladding"). In addition, owing to a much better inner-outer refractive index ratio in the NPM-based fibers, such cables will be characterized by very low signal attenuation.
- Another derivative application is sensor systems. The NPM-based fibers can be used in optical sensor systems that employ optical fibers embedded in a laminar or a cast material. Such systems are used, for instance, in composite airframes. With the NPM used as a cladding or, at least, as a coating of the silica optical fiber, the optical performance and the mechanical reliability of the light-guide will improve dramatically compared with the conventional systems.
- Ultra-thin planar light-guides are another derivative application of the NPM. In the new generation of the planar light-guides, NPM can be used as the top cladding material. It will replace silicon or polymer claddings, which are considered in today's planar light-guides. All the advantages of the NPM cladding material discussed above for optical fibers are equally applicable to planar light-guides. These are thought to have a "bright" future in the next generation of computers and other photonic devices.

7.3.2. NPM-Based Optical Silica Fibers

A modification of the NPM has been developed and tested as an attractive substitute for the existing hermetic and non-hermetic optical fiber coatings. The following major activities were undertaken and the following results were obtained:

- The drawing (manufacturing) process and the drawing tower were adequately retrofitted to adjust them to the characteristics of the developed NPM and to the NPM layer application procedure
- The conducted mechanical tests have demonstrated remarkable strength (up to 7.5 GPa = 765 kgf/mm sq = 1088 kpsi) and attractive quality (low strength variability) of the manufactured NPM-based fibers. Such high strength characteristics have been never achieved before, even in the lab conditions.
- The environmental tests have shown that even at the humidity level of 100% (samples were immersed into water for 24 hours) the mechanical strength of these fibers is on the order of the strength of the best quality fibers at the "dry" conditions in the previous tests. There is reason to believe that the achieved performance is still not a limit of the NPM-based technology and that the higher fibers strengths and better environmental stability are feasible by further "fine tuning" and further optimization of the NPM and the drawing procedure.
- The optical performance of the NPM-based fibers (in terms of the attenuation level) is almost two-fold better than the optical performance of the reference (existing) samples. The estimated lower limit of the NPM based optical fibers with silica glass core and stepwise refractive index change, can potentially get a record values for the tested type of multi-mode fibers (getting even below 1 dB/km in a specific spectral "window").

The obtained results clearly demonstrated the performance superiority of the developed technology and a great potential (scientific, technological and commercial) of the future products, which makes the project attractive for the commercialization. The commercialization phase of the project will allow one to broaden the scope and the range of the NPM compounds modeling leading to much more "smarter" and better "self-programmable" claddings/coatings, including transparent and other special properties NPM materials for claddings and coatings. Some NPM coated samples are shown in Figures 7.1 and 7.2. Figures 7.3 and 7.4 illustrate the significant advantages, in terms of the

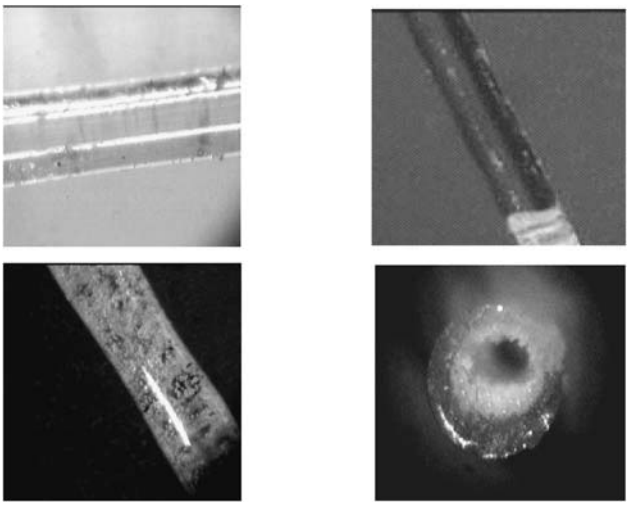
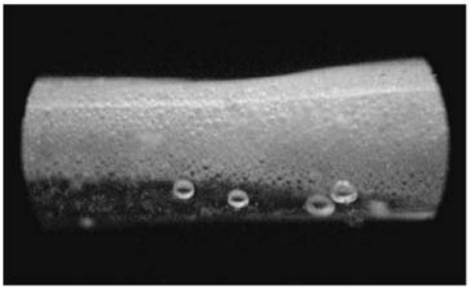


FIGURE 7.1. NPM coated samples.



Advantages of this structure are:

- **Good strippability**
- **Tensile and twisting stresses applied to the fiber are, actually, applied to the coating only**
- **Reduced bending stresses applied to the core/cladding for the same bending radii**

FIGURE 7.2. NPM-based containing structures-1.

Strength Distributions of NPM -Coated (Blue) and Reference (Red) Fibers

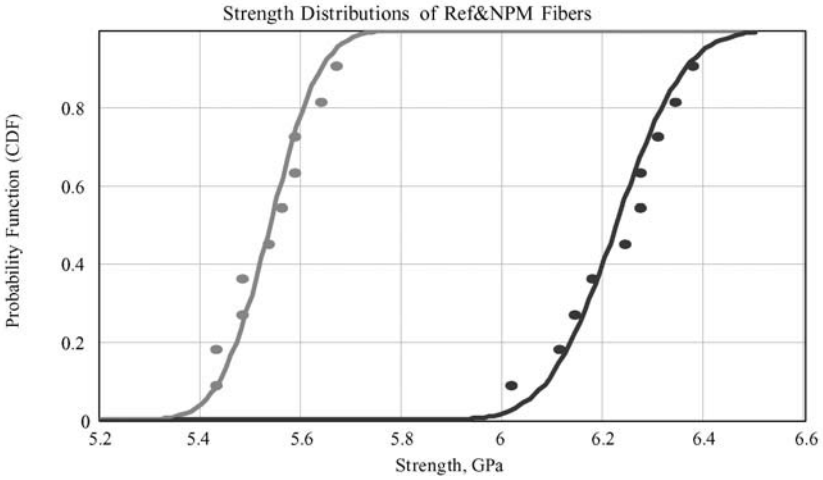
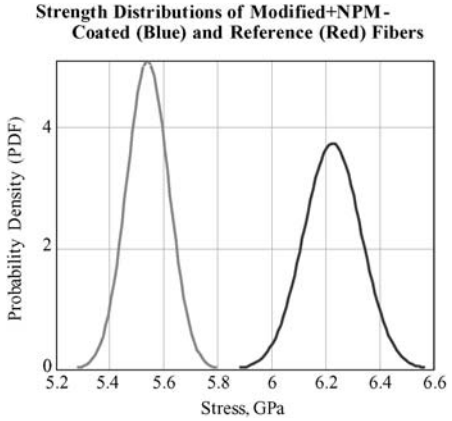


FIGURE 7.3. Samples manufactured using improved technology environmental tests (RH = 100%).



At a high humidity level, the gap between the strengths of the NPM-coated and the reference population is even greater: the weakest NPM-coated sample turned out to be substantially stronger than the strongest reference fiber sample

FIGURE 7.4. Samples manufactured using improved technology.

fiber strength, of the NPM-based coating technology over the conventional, polymer-based, optical fiber coatings.

7.4. CONCLUSIONS

The following conclusions can be drawn from the carried out study:

- The application of the methods and approaches of Fiber Optics Structural Mechanics (Structural Analysis of fiber optics systems) can be very helpful in creating a viable and reliable fiber optics products and networks.
- A viable and effective advanced technology for drawing optical silica fibers with the NPM-based coatings has been developed. This technology has many important advantages over the existing "non-hermetic" (polymer) and "hermetic" (metalization, carbon, etc.) coating materials.
- The conducted mechanical tests have demonstrated remarkable strengths (up to 7.5 GPa) and attractive quality (low local strength variability) of the manufactured NPM-Coated fibers.
- Environmental tests showed that even at the humidity level of 100% (samples were immersed into water for 24 hours) the mechanical strength of these fibers is on the order of the strength of the best quality fibers at dry conditions in the previous tests.
- There is a reason to believe that the achieved performance is still not a limit of the NPM coating technology and that the higher fibers strengths and environmental stability are feasible by further "fine tuning" and optimization of the NPM coating compound content and the drawing procedure.
- The achieved results clearly demonstrated the performance superiority of the developed technology and a great potential (scientific, technological and commercial) of the future products, which makes the project extremely attractive for commercialization.

ACKNOWLEDGMENT

The author acknowledges, with thanks, the support of the project on nano-material by the DARPA and the Nave-Air agencies.

REFERENCES

Fiber optics structural mechanics: design for reliability

1. E. Suhir, Structural Analysis in Microelectronics and Fiber Optics, Van-Nostrand, New York, 1991.
2. E. Suhir, Applied Probability for Engineering and Scientists, McGraw Hill, New York, 1997.
3. E. Suhir, M. Fukuda, and C.R. Kurkjian, Eds., Reliability of photonic materials and structures, Materials Research Society (MRS) Symposia Proceedings, Vol. 531, 1998.
4. E. Suhir, Fiber optics structural mechanics—brief review, Editor's Note, ASME Journal of Electronic Packaging, September 1998.
5. E. Suhir, Thermal stress failures in microelectronics and photonics: prediction and prevention, Future Circuits International, Issue #5, 1999.
6. E. Suhir, Microelectronics and photonics—the future, Microelectronics Journal, 31(11–12) (2000).

7. E. Suhir, The future of microelectronics and photonics, and the role of mechanical, materials and reliability engineering, Proceedings of the International Conference on Materials in Microelectronics, MicroMat 2000, April 17–19, Berlin, Germany, 2000.
8. E. Suhir, Modeling of the mechanical behavior of microelectronic and photonic systems: attributes, merits, shortcomings, and interaction with experiment, Proceedings of the 9-th Int. Congress on Experimental Mechanics, Orlando, FL, June 5–8, 2000.
9. E. Suhir, Thermal stress modeling in microelectronics and photonics packaging, and the application of the probabilistic approach: review and extension, IMAPS International Journal of Microcircuits and Electronic Packaging, 23(2) (2000) (invited paper).
10. E. Suhir, Thermomechanical stress modeling in microelectronics and photonics, Electronic Cooling, 7(4) (2001).
11. E. Suhir, Accelerated life testing (ALT) in microelectronics and photonics: its role, attributes, challenges, pitfalls, and interaction with qualification tests, Keynote address at the SPIE's 7-th Annual International Symposium on Nondestructive Evaluations for Health Monitoring and Diagnostics, 17–21 March, San Diego, CA, 2002.
12. E. Suhir, Analytical stress-strain modeling in photonics engineering: its role, attributes and interaction with the finite-element method, Laser Focus World, May 2002.
13. E. Suhir, How to make a photonic device into a product: role of accelerated life testing, Keynote Address at the International Conference of Business Aspects of Microelectronic Industry, Hong-Kong, January 2003.
14. E. Suhir, Microelectronic and photonic systems: role of structural analysis, InterPack'2005, San Francisco, July 2005.
15. E. Suhir, Analytical thermal stress modeling in physical design for reliability of micro- and opto-electronic systems: role, attributes, challenges, results, Invited Talk, Therminic, 2005, Lago Maggiore, Italy, September 27–30, 2005.

Bending of bare fibers

16. E. Suhir, How long should a beam specimen be in bending tests? ASME Journal of Electronic Packaging, 112(1) (1990).
17. E. Suhir, Analysis and optimization of the input/output fiber configuration in a laser package design, ASME Journal of Electronic Packaging, 117(4) (1995).
18. E. Suhir, Predicted curvature and stresses in an optical fiber interconnect subjected to bending, IEEE/OSA Journal of Lightwave Technology, 14(2) (1996).
19. E. Suhir and J.J. Vuillamin, Jr., Effects of the CTE and Young's modulus lateral gradients on the bowing of an optical fiber: analytical and finite element modeling, Optical Engineering, 39(12) (2000).
20. E. Suhir, Silica optical fiber interconnects: design for reliability, Proceedings of the Annual Conference of the American Ceramic Society, St.-Louis, MO, May 3, 2000.
21. E. Suhir, Optical fiber interconnect with the ends offset and axial loading: what could be done to reduce the tensile stress in the fiber? Journal of Applied Physics, 88(7) (2000).
22. E. Suhir, Method of improving the performance of optical fiber, which is interconnected between two misaligned supports, U.S. Patent #6,314,218, 2001.
23. E. Suhir, Interconnected optical devices having enhanced reliability, U.S. Patent #6,327,411, 2001.

Bare fibers under the combined action of bending and tension

24. E. Suhir, Bending performance of clamped optical fibers: stresses due to the ends off-set, Applied Optics, 28(3) (1989).
25. E. Suhir, Pull testing of a glass fiber soldered into a ferrule: how long should the test specimen be?, Applied Optics, 33(19) (1994).
26. E. Suhir, Optical fiber interconnect subjected to a not-very-small ends off-set, MRS Symp. Proc., Vol. 531, 1998.
27. E. Suhir, Method and apparatus for prooftesting optical fibers, U.S. Patent #6,119,527, 1998.
28. E. Suhir, Optical fiber interconnect with the ends offset and axial loading: what could be done to reduce the tensile stress in the fiber?, Journal of Applied Physics, 88(7) (2000).
29. E. Suhir, Method for determining and optimizing the curvature of a glass fiber for reducing fiber stress, U.S. Patent #6,016,377, 2000.
30. E. Suhir, Apparatus and method for thermostatic compensation of temperature sensitive devices, U.S. Patent #6,337,932, 2002.

31. E. Suhir, Optical fiber interconnects having offset ends with reduced tensile strength and fabrication method, U.S. Patent #6,606,434, 2003.

Consideration of the structural and materials nonlinearity

32. J. Murgatroyd, The strength of glass fibers, Journal of the Society of Glass Technology, 28 (1944).
33. D. Sinclair, A bending method for measurement of the tensile strength and Young's modulus of glass fiber, Journal of Applied Physics, 21 (1950).
34. J.T. Krause, L.R. Testardi, and R.N. Thurston, Deviations FROM linearity in the dependence of elongation upon force for fibers of simple glass formers and of glass optical lightguides, Physics and Chemistry of Glasses, 20 (1979).
35. P.W. France, M.J. Paradine, M.H. Reeve, and G.R. Newns, Liquid nitrogen strength of coated optical glass fibers, Journal of Materials Science, 15 (1980).
36. S.F. Cowap and S.D. Brown, Static fatigue testing of a hermetically sealed optical fiber, American Ceramic Society Bulletin, 63(3) (1984).
37. J.N. McMullin and J.E. Freeman, On the shape of a bent fiber, IEEE/OSA Journal of Lightwave Technology, 8(7) (1990).
38. E. Suhir, Predicted bending stresses in an optical fiber interconnect experiencing significant ends off-set, MRS Symp. Proc., Vol. 531, 1998.
39. E. Suhir, Elastic stability, free vibrations, and bending of optical glass fibers: the effect of the nonlinear stress-strain relationship, Applied Optics, 31(24) (1992).
40. E. Suhir, The effect of the nonlinear behavior of the material on two-point bending in optical glass fibers, IEEE/OSA Journal of Electronic Packaging, 114(2) (1992).
41. E. Suhir, Predicted stresses and strains in fused biconical taper couplers subjected to tension, Applied Optics, 32(18) (1993).
42. E. Suhir, Optimized configuration of an optical fiber "Pigtail" bent on a cylindrical surface, in T. Winkler and A. Schubert, Eds., Materials mechanics, fracture mechanics, micromechanics, an Anniversary Volume in Honor of B. Michel's 50-th Birthday, Fraunhofer IZM, Berlin, 1999.
43. E. Suhir, Method for determining and optimizing the curvature of a glass fiber for reduced fiber stress, U.S. Patent #6,016,377, 2000.
44. M. Muraoka, The maximum stress in optical glass fibers under two-point bending, ASME Journal of Electronic Packaging, 123(March) (2001).

Coated fibers

45. E. Suhir, Stresses in dual-coated optical fibers, ASME Journal of Applied Mechanics, 55(10) (1988).
46. E. Suhir, Stresses in a coated glass fiber stretched on a capstan, Applied Optics, 29(18) (1990).
47. O.S. Gebizioglu and I.M. Plitz, Self-stripping of optical fiber coatings in hydrocarbon liquids and cable filling compounds, Optical Engineering, 30(6) (1991).
48. E. Devadoss, Polymers for optical fiber communication systems, Journal of Scientific and Industrial Research, 51(4) (1992).
49. E. Suhir, Can the curvature of an optical glass fiber be different from the curvature of its coating? International Journal of Solids and Structures, 30(17) (1993).
50. E. Suhir, Buffering effect of fiber coating and its influence on the proof-test load in optical fibers, Applied Optics, 32(7) (1993).
51. E. Suhir, Analytical modeling of the interfacial shearing stress during pull-out testing of dual-coated lightguide specimens, Applied Optics, 32(7) (1993).
52. E. Suhir, Analytical modeling of the interfacial shearing stress in dial-coated optical fiber specimens subjected to tension, Applied Optics, 32(16) (1993).
53. E. Suhir, Approximate evaluation of the interfacial shearing stress in circular double lap shear joints, with application to dual-coated optical fibers, International Journal of Solids and Structures, 31(23) (1994).
54. W.W. King and C.J. Aloisio, Thermomechanical mechanism for delamination of polymer coatings from optical fibers, ASME Journal of Electronic Packaging, 119(2) (1997).
55. E. Suhir, Bending of a partially coated optical fiber subjected to the ends off-set, IEEE/OSA Journal of Lightwave Technology, 12(2) (1997).
56. E. Suhir, Predicted thermal mismatch stresses in a cylindrical bi-material assembly adhesively bonded at the ends, ASME Journal of Applied Mechanics, 64(1) (1997).

57. E. Suhir, Thermal stress in a polymer coated optical glass fiber with a low modulus coating at the ends, Journal of Materials Research, 16(10) (2001).
58. E. Suhir, Coated optical glass fiber, U.S. Patent #6,647,195, 2003.
59. E. Suhir, Polymer coated optical glass fibers: review and extension, Proceedings of the POLYTRONIK'2003, Montreaux, October 21–24, 2003.
60. E. Suhir, Modeling of thermal stress in microelectronic and photonic structures: role, attributes, challenges and brief review, Special Issue, ASME Journal of Electronic Packaging, 125(2) 2003.
61. E. Suhir, V. Ogenko, and D. Ingman, Two-point bending of coated optical fibers, Proceedings of the PhoMat'2003 Conference, San-Francisco, CA, August 2003.
62. E. Suhir, Mechanics of coated optical fibers: review and extension, ECTC'2005, Orlando, FL, 2005.

Elastic stability and microbending

63. E. Suhir, Effect if the initial curvature on the low temperature microbending in optical fibers, IEEE/OSA Journal of Lightwave Technology, 6(8) (1988).
64. E. Suhir, Spring constant in the buckling of dual-coated optical fibers, IEEE/OSA Journal of Lightwave Technology, 6(7) (1988).
65. E. Suhir, Mechanical approach to the evaluation of the low temperature threshold of added transmission losses in single-coated optical fibers, IEEE/OSA Journal of Lightwave Technology, 8(6) (1990).
66. S.T. Shiue and S.B. Lee, Thermal stresses in double-coated optical fibers at low temperature, Journal of Applied Physics, 72(1) (1992).
67. S.T. Shiue and S.B. Lee, Thermal stresses in double-coated optical fibers at low temperature, Journal of Applied Physics, 72(1) (1992).
68. S.T. Shiue, Design of double-coated optical fibers to minimize hydrostatic-pressure-induced microbending losses, IEEE Photonics Technology Letters, 4(7) (1992).
69. E. Suhir, Elastic stability, free vibrations, and bending of optical glass fibers: the effect of the nonlinear stress-strain relationship, Applied Optics, 31(24) (1992).
70. S.T. Shiue, Axial strain-induced microbending losses in double-coated optical fibers, Journal of Applied Physics, 73(2) (1993).
71. F. Cocchini, Double-coated optical fibers undergoing temperature variations-the influence of the mechanical behavior on the added transmission losses, Polymer Engineering and Science, 34(5) (1994).
72. S.T. Shiue, Thermal stresses in tightly jacketed double-coated optical fibers at low temperature, Journal of Applied Physics, 76(12) (1994).
73. S.T. Shiue, The axial strain-induced stresses in double-coated optical fibers, Journal of the Chinese Institute of Engineers, 17(1) (1994).
74. S.T. Shiue, Thermally induced microbending losses in double-coated optical fibers at low temperature, Materials Chemistry and Physics, 38(2) (1994).
75. S.T. Shiue, The hydrostatic pressure induced stresses in double-coated optical fibers, Journal of the Chinese Institute of Engineers, 17(4) (1994).
76. P. Ostojic, Stress enhanced environmental corrosion and lifetime prediction modeling in silica optical fibers, Journal of Materials Science, 30(12) (1995).
77. E. Suhir, V. Mishkevich, and J. Anderson, How large should a periodic external load be to cause appreciable microbending losses in a dual-coated optical fiber? in E. Suhir, Ed., Structural Analysis in Microelectronics and Fiber Optics, ASME Press, 1995.
78. M. Uschitsky and E. Suhir, Epoxy-bonded optical fibers: the effect of voids on stress concentration in the epoxy material, in E. Suhir, Ed., Structural Analysis in Microelectronic and Fiber-Optic Systems, ASME Press, 1995.
79. S.T. Shiue, The spring constant in the buckling of tightly jacketed double-coated optical fibers, Journal of Applied Physics, 81(8) (1997).
80. S.T. Shiue and W.H. Lee, Thermal stresses in carbon coated optical fibers at low temperature, Journal of Materials Research, 12(9) (1997).
81. E. Suhir, Coated optical fiber interconnect subjected to the ends offset and axial loading, Int. Workshop on Reliability of Polymeric materials and Plastic Packages of IC Devices, Paris, Nov. 29–Dec. 2, 1998, ASME Press, 1998.
82. E. Suhir, Critical strain and postbuckling stress in polymer coated optical fiber interconnect: what could be gained by using thicker coating? Int. Workshop on Reliability of Polymeric materials and Plastic Packages of IC Devices, Paris, Nov. 29–Dec. 2, 1998, ASME Press, 1998.

Solder materials and joints

83. E. Suhir, Thermally induced stresses in an optical glass fiber soldered into a ferrule, IEEE/OSA Journal of Lightwave Technology, 12(10) (1994).
84. E. Suhir, Solder materials and joints in fiber-optics: reliability requirements and predicted stresses, Proceedings of the International Symposium Design and Reliability of Solder Joints and Solder Interconnections, Orlando, FL, 1997.

Dynamic response

85. E. Suhir, Vibration frequency of a fused biconical taper (FBT) lightwave coupler, IEEE/OSA Journal of Lightwave Technology, 10(7) (1992).
86. E. Suhir, Free vibrations of a fused biconical taper lightwave coupler, International Journal of Solids and Structures, 29(24) (1992).
87. E. Suhir, Is the maximum acceleration an adequate criterion of the dynamic strength of a structural element in an electronic product? IEEE Transactions on Components, Packaging and Manufacturing Technology, 20(4) (1997).
88. E. Suhir, Dynamic response of microelectronics and photonics systems to shocks and vibrations, INTERPack'1997 Proc., Hawaii, June 15–19, 1997.
89. E. Suhir, Could shock tests adequately mimic drop test conditions? IEEE ECTC Conference Proceedings, San-Diego, CA, May 28–31, 2002.
90. E. Suhir, New nano-particle material (NPM) for micro- and opto-electronic packaging applications, IEEE Workshop on Advanced Packaging Materials, Irvine, March 2005.

Nano-technology based new generation of fiber coatings

91. D. Ingman and E. Suhir, Optical fiber with nano-particle cladding, Patent Application, 2001.
92. E. Suhir, Strain free planar optical waveguides, U.S. Patent #6,389,209, 2002.
93. E. Suhir and D. Ingman, New hermetic coating for optical fiber dramatically improves strength: new nanoparticle material (NPM) and NPM-based new generation of optical fiber claddings and coating, U.S. Navy Workshop, St. Louis, MO, 2003.
94. E. Suhir, Polymer coated optical glass fiber reliability: could nano-technology make a difference? Polytronic'04, Portland, OR, September 13–15, 2004.
95. D. Ingman, T. Mirer, and E. Suhir, Dynamic physical reliability in application to photonic materials, Chapter 17, present book.

8

Area Array Technology for High Reliability Applications

Reza Ghaffarian

Jet Propulsion Laboratory, California Institute of Technology, CA, USA

8.1. INTRODUCTION

Commercial-off-the-shelf (COTS) area array packaging technologies in high reliability versions are now being considered for applications, including use a number of NASA electronic systems being utilized for both the Space Shuttle and Mars Rover missions. Understanding process and quality assurance (QA) indicators for reliability are important for low risk insertion of these newly available packages in high reliability applications. This chapter is based on a survey of those packages with the greatest potential utilization for high reliability applications. It provides a body of knowledge (BOK) literature survey for designing, manufacturing, and testing area array packages on printed wiring assemblies (PWAs a.k.a. CCAs). In addition, it summarizes lessons learned from previous activities and the most recent information gathered by reviewing package suppliers' data sheets, presentations at recent conferences, papers in proceedings, and literature search on inspection, quality assurance, and reliability, with special emphasis on packaging for board level (2nd) systems.

A large number of papers and books are now published in area array technology [1–5]. Data for assemblies with lead-free solder alloy interconnections are now being gathered [6]. In addition an industry standard was recently released specifically addressing the issues associated with qualification of area array technology. This specification was published by the IPC, a worldwide trade association of the interconnection electronics industries (IPC). This standard is IPC-9701 [7], "Performance Test Methods and Qualification Requirements for Surface Mount Solder Attachments." The updated revision of this specification, IPC-9701A, includes guidelines for lead-free solder interconnections. Activities carried out with team members from different high reliability sectors have contributed significantly to the information found in this chapter. Specifically, the thermal cycle test results for plastic wire-bond ball grid array (PBGA), flip chip BGA (FCBGA), ceramic BGA, and ceramic column grid array (CCGA) assemblies on printed wiring boards (PWBs) are discussed in details. Corner staking adhesives are generally used to enhance resistance to mechanical shock and vibration. Effects of such adhesives on reliability are also presented.

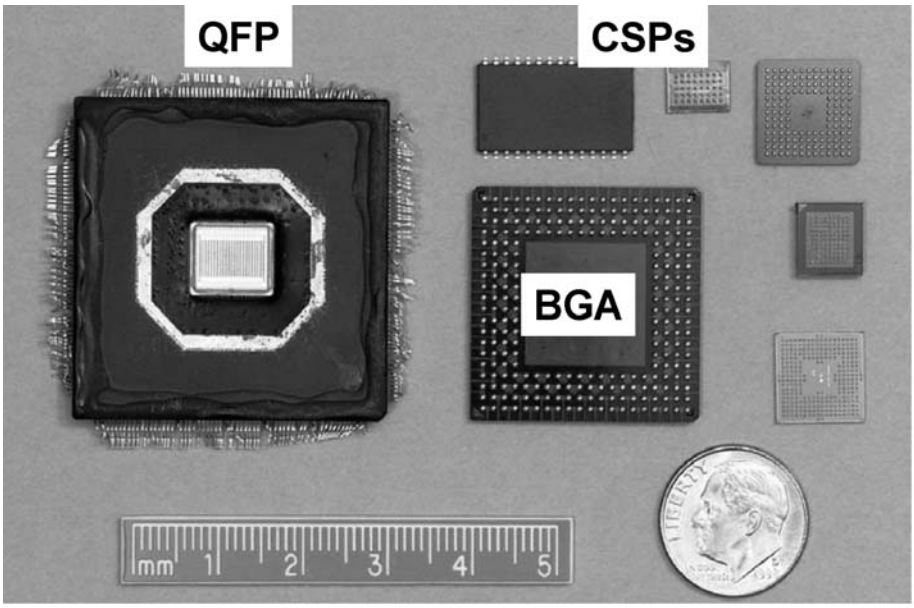


FIGURE 8.1. Miniaturization trends in surface mount electronics packaging—from quad flat pack (QFP) to ball grid array (BGA) and chip scale package (CSP).

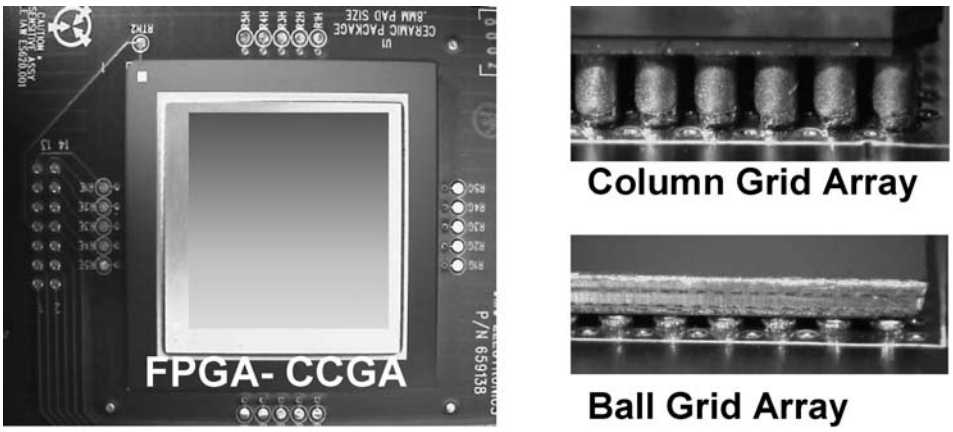


FIGURE 8.2. Photos of a column grid array (left) with details on column (top right) and ball interconnections.

8.2. AREA ARRAY PACKAGES (AAPS)

Figure 8.1 compares a conventional leaded package with a ball grid array (BGA) package and a package of even smaller configuration, a chip scale package (CSP). Figure 8.2 shows higher magnification of the solder balls of a BGA and the columns of a ceramic column grid array (CCGA or CGA). Area array packages, e.g., BGAs and CCGAs, with 1.27 mm pitch (pitch = the distance between adjacent ball centers) and finer pitch versions with 1 mm pitch are the only choice for packages with higher than 300 I/O (input/output) counts, replacing leaded packages such as the quad flat pack (QFP). The area

array packages also provide improved electrical and thermal performance, more effective manufacturing (that is, improved manufacturability), and ease-of-handling compared to conventional surface mount (SMT) leaded parts.

Finer pitch area array packages (FPBGA a.k.a. CSPs) are more miniaturized versions of BGAs, or smaller configurations of leaded and leadless packages with pitches generally less than 1 mm. These electronics packages have low mass and small chip sizes. They are typically used for low I/O (<100) memories and have the potential for use at higher I/Os (>300). In low I/O versions, CSPs replace conventional surface mount leaded parts such as thin small outline packages (TSOPs). In spite of the advantages of area array packages, several issues exist regarding their implementation for high reliability applications mainly inspection and reliability. Risk can be partially mitigated by prudent selection from a variety of the COTS available packages and control of manufacturing processes.

8.2.1. Advantages of Area Array Packages

Area array packages offer several distinct advantages over fine pitch surface mount components having gull wing leads, including the following:

- High I/O capability, 100s to 1000s of balls can be built and manufactured; gull-wing leads are limited to less than 300 I/Os.
- Larger lead pitches are possible, significantly reducing the manufacturing complexities for high I/O parts.
- Higher packaging densities are achievable since the lead periphery envelope limit for the gull wing leads does not apply to area array; hence, it is possible to mount more packages per board.
- Faster circuitry speed than gull wing SMCs (surface mount components) because the terminations are much shorter and therefore less inductive and resistive.
- Better heat dissipation.
- Conventional SMT manufacturing and assembly technologies—such as stencil printing and package mounting—can be employed.

Area array packages are also robust in processing. This stems from their higher pitch (1.0–1.27 mm, typical), better lead rigidity, and self-alignment characteristics during reflow processing. This latter feature, self-alignment during reflow (attachment by heat), is very beneficial and opens considerably the manufacturing process window.

8.2.2. Disadvantages of Area Arrays

Area array packages, however, are not compatible with multiple solder processing methods, and individual solder joints cannot be inspected and reworked using conventional methods. For high reliability applications of SMT assemblies, especially those of NASA, the ability to inspect all solder joints visually has been a standard inspection requirement and is considered a key factor for providing confidence in the solder joint reliability. Continuous advanced inspection techniques, such X-ray and C-SAM, need to be developed to provide similar confidence levels for area array packages.

The four chief drawbacks of area array packages are:

- Lack of direct visual inspectability.
- Lack of individual solder joint reworkability.

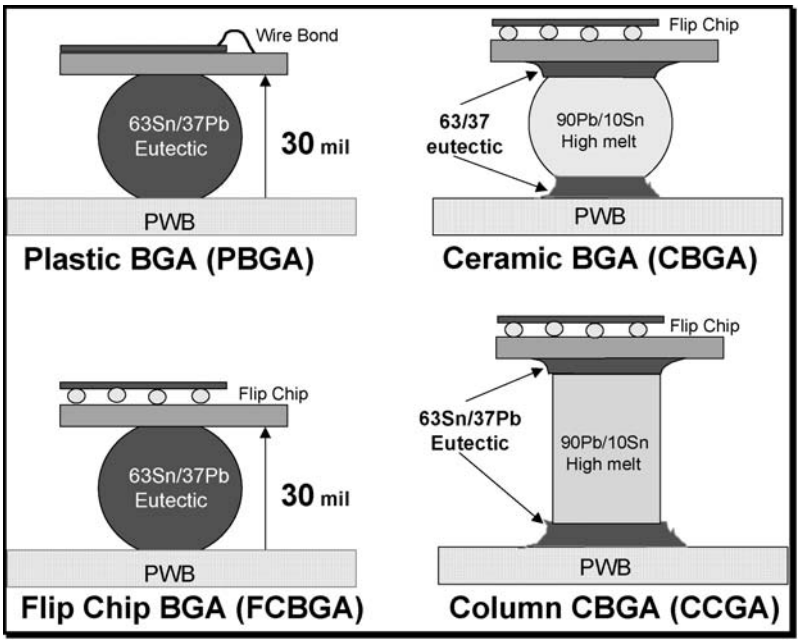


FIGURE 8.3. Plastic and ceramic package configurations.

- Interconnect routing between the chip and the PWB necessitates a multilayer PWB which are generally more difficult and more costly to produce.
- Reduced resistance to thermal cycling due to the use of rigid balls/columns.

8.2.3. Area Array Types

Examples of typical area array packages are shown in Figure 8.3. These include plastic ball grid arrays with a ball composition of eutectic Sn63Pb37 alloy or its slight variation. The ceramic BGA package uses a higher melting ball (Pb90Sn10) with eutectic attachment to the die and board. The column grid array (CCGA) is similar to the BGA except that it uses column interconnects instead of balls. The flip chip BGA (FCBGA) is similar to the BGA, except that the flip chip is internal to the package and a flip chip die is used.

8.3. CHIP SCALE PACKAGES (CSPS)

As mentioned previously, the trend is towards ever increasing I/Os on packages, and so this is driving the packaging configuration of semiconductors. Because of these advantages, chip scale packages (CSPs) are already making their appearance and are competing with bare die assemblies. Unlike conventional BGA technology at typically 1–1.27 mm pitch, CSPs utilize lower pitches, e.g., currently 0.8 to 0.4 mm. Hence, they have smaller sizes and their own set of challenges. Compared to bare die, their key advantages/disadvantages are listed in Figure 8.4. A photo comparing a wafer-level CSP to a QFP with the same I/O count is shown in Figure 8.5. The tiny CSP is located at the center of the QFP. Note the CSP's bottom side with solder balls is also apparent in the photo.

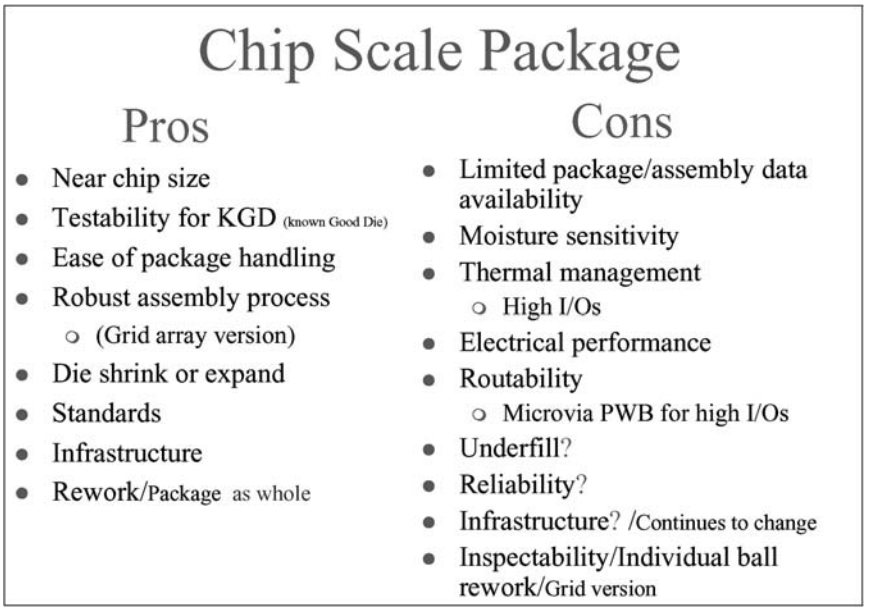


FIGURE 8.4. Pros and cons of the chip scale package (CSP).

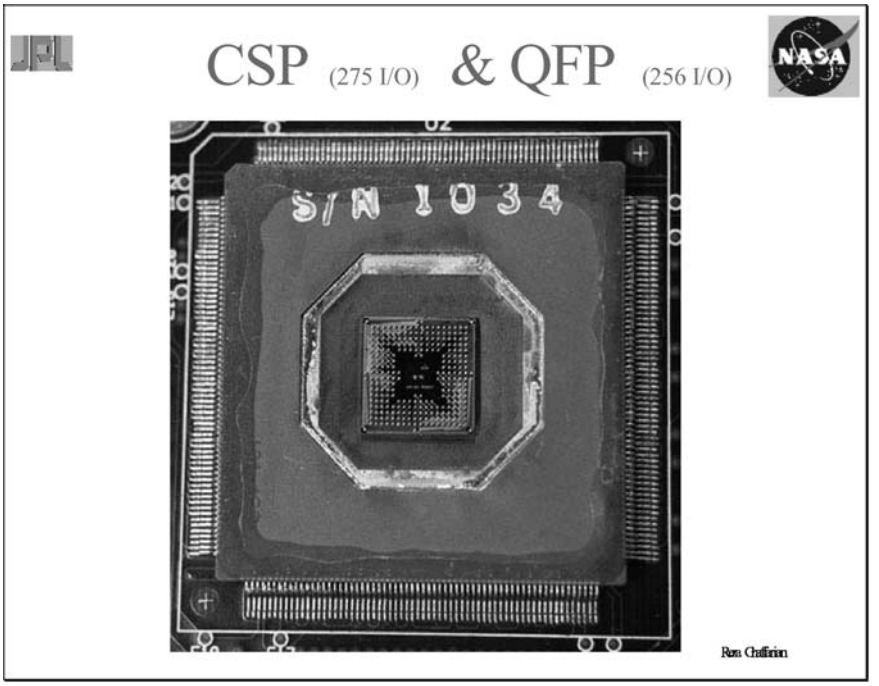


FIGURE 8.5. Size comparison of QFP with a wafer-level CSP—same I/O count.

CSPs, especially their area array versions, a.k.a. FPBGAs, are now widely used for many electronic applications, including portable and telecommunication products. The BGA version of the area array package, introduced in the late ‘80s and implemented with great caution in the early ‘90s, further evolved in the mid ‘90s to the CSP having a much finer pitch. Distinguishing between sizes and pitches have become difficult for the various array package versions. These are now categorized as area array packages in order to be able to distinguish them from the flip chip, bare die category. The bare dies have been around for a longer time, but their associated issues, including known good die and difficulty in direct attachment to printed wiring boards (PWBs), have limited their wide implementation.

The definition of CSP has evolved as the technology has matured. CSP now refers to a package with 0.8 mm or less pitch with some current packages having as low as 0.4 mm pitch. Packages with fine pitches, especially those with less than 0.8 mm and high I/Os, may require the use of microvia PWBs and are costly and may perform poorly when assembled onto these boards. One approach has been to increase functionality through systems-in-a-package technology, i.e., stacking die/packages while keeping the pitch within the limitation of current board technology.

The CSP technology continues to emerge with new packages, smaller pitches, and lower reliability. Thus, CSP technology is more challenging for use in high reliability space applications. Each individual package needs to be evaluated for design, construction, materials, ball shear, moisture, and so forth prior to acceptance and evaluation for 2nd level assemblies. IPC-9701 should be used as a guideline for solder joint qualification.

8.4. PLASTIC PACKAGES

8.4.1. Background

Plastic BGAs with a variety of sizes and shapes are abundantly available, and are used widely by commercial industry for a variety of applications from benign office environments to high-end server applications. The military and avionic industries are also using them selectively after they have reached an acceptable level of maturity. Because of their wider applications, reliability of these packages is generally characterized by suppliers and verified by industry. Numerous excellent publications and references are available through the IEEE CPMT, SMTA, IMAPS, and the IPC proceedings and through journals such as the Microelectronic Reliability Journal.

In what follows, first, plastic package types will be described followed by a discussion on 2nd level assembly reliability. Data from researchers as well as suppliers will be tabulated for a number of packages from low to very high I/Os. Also, reliability test results performed by a team of high reliability industry representatives will be presented and compared to those from the literature for single- and double-sided assemblies. Second, discussion will be given on the types of ceramic area array packages, ball and column arrays, and their assembly reliability. Test results from these team activities and those performed specifically for NASA use will be presented and compared to literature data.

8.4.2. Plastic Area Array Packages

Three BGA package configurations are popular. These are shown in Figure 8.6. These are:

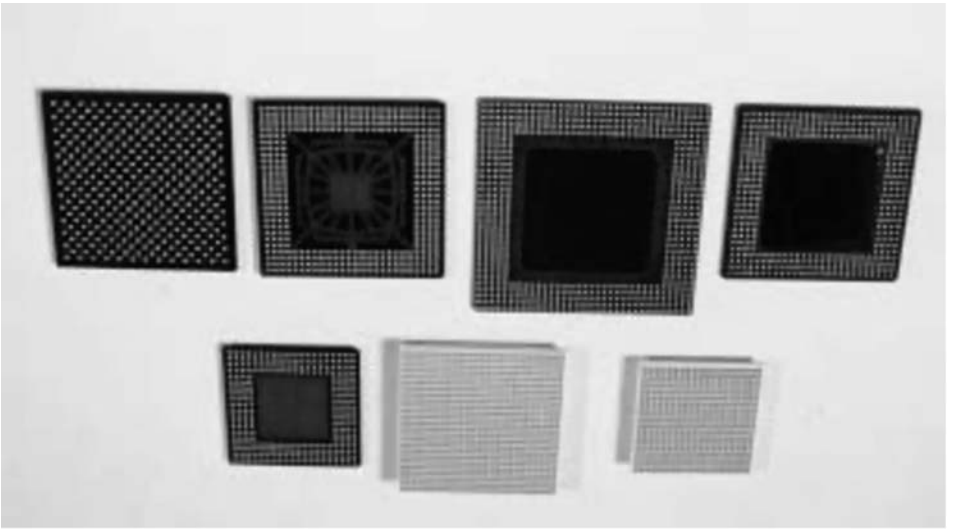


FIGURE 8.6. Ceramic (two on the bottom right) and plastic BGA packages of test vehicles.

- Full array.
- Staggered array.
- Peripheral array.

Plastic packages come in all styles whereas ceramic packages are generally limited to the full array configuration.

Fully populated BGA packages present some significant routing challenges if conventional printed wiring board with plated-through-hole (PTH) vias are considered for design. The staggered grid BGA uses only half of the available sites and therefore, similar to peripheral arrays, eases routability. A package with the staggered grid pattern of 1.27 mm pitch provides an effective minimum pitch of 1.8 mm along the diagonal of the package.

A peripheral array format for plastic packages, with the die at the package center, was developed to increase thermal cycle reliability as well as to ease routability. Premature failures of solder joints under the die, due to a large CTE mismatch between the die and the PWB, have been observed. Removal of the center solder balls, however, will slightly degrade thermal performance of the package. To improve thermal dissipation, generally, a number of non-functional thermal balls are generally added to the middle of the package.

Die size in the plastic package is an important factor influencing solder joint performance. Experimental results indicate that solder joints close to the perimeters of the die fail first under temperature cycling. It has been shown that the number of cycles to failure is not only sensitive to the I/O count, but also to the die size [5]. The significant CTE differences between a ceramic die (2.3 ppm/°C) and the PWB (about 15 ppm/°C) and encapsulating epoxy ($>$70 ppm/°C) are the key contributing factors to the failures associated with the die.

8.4.3. Plastic Package Assembly Reliability

Table 8.1 lists cycles to failure for a number of plastic packages with different configurations, selected from those reported in the literature [8–13]. Data were chosen to illustrate the effects of a few key parameters on reliability. The following parameters were consid-

TABLE 8.1.
Cycles-to-failure data illustrating the effect of a number of key variables.

Case number	Package I/Os, pitch	Pkg size (die size) mm	Thermal cycle condition (ramp, dwell, cycle/hr)	First failure	Mean life ($N_{63.2\%}$)	Comments
1	PBGA-119-1.27	27 × 27 (17.8 × 17.8 × 0.3)	0°C/100°C (10 min, 5 min, 2)	6260 (1% failure)	12215	27 Pkg, Ref. A. Mawer [8,13]
2	PBGA-256-1.27	27 × 27 (10 × 10)	0°C/100°C (10 min, 5 min, 2)	3200 (1% failure)	(6195)	Ref. Amkor [9]
3	FCBGA-1849-1.27	??	0°C/100°C	3095 (1% failure)	4710	Ref. Shiah [12]
4	PBGA-256-1.0	17 × 17 (8.80 × 7.9)	0°C/100°C (10 min, 5 min, 2)	3687 (1% failure)	NA	Full Array PWB, 2.3 mm Thk, Ref. Altera [10]
5	PBGA-676-1.0	27 × 27 (17.8 × 17.8 × 0.3)	0°C/100°C (10 min, 5 min, 2)	5909	9267	15/32 Fail, PWB, 1.6 mm Thk, Ref. Xilinx [11]
6	SBGA-860-1.0	42.5 × 45.5 (22.45 × 21.44 × 0.3)	0°C/100°C (10 min, 5 min, 2)	>8824	N/A	0/32, No failure to 8824 cycle, PWB 1.6 mm Thk, Ref. Xilinx [11]
7	FCBGA-1020-1.0	33 × 33 (22.6 × 19.9)	0°C/100°C (2 cycles/hr)	5670 (1% failure)	NA	PWB Thk 2.3 mm 6 layer build up BT, Ref. Altera [10]
8	FCBGA-1020-1.0	33 × 33 (17.9 × 16.7)	0°C/100°C (2 cycles/hr)	2770 (1% failure)	NA	PWB Thk 2.3 mm 6 layer build up BT + Cu heat sink, Ref. Altera [10]
9	PBGA-1156-1.0	35 × 35 (23.11 × 21.13 × 0.3)	0°C/100°C (10 min, 5 min, 2)	7289	9350	8/32 Fail, 1.6 mm Thk, Ref. Xilinx [11]
10	PBGA-313-1.27	35 × 35 (13 × 13)	−30°C to 100°C (25 min, 15 min, 0.75)	3310 (1% failure)	4000	13 Pkg, Ref. Evans [13]
11	PBGA-256-1.27	27 × 27 (10 × 10)	−40°C–125°C (15 min, 15 min, 1)	∼2000 (1% failure)	(3164)	Ref. Amkor [9]
12	PBGA-676-1.0	27 × 27 (17.8 × 17.8 × 0.3)	−40°C–125°C (15 min, 15 min, 1)	1341	1830	27/32 Fail, PWB 1.6 mm Thk, Ref. Xilinx [11]
13	PBGA-1156-1.0	35 × 35 (23.11 × 21.13 × 0.3)	−40°C–125°C (15 min, 15 min, 1)	1601	2386	30/32 Fail, PWB 1.6 mm Thk, Ref. Xilinx [11]

ered when test data were tabulated even though in some cases specific information was not reported.

Thermal cycle range, ramp rate, dwell times

- For example, the CT1%F (cycles-to-one percent-failure) for the PBGA 256 in the range of 0°/100°C was 3200 cycles (Case #2); and was reduced to about 2000 cycles when the temperature range increased to −40/125°C (Case #11).

Package size, thickness, configuration, and I/Os

- For example, comparing Case #2 to Case #3, a small reduction in cycles-to-first failure is shown when package I/O is increased from the 256 I/Os, 1.27 mm, to 1849 I/Os (3200 vs 3095 cycles in the range of 0°/100°C). A slightly higher reduction is shown when test results for the 256 I/Os (Case #11) package with 1.27 mm pitch are compared to the 1156 I/Os (Case #13) with 1.0 mm pitch, 2000 vs 1601 cycles in the range of −40° to 125°C.

Die size and its relation to package size and ball configuration

- The effects of these parameters on reliability are not apparent from the cases presented in Table 8.1, but the die and package sizes are listed in the table for the purpose of identifying such correlation. It is seen that as the die size increased, the cycles-to-failure decreased. Comparing Case #7 and #8 illustrates an increase in cycles to failure when the die size is increased. The opposite effect shown here might be due to the confounding effects of added a heat sink for Case #8.

PWB thickness, definition of pad, surface finish

- Preferred thickness was defined as 2.3 mm in IPC-9701 [7] since it is known that generally packages assembled on thinner PWBs show higher cycles to failure. So comparing Case #4 to #5, one may conclude that one possible reason that the PBGA with 676 I/Os exhibits higher thermal resistance than the 256 I/Os, is because of the use of a thinner board for the 676 I/O package assembly.

Single side or double side, relative offset of package on top and bottom

- This is not shown in Table 8.1, but refer to Table 8.2 for significant reduction when the double-sided mirror image packages assembly is considered.

8.4.4. Reliability Data for BGA, Flip Chip BGA, and CSP

Reliability of numerous area array package technologies for 2nd level assemblies were investigated [1–6,13,21]. These included ball grid arrays (BGAs), fine pitch BGAs (FPBGAs), flip chip ball grid arrays (FCBGAs), chip scale packages (CSPs), flip chip die, and thin small outline packages (TSOPs). Packages having I/Os ranging from 48 to 784 and pitches varying from 0.5 to 1.27 mm were included in the investigation. The packages were mounted on multilayer FR-4 PWBs or the polyimide version. The test vehicles were subjected to numerous thermal cycling conditions including −55° to 125°C with a near thermal shock condition. Cycles-to-failure (CTF) test results up to 1,500 cycles are compared under this condition for 784 I/O FCBGAs, 175 I/O FPBGAs, and 313 I/O PBGAs. Key test results from the investigation are summarized and compared to those given in the literature, including reduction in CTF due to double-sided mirror image assembly.

8.4.4.1. Thermal Cycle Profiles Numerous cycle profiles have been used over the years of thermal cycling testing. These are summarized below and references are made to them in this and subsequent sections. These are:

- Cycle A: The cycle A condition ranges from −30° to 100°C and has an increase/decrease heating rate of 2° to 5°C/min and dwell of about 20 minutes at the high temperature to assure near complete creeping. The duration of each cycle is 82 minutes.
- Cycle B: The cycle B condition ranges from −55° to 100°C with long time duration. The heating and cooling rates are 2° to 5°C per minute with an oven dwell setting of 45 minutes at the two extreme temperatures. The duration of each cycle is 246 minutes.
- Cycle C: The cycle C condition ranges from −55° to 125°C with a 2° to 5°C/min heating/cooling rate. Dwells at extreme temperatures are at least 10 minutes with a duration of 159 minutes for each cycle.
- Cycle D: The cycle D condition ranges from −55° to 125°C, the same as condition C, but with a very high heating/cooling rate. It can also be considered a thermal shock since it uses a three-region chamber: hot, ambient, and cold. Heating and cooling rates are nonlinear and vary between 10° to 15°C/min with dwells at the extreme temperatures of about 20 minutes. The total cycle lasted approximately 68 minutes.
- Cycle E: The cycle E condition ranges from −50° to 75°C with a 2° to 5°C/min heating/cooling rate. Dwell at extreme temperatures are at leasts 10 minutes with a duration of 105 minutes for each cycle.
- Cycle F: The cycle F condition ranges from −55° to 100°C with a 2° to 5°C/min. Dwells at extreme temperatures are 30 minutes.

The criteria for an open solder joint specified in IPC-9701 and IPC-SM-785 were used as guidelines to interpret electrical interruptions. Generally, once the first interruption was observed, there were many additional interruptions within 10% of the cycle life. This was especially true for ceramic packages, but it was not apparent for plastic packages.

8.4.4.2. Results of CTF for −55°/125°C to the Literature for 0°/100°C Table 8.2 provides a comparison of the CTF data given above for 175 I/O FPBGAs to CTF data published in the literature for solder joint reliability in the range of 0° to 100°C [14]. The referenced papers provide an in-depth characterization of the effect of the package substrate thickness and rework on the reliability for both single and double-sided mirror image assemblies. Here, the comparison was made to data for a new, modified package with a substrate thickness of 0.25 mm, rather than to a previous version with 0.11 mm thick substrates. An acceleration ratio value of 3.8 was calculated when the two N_o values are compared. This is a reasonable ratio and approximately predicted by the modified Coffin-Manson relationship considering that other differences are not included in this relationship including the thermal profile and the PWB thickness. Here TV means test vehicle.

8.4.4.3. Summary of Test Results for FPBGAs and FCBGAs The following points summarize the test results based on a limited number of test vehicles that were subjected to the various thermal cycle conditions.

- Cycles-to-failure (CTF) data for a 784 I/O full array FCBGA were between 1251 and 1489 cycles under condition D (−55°/125°C, near thermal shock) for the best board and assembly conditions. The CTF decreased significantly to between 607 and 829

TABLE 8.2.
Comparison of TV-2 FPBGA thermal cycle test results (−55°/125°C) to literature data for 0°/100°C thermal cycle range.

TV ID	Board thickness (NSMD pad size)	Via location (diameter)	Thermal cycle range (total time)	Weibull scale ($N_{o)}$	Weibull shape (m)	Acceleration ratio
175 I/O FPBGA*	1.57 ± 0.2	On Pad	0°C to 100°C			
Single-Side		(125 μm)	(32 min)			
Condition 1	(300 μm)			4331	11.1	N/A
Condition 2	(400 μm)			3525	9.1	
175 I/O FPBGA*						
Double-Side						
Condition 1	(300 μm)			1616	17.6	N/A
Condition 2	(400 μm)			1163	10.5	
TV-2 175 I/O FPBGA	1.27	On Pad	−55°C to 125°			
9 data points	(300 μm)	(100 μm)	(68 min)			
8 data points				1126	6	3.8
				1134	11.9	

cycles for another PWB batch from the same manufacturer. CTF for this package were lower than for another assembly populated with a PBGA with 313 I/Os, wire bond/molded compound and depopulated full array package.

- The CTF data for a 175 I/O FPBGA with a 0.8 mm pitch was between 691 and 1231 cycles under condition D. The trend was approximately the same as those reported in the literature for the identical package, but having a thicker substrate, when they were subjected to another thermal cycle condition, and the cycles to failures were adjusted for an acceleration factor using a modified Coffin-Manson relationship.
- The CTF data generated for these assemblies using a second board manufacturer were poor for both the C (−55°/125°C) and D thermal cycle conditions. The trend in failures appears to be in agreement with inspection observations performed qualitatively for PWB warpage. No specific quantitative numerical values for warpage distribution surface plots were gathered for each individual board and package site. These values are needed to be able to draw a strong relationship between package size and local board warpage.

8.5. CERAMIC PACKAGES

8.5.1. Background

Conventional ceramic ball grid arrays on polymeric boards have very limited assembly level thermal cycle resistance. Ceramic packages have been widely used for high-end commercial microprocessors; however, their use for the desktop is limited because of added on/off thermal cycle requirements due to the need for energy conservation. The Advanced Configuration Power Interface (ACPI) defines low power system states that the operating system controls [15]. These ACPI systems are directed at reducing the energy consumption of desktop computer equipment during periods of inactivity as well as providing quick resumption of activity. To conform to ACPI, systems will power down more often when not in use during typical daily conditions. Each power down cycle may result in a temperature

excursion and affects the fatigue life of area array solder joints. This requirement results in the need for an increased reliability of the soldered surface mount ceramic packages.

At high pin counts and a high number of on/off cycles due to ACPI, conventional ceramic ball grid arrays (CBGAs) marginally meet the reliability requirements of desktop personal computers. Typical desktop requirements prior to ACPI ranged from 1,250–1,875 cycles at 1–1.5 cycles per day [15]. The temperature excursion that a package experiences due to a power down state is about 40°C. During a typical day, there may be 4–8 ACPI cycles in addition to the full cycles. The power swing during transition from certain ACPI states can be as high as 70% of a full power down. Because of this new requirement, commercial industries have sought ceramic packaging technologies with a greater reliability margin than obtained with CBGAs including CCGAs.

8.5.2. Ceramic Package Assembly Reliability

The key factors that affect 2nd level reliability of CBGAs and CCGAs for a given temperature excursion are:

- CTE mismatch between the ceramic package and the FR-4 board.
- Rigidity of the ceramic substrate in ceramic BGA. The more rigid and thicker the substrate, the lower the reliability. This effect may not be as severe for CCGAs as it is for CBGAs.
- Standoff between the package and the board.
- Type of solder joint geometry, e.g., ball or columns which determines the compliance of the joint. Variations include cast solder column and solder column interposer.

A few of these variables are discussed in further detail below.

8.5.2.1. High CTE Ceramics For ceramic area array packages, generally, the CTE difference between the substrate and the PWB is large, thus, these packages experience severe shear strain resulting in damage to the solder joints. For the conventional ceramic package mounted on a polymeric PWB, this shear strain is of concern, since the CTE of alumina ceramics is approximately about 7 ppm/°C whereas a typical PWB (FR-4 board) is approximately 12–16 ppm/°C. In reference [15], a FEM (Finite Element Method) analysis was performed to determine the optimum CTE taking into account the effects of CTE mismatch between the package, PWB, and the silicon IC. The optimum CTE was found to depends on the package configuration and for common package sizes. It ranged from 9–13 ppm/°C.

8.5.2.2. Solder Column Interposer Column grid arrays come in many forms. The most popular versions are those with the floating columns during solder reflow [4]. Columns rigidly connected (cast) at the fixed sites is another version of the column grid array package. Use of a ceramic interposer is a new alternative that was developed with the aim of improving attachment reliability by using a thin ceramic between the package and the columns. The solder interposer is a thin ceramic body 0.3 mm thick with an array of holes at 1.27 mm pitch. The holes are punched in the green state, metalized with molybdenum (Mo), and sintered using a conventional ceramic process. The metalized holes are then plated with nickel/gold (Ni/Au). The Pb90/Sn10 solder columns are then attached by reflow to form a metallurgical bond between the solder and the plated hole in the ceramic. Eutectic solder is then applied on the opposite side of the columns. This eutectic solder allows the subsequent attachment to the package. Figure 8.7 shows a land grid array package attached to a solder column interposer. Added to this figure are photomicrographs of

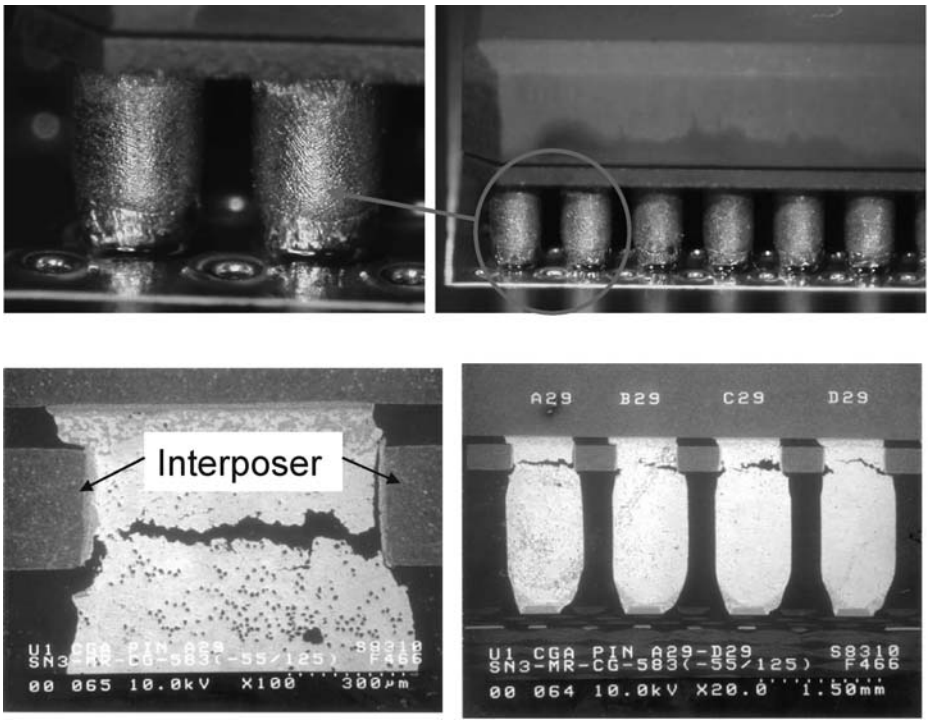


FIGURE 8.7. Ceramic column grid array with interposer technology and failure after thermal cycling.

cross-sections of the column interposers. Similar failures are also found in reference [15] showing construction of a similar interposer and failures due to thermal cycling.

8.5.2.3. Effect of PWB Thickness on CBGA Reliability It is now well established that the PWB thickness generally affects solder joint reliability of area array package assemblies, the reliability decreasing with increasing thickness. For this reason, a relatively higher PWB thickness was recommended to be used as a control by IPC 9701 committee, i.e., 0.093 inch and other thicknesses specific to application. Finite element analysis (FEA) projections for the effect of PWB (card) thickness were compared to the test results for the CBGA [16]. Because of reasonably good agreement of projections with cycles-to-failure test data, the model was used to estimate the effect of thickness. It was shown that the relationship of CTF with board thickness is nonlinear. CTF for board thicknesses of 0.040, 0.06, 0.08, and 0.10 inches were projected to be about 2,600, 1,600, 1,200, and 1,000 cycles, respectively. It is not known whether a similar effect is applicable to ceramic column grid arrays.

8.5.3. Literature Survey on CBGA/CCGA Assembly Reliability

Table 8.3 lists cycles to failure for a number of CCGAs/CBGAs with different configurations, selected from the very limited data set reported on in the literature [15–20]. Data were chosen to be able illustrate the effects of a few key parameters on the reliability. The following parameters were considered when test data were tabulated even though in some cases specific information was not reported and in fact is missing.

TABLE 8.3.
Cycles-to-failure data illustrating the effect of a number of key variables.

Case number	Package I/O-pitch	Pkg size (die size) mm	Thermal cycle condition (ramp, dwell, cycle/hr)	First failure	Mean life ($N_{63.2\%}$)	Comments
1	CBGA-255-1.27	21 × 21 (1 mm substrate)	0°C to 100°C (10, 5, 2 cycles/hr)	1980 (1% failure)	2426 ($N_{50\%}$)	PWB, 1.55 mm Thk, Ref. Burrnette [17]
2	CBGA-361-1.27	25 × 25 (substrate 0.8 mm Thk)	0°C to 100°C (3 cycles/hr)	NA	4535 ($N_{50\%}$)	Avg solder paste vol 5,900 mil^3 PWB, 1.57 mm Thk, Die 15 × 10 mm
3	CBGA-361-1.27	25 × 25 (substrate 1.2 mm Thk)	0°C to 100°C (3 cycles/hr)	NA	2700 ($N_{50\%}$)	Note: Increase from die Thk .8 to 1.2 and 2.9, reliability reduction by 1.8 and 3.2 times, Ref. [17]
5	CBGA-625-1.27	32.5 × 32.5 (substrate 0.8 mm Thk)	0°C to 100°C (3 cycles/hr)	NA	2462 ($N_{50\%}$)	PWB, 1.57 mm Thk, Ref. [18]
6	CBGA-361-1.27	25 × 25 (substrate 0.8 mm Thk)	−55°C to 110°C (2 cycles/hr)	890 (100 ppm)	1,190 ($N_{50\%}$)	PWB, 1.57 mm Thk, Ref. [15]
7	CBGA-361-1.27-HiCTE Substrate	25 × 25 (substrate 0.8 mm Thk)	−55°C to 110°C (2 cycles/hr)	1,310 (100 ppm)	2,160 ($N_{50\%}$)	Substrate CTE, 12.2 ppm, Ref. [15]
8	CGA-361-1.27-Interposer	25 × 25 (substrate 0.8 mm Thk)	−55°C to 110°C (2 cycles/hr)	1,350 (100 ppm)	2,320 ($N_{50\%}$)	NTK Interposer CGA PWB, 1.57, Ref. [15]
9	CGA-361-1.27-IBM	25 × 25 (substrate 0.8 mm Thk)	−55°C to 110°C (2 cycles/hr)	1,080 (100 ppm)	1,520 ($N_{50\%}$)	Ref. [15]
10	CBGA-1681-1.27-HiTCE	42.5 × 42.5 × 1.85 (substrate)	−25°C to 125°C (1, 9 min, 3 cycles/hr)	613 (1st failure)	1142 ($N_{50\%}$)	PWB, 93 mm Thk Ref. [19]
11	CBGA-625-1.0	32 × 32 × 2.4 mm (substrate)	0°C to 100°C (2 cycles/hr)	NA	740 ($N_{50\%}$)	Ref. [16], IBM-2003
12	CBGA937-1.0	32 × 32 × 1.5 (substrate) 32 × 32 × 2.4	0°C to 100°C (2 cycles/hr)	NA	1,860 ($N_{50\%}$) 1,310 ($N_{50\%}$)	Ref. [16]
13	CCGA1657-1.0	42 × 42 × 1.5 42 × 42 × 2.55 42 × 42 × 3.7	0°C to 100°C (2 cycles/hr)	NA	1,530 ($N_{50\%}$) 990 ($N_{50\%}$) 620 ($N_{50\%}$)	Ref. [16]
14	CCGA 1657-1.0 Cu	42.5 × 42.5 × 2.55	0°C to 100°C (2 cycles/hr)	1660 (1st failure)	2410 ($N_{50\%}$)	Cu Column, solder paste 96.5 Sn3.5Ag, Ref. [20]

Thermal cycle range, ramp rate, dwell times

- For example, the CT50%F (cycles-to-fifty percent-failure) for the CBGA 361 over the range of 0°/100°C was 4,535 cycles (Case #2); it diminished to 1,190 cycles when the temperature range increased to −55°/110°C (Case #6)

Package size, thickness, materials, configuration, and I/Os

- Comparing Case #2 to Case #3, a relatively large reduction in CT50%F is shown when the package thickness is increased from 0.8 mm to 1.2 mm (4,535 vs 2,700 cycles). When the package thickness is further increased to 2.9 mm, the CT50%F is further reduced, a reduction by 3.2 times relative to the package with 0.8 mm thickness. A similar reduction was observed for the CCGA 1657 I/Os when the package thickness is increased from 1.5 mm to 3.7 mm (Case #13). The reliability will decreases by increasing the package I/O since the distance to neutral point is increased. CT50%F is reduced from 4,535 cycles to 2,462 cycles when the I/Os for the 0.8 mm thick package increased from 361 to 625 (Case #2 vs Case #5). Use of higher CTE ceramic materials—to better match the ceramic CTE to the PWB—will also improve the reliability. For example, compare Case #6 to the Case #7 for the 361 I/O CBGA assemblies. The CT50%F increased from 1,190 to 2,160 cycles for the HiCTE package.

Die size and its relation to the package size and ball configuration

- The effects of die size and package configuration (full vs peripheral) arrays on reliability are more pronounced for plastic than for ceramic package assemblies.

PWB thickness, definition of pad, surface finish

- The preferred thickness was defined as 2.3 mm in IPC-9701 [7] since generally plastic packages assembled on thinner PWBs show higher cycles to failure. The effect of board thickness for ceramic packages is not well established yet, but its effect may be less critical for column grid array than plastic package assemblies, especially when the dominant failure is the columns rather than solder joints.

Single side or double side, relative offset of package on top and bottom

Previously, an example was give for reduction in life cycle due to double-sided mirror image assemblies. The effect of mirror image assemblies on reliability for CBGAs/CCGAs is not presently known.

8.5.4. CBGA Thermal Cycle Test

The reliability of plastic and ceramic ball grid arrays was assessed as part of a previous JPL-lead BGA Consortium effort. Nearly 200 test vehicles, each with four packages, were assembled and tested using an experiment design [3]. The most critical variables incorporated in the experiment were package types both ceramic and plastic; board materials both FR-4 and polyimide; surface finishes such as OSP, HASL, and Ni/Au; solder volumes including low, standard, and high; and various environmental conditions.

Ceramic and plastic packages up to 625 I/Os were assembled on printed wiring boards and were subjected to various thermal cycle profiles to determine the solder joint reliability and failure mechanisms. Cycles-to-failure test results for the CBGA with 625

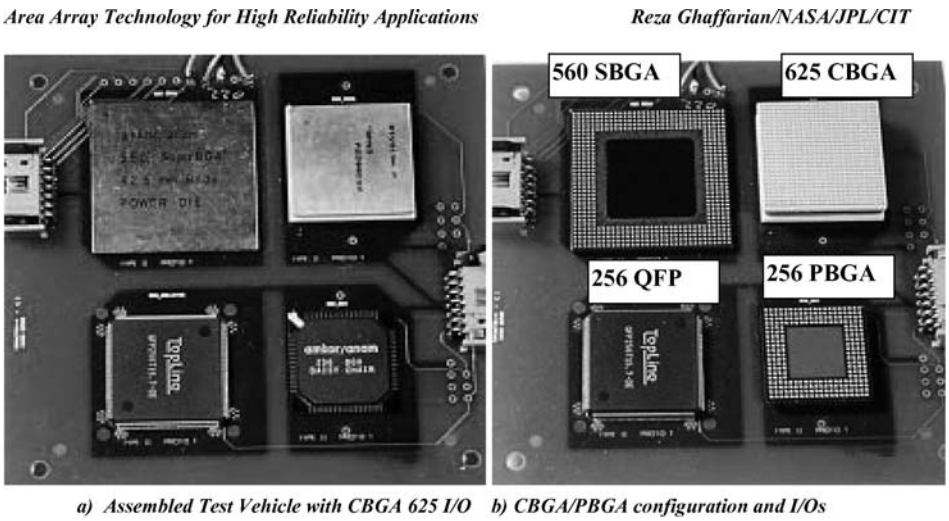


FIGURE 8.8. Test vehicle assembly and one with unassembled parts topping assembled ones.

I/Os under four different cycle regimes were presented. The failure data were used to determine validity of the Coffin-Manson relationship using a projection from one thermal cycle condition to other.

8.5.4.1. Test Vehicle for CBGA assemblies The test vehicles were populated with plastic (PBGA) and ceramic (CBGA) packages. Both FR-4 and polyimide PWBs with six layers (1.57 mm thick) were used. Only data for the CBGA 625I/O full arrays shown in Figure 8.8 are presented in the following section.

Ceramic solder balls (Pb90/Sn10) with 0.035-inch diameters having a higher melting temperature were attached to the ceramic substrate with a lower melting eutectic solder (Sn63/Pb37). During BGA assembly, eutectic solder on the package side and the eutectic paste on the PWB side melt during reflow to provide the electro-mechanical interconnects.

Full assembly was implemented after process optimization from the trial test. The following procedures were followed:

- The PWBs were baked at 125°C for four hours prior to screen printing.
- Two types of solder pastes were used: an RMA and a water soluble one.
- Pastes were screen printed and the heights were measured by a laser profilometer. Three levels of paste volumes were included in the evaluation: standard, high, and low. Stencils were stepped down to 50% in order to be able to accommodate a mixed assembly of ceramic, plastic, and fine pitch QFP packages on the Type 2 test vehicle assembly.
- A 10-zone convection oven was used for reflowing the solder paste.
- The first assembled Test Vehicle (TV) using an RMA reflow process was visually inspected and X-rayed to check solder joint quality.
- All assemblies were X-rayed.

8.5.4.2. Thermal Test Results

8.5.4.2.1. CBGAs Failure Mechanisms. Both board and package interface cracking was observed as the number of cycles increased. Figure 8.9 shows typical failures for two cycling conditions for a ceramic package. Failures under the A condition were generally from

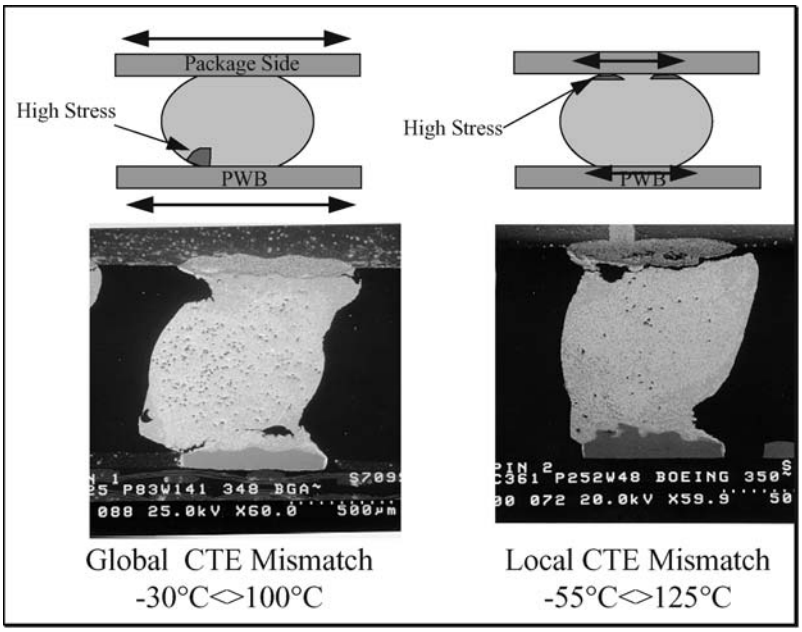


FIGURE 8.9. Cross-sections of failure sites for CBGA 625 after 350 cycles under A and D conditions.

the PWB site whereas those for the D condition were from the package sites. Failure mechanism differences may be explained by the existence of the stress state during cycling due to either global or local conditions.

Modeling indicates that the high stress regions shifted from the board to the packages when the stress conditions changed from the global to local. For the A cycling, with slow heat/cooling ramping, which allowed the system to reach uniform temperature, damage features indicate a global stress condition. On the other hand, for the D cycle with its rapid heating/cooling (−55°/125°C), the damage seems to indicate a local stress condition.

8.5.4.2.2. CBGA Cycles-to-Failure. Figure 8.10 shows cycles to first failure for the CBGA 625 under four different thermal cycling conditions. These plots were generated by ranking CTF from low to high and then approximating the failure distribution percentiles using a median plotting position, $F_i = (i - 0.3)/(n + 0.4)$. Weibull parameters were also generated and plotted in continuous graphs. Weibull parameters for the four conditions are listed in Table 8.4. In addition to cycling conditions, they include manufacturing and assembly variables that possibly affected the m values.

The widest spread, i.e., $m = 4$, was observed for those assemblies cycled under condition D. The other three conditions had m values ranging from 8.4 to 11.7 with the narrowest spread, $m = 11.7$, for condition C. Near-thermal shock with two hot and cold extreme temperatures and the use of a limited number of test samples are possible causes of the low values for m under the D condition. Only seven test vehicles were tested, three on polyimide, three on FR-4, and one with solder volume above the norm. Test vehicles tested under condition C were manufactured at another facility under different manufacturing conditions.

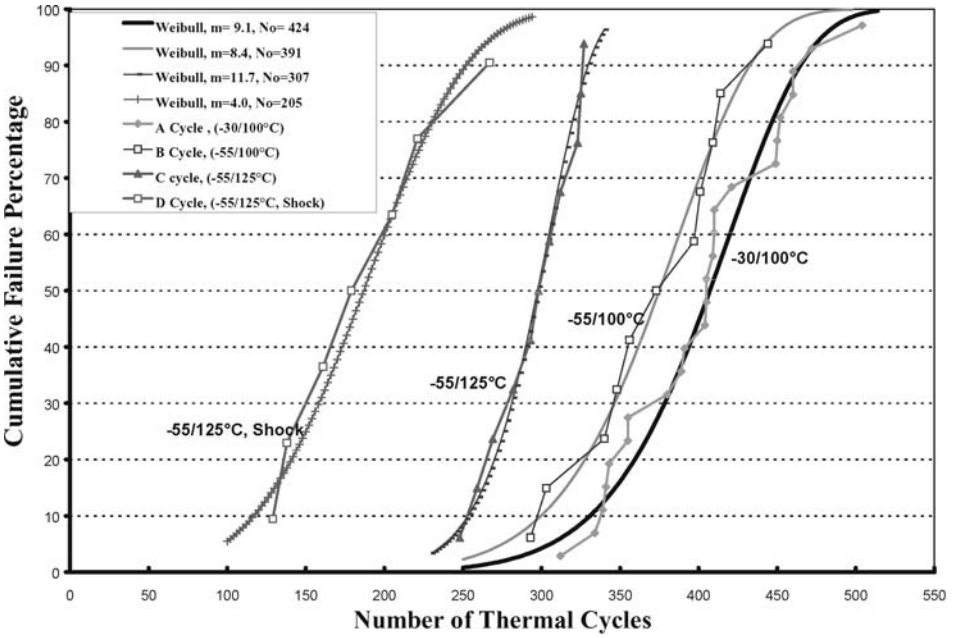


FIGURE 8.10. Cycles-to-failure for CBGA 625 assemblies under different cycling conditions.

TABLE 8.4.
Weibull parameters for CBGA 625 under different thermal cycling conditions.

Cycling condition	Weibull N_o ($N_{50\%}$)	Weibull m	Comments
A (−30/100°C)	424 (407)	9.1	23 test vehicles (TVs), Includes FR-4, polyimide PWB, and 3 solder volumes
B (−55/100°C)	391 (373)	8.4	11 TVs, FR-4 and polyimide
C (−55/125°C)	307 (298)	11.7	11 TVs, different assembler
D (−55/125°C, near thermal shock)	205 (189)	4.0	7 TVs, 3 on FR-4, 3 on polyimide, one with high solder volume

8.5.4.2.3. Coffin-Manson Relationship In the Coffin-Manson relationship, CTF is inversely proportional to creep strain. Its modified version includes the effects of frequency as well as the maximum temperature in the form of

$$(N_1/N_2) \propto (\Delta\gamma_2/\Delta\gamma_1)^{\beta} (f_1/f_2)^{\kappa} \exp[1414(1/T_1 - 1/T_2)].$$

- N_1 and N_2 represent cycles to failure under two plastic strain conditions. β is the fatigue exponential and is generally assumed to be equal to 1.9 [5].
- $\Delta\gamma$ is proportional to $(DNP/h)\Delta\alpha\Delta T$, where DNP is the distance from the neutral point at the center of package, h is equal to the solder joint height, $\Delta\alpha$ is the difference in the coefficient of thermal expansion of the package and PWB, and ΔT is the cycling temperature range.
- f_1 and f_2 are fatigue frequencies. κ is the frequency exponential varying from 0 to 1, with value 0 for no frequency effect and 1 for the maximum effect depending

on the materials and testing conditions. A value equal to 1/3 is commonly used to extrapolate the laboratory accelerated thermal cycles-to-failure data with short duration (high frequency) to on/off field operating cycle with long duration (low frequency), i.e., a shorter field cycles-to-failure projection.

- T_1 and T_2 are absolute maximum temperatures under the two cycling conditions.

The above relationship was used to correlate the test results for CBGA625 for the above four thermal cycling conditions. Cycles to 50% failure data are shown in Figure 8.11. Projections to lower temperature ranges from two data points and a single data point using the above relationship were also plotted. Projections were compared to a test result for the same package thermal cycled in the range of 0° to 100°C by IBM (Martin, et al., SMTA-International, 1997). The following key points were considered:

- Effect of frequency. The effect of frequency was considered to be minor since frequency ranges in our study were very similar and their dwells at maximum temperatures were at least 10 minutes. Creep was assumed to be almost complete for the two maximum cycling temperatures of 100°C and 125°C. These temperatures are well above the creep temperature of the eutectic Pb/Sn solder.
- Projection from C and B (−55°/125°C and −55°/100°C) to lower cycling ranges. The projection lines were calculated using the Coffin-Manson relationship. These include projection with and without considering the effect of maximum temperature. As shown, projections with a correction for maximum temperature better approximate the test results under conditions C and A as well as an IBM data point.
- Projection from near-thermal shock data CTF under thermal shock conditions were much lower than those under low heating/cooling rate cycling. Therefore, projec-

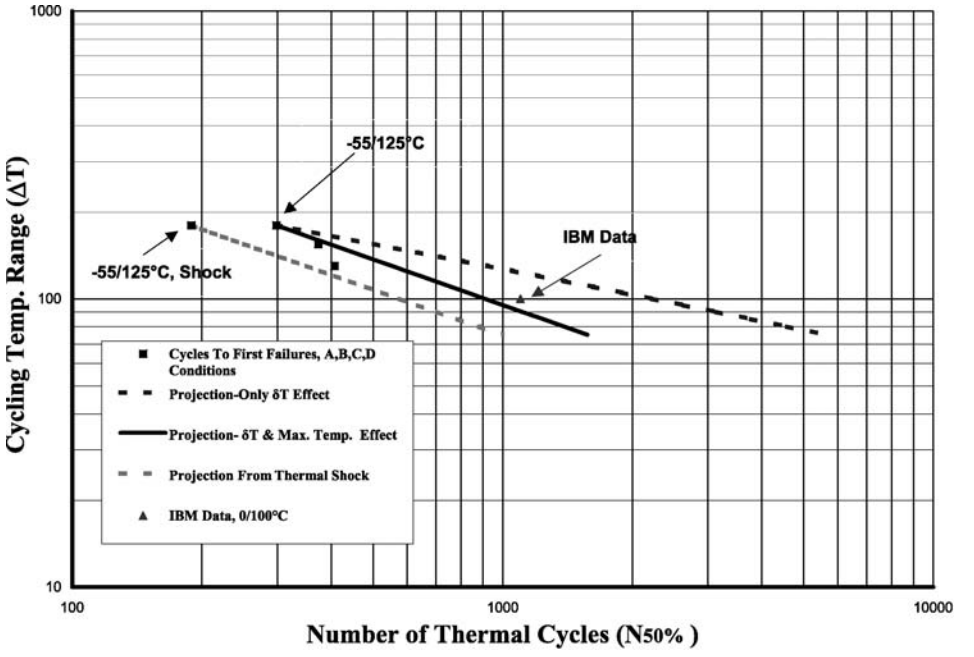


FIGURE 8.11. Effect of temperature range on fifty percentile failure cycles for CBGA 625 and projection using a modified Coffin-Manson relationship.

tion to other temperature ranges provides conservative failure cycles. Characteristic changes in solder behavior under rapid heating and cooling as well as failure mechanism changes for thermal shock conditions may not realistically represent most field applications.

- Projection from −55°/100°C and −30°/100°C to lower temperatures. This projection is not shown in Figure 8.11, but it is conservative since the slope between the two data points is much steeper than the slope connecting the data sets for −55°/125°C and for −55°/100°C cycles. This means that damage induced by fatigue due to a decrease in the cold regime is not as severe as creep induced due to the temperature increase at hot temperatures.

8.5.5. Comparison of 560 I/O PBGA and CCGA assembly reliability

This section includes thermal cycle test results for assemblies that included the 560 I/O ceramic column grid array and its plastic ball grid array counterpart with the identical peripheral package configuration. A designed experiment was utilized to cover many aspects that are considered to be unique for the potential use of these packages for NASA systems. Solder joint reliability is affected by many variables as briefly discussed in the previous section. The following parameters were either characterized or evaluated as part of the DOE implementation.

- Two pad designs, one for the CCGA and smaller pads for PBGA attachment. PBGAs were assembled on both pad sizes to evaluate PBGA interchangeability with the CCGA.
- Two stencil designs, a relatively thicker mini-stencil especially designed for CCGAs and a standard stencil to assemble PBGAs and other surface mount packages. Solder paste print volumes were measured and their variation shown in graphs.
- CCGA and PBGA assemblies without and with corner stake adhesive bonds. Corner adhesive bonds are used to improve resistance to mechanical vibration and shock loads.
- Added heat straps to the top of PBGAs to determine bonding attachment durability of the heat strap.

The assemblies were subjected to three types of thermal cycles. The process and results for CCGA package assemblies are discussed and compared to their PBGA counterparts in the following.

8.5.5.1. Test Vehicles Polyimide PWBs were designed to accommodate two pads configuration, one for the PBGAs and the other for the CCGAs. The pad size for PBGA was 24 mils, whereas for CCGA was 33 mils attached with traces to plated-through-hole, (PTH) vias with 24 mil diameter. Specific pairs of pads were connected through PTHs and these connections within a package pair completed a daisy chain to be used for monitoring. Four daisy chains for each package were used for continuous monitoring. Four additional pads were added at each side of the package for manual probing and failure identification to a narrower region after failure detection thorough continuous monitoring.

Similarly, two daisy chain sets—each with three rows using PTH vias representative of the test design—were added to monitor the behavior of the PTHs during thermal cycling. Continuity of these PTH daisy chains was performed manually at each interval when the assemblies were removed from the chamber for inspection. The PWB pads had HASL (Hot Air Solder Level) surface finish. HASL and OSP (Organic Solderability Preservative)

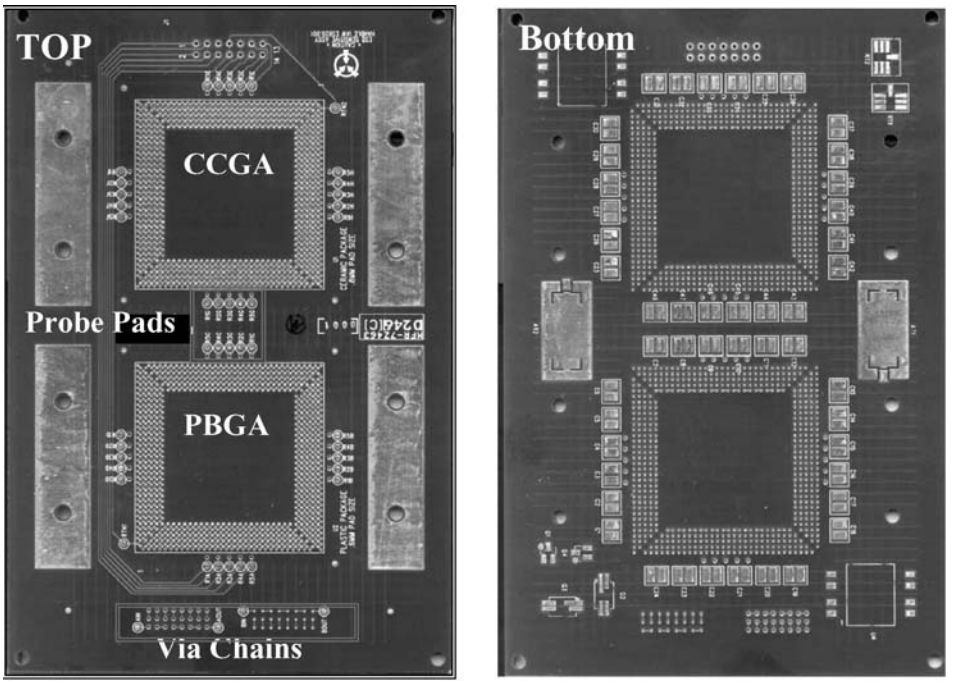


FIGURE 8.12. Printed wiring board design showing package daisy chain, probing pads, and via daisy chains.

surface finishes are specified in IPC-9701 as the recommended surface finishes for solder attachment reliability evaluation. The key reason for such recommendation is to avoid potential premature intermetallic failures such as those occasionally observed for the ENIG (Electrolytically Plated Ni/Au) surface finish. Figure 8.12 shows the top and bottom of the board design showing the daisy chain configurations for the packages, probing pads, and PTH vias.

8.5.5.2. Stencil Design, Paste Print, and Volume Measurement In a previous study, only the 8 mil thick stencil thickness was used to assemble both CCGA and surface mount assemblies. However, to achieve optimum solder paste volume for each part assembly, two stencil types with two different thicknesses were used to meet two solder volume requirements. Much higher solder volume was recommended to be used for CCGAs by the package supplier. Table 8.5 lists solder paste volumes that can be achieved with different stencil thicknesses and aperture openings. The 7 mil stencil thickness represents the general stencil that could be used for paste application on PBGA and other package pad patterns. The mini-stencil with 10.5 mil thickness was used only for the manual paste print application in order to achieve the higher paste volume recommended by the CCGA package supplier. Previously, it was shown that a higher solder paste volume could improve the reliability of CBGAs (ceramic ball grid array) and CCGAs with 1.27 mm pitch.

An RMA paste, type III $(-325 + 500)$ mesh paste was used for paste printing using automatic and normal manufacturing parameter setup for the case of the 7 mil stencil thickness. Manual paste printing was performed when the mini-stencil was used. Each paste print on the PWB was visually inspected after printing for gross defects such as bridging or insufficient paste. Paste print quality was improved when needed by adding solder

paste when insufficient or removing bridges for bridging conditions. Solder paste heights were measured using a laser profilometer. Measurement was carried out at 16 locations—as well as at corner and center pads—to accumulate solder volume data and their distributions.

Figure 8.13 shows plots of solder volume distribution for two different stencil thicknesses (7 and 10.5 mil) and pad opening for the 7 mil stencil thickness. Note the solder paste height measurement was done relative to the PWB surface rather than the Cu (copper) pad when made manually; therefore, the heights are a Cu thickness higher than their actual values. No adjustments were made when the solder volume was calculated based on the height and the pad diameter. It is apparent that the mini-stencil produced a wider distribution in the solder paste volume that can be achieved through automatic printing. Improvement in the distribution was improved slightly after a 2nd manual printing, but the distributions are still much wider than in the automatic version.

TABLE 8.5.
Stencil parameter and solder volume.

Stencil thickness	Solder volume $(mil)^3$	Option
BGA-Aperture 23 mil-stencil 7 mil	2909	Stencil
BGA-Aperture 24 mil-stencil 7 mil	3168	No
CCGA-Aperture 32 mil dia-stencil 7 mil	5632	Stencil
CGA-Aperture 33 mil dia-stencil 7 mil	5990	No
CCGA-Aperture 32 mil dia-stencil 8 mil	6430	Stencil (Previous study)
CCGA-Aperture 32 mil dia-stencil 10.5 mil	8440	Mini stencil

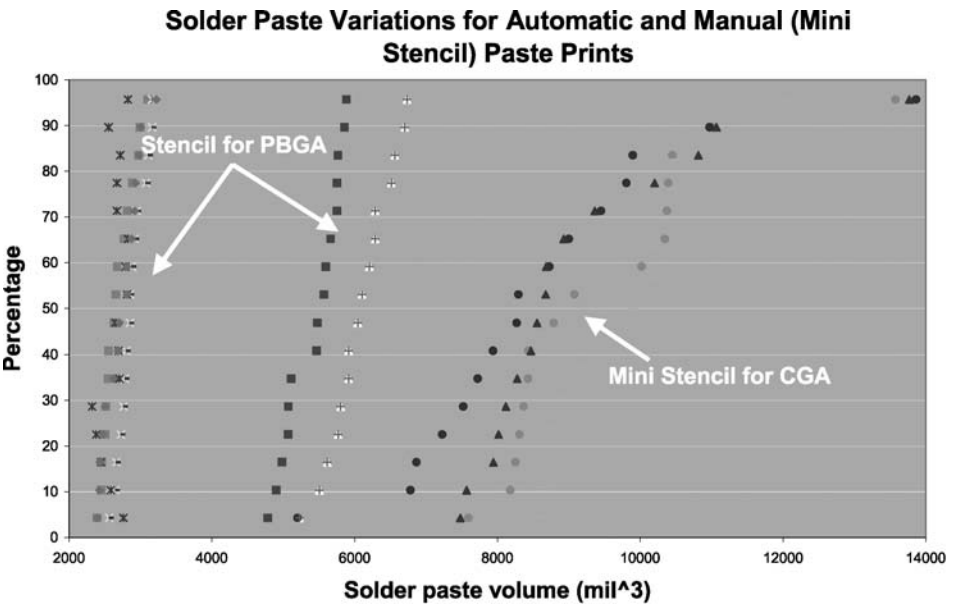


FIGURE 8.13. Paste volume variations and the effect of automatic and manual using mini stencil for printing.

8.5.6. Designed Experiment for Assembly

The objectives and parameters considered in the designed experiment were discussed previously. Packages are 42.5 by 42.5 mm in dimension; they are peripheral array with 5 rows and 1.27 mm pitch. The CCGA has a 0.2 mm ceramic interposer between the package and the column and a gap of 0.1 mm between the interposer and the package. The Solder column is a high melt solder with a diameter of 0.89 mm and a height of 1.62 mm. The plastic package with a heat spreader has eutectic tin/lead solder balls 0.75 mm in diameter.

8.5.6.1. Inspection before Environmental Tests For high reliability electronic applications, visual inspection is traditionally performed by Quality Assurance personnel at various levels of the packaging and assembly. Solder joints are inspected and accepted or rejected based on specific sets of requirements. Further assurance is gained by subsequent short-time environmental exposure, by thermal cycling, vibration, and mechanical shock, and so forth. These screening tests also allow detection of anomalies due to workmanship defects or design flaws at the system level. For NASA systems, generally 100% visual inspection is performed at prepackage prior to its closure (precap) and after assembly prior to shipment.

Visual inspection provides some usefulness for the area array packages, but obviously is of no value for the hidden balls and columns under the package. X-ray inspection is needed for area array packages. However in the case of CCGAs, the hidden solder joint could not be distinguished because of the heavy ceramic lid that inhibited X-ray penetration [21]. Visual inspection has a higher value for CBGA and CCGA assemblies since generally the solder fails at the exposed corners or periphery ball attachments. Peripheral balls and columns were inspected visually using an optical microscope at the start and during thermal cycling to document damage progress. Figure 8.14 shows photomicrographs of solder joints of PBGA and CCGA assemblies prior to thermal cycling.

8.5.6.2. Thermal Cycle Test An industry-wide guideline document, IPC SM-785, for accelerated reliability testing of solder attachment has been in existence around for more than a decade. Only recently, industry has agreed to release an industry-wide specification,

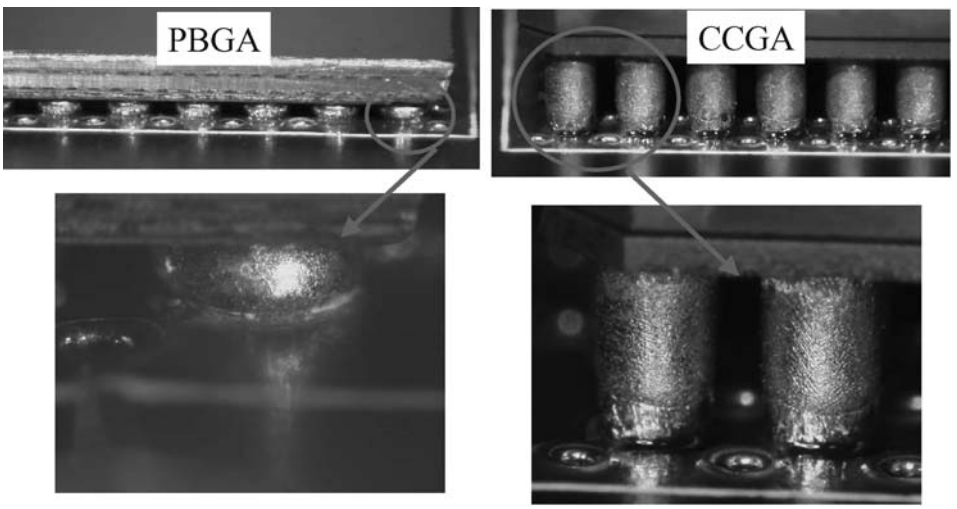


FIGURE 8.14. Optical photomicrographs of PBGA and CCGA after assembly.

IPC-9701, in response to BGA and CSP technology implementation [7]. Although very valuable and still valid, the IPC SM-785 guideline did not answer the key question of what the data means in terms of product application and data comparison. As is well established by industry, many variables could be manipulated to either favor or disfavor test results.

Also, in some cases, considerable resources and time could be wasted to generate failure data not related to solder attachment. An example is the use of a surface finish having the potential of inducing intermetallic rather than solder joint failure. This mishap is especially to likely to occur by a novice user/supplier.

The IPC-9701 specification addresses how thermal expansion mismatch between the package and the PWB affects solder joint reliability. In order to be able to compare solder joint reliability for different package technologies, PWB materials (e.g., FR-4) using a relatively larger nominal control thickness to minimize bending (0.093 inch). Surface finish choices to eliminate intermetallic failure (OSP, HASL), pad configuration to eliminate failure due to stress riser (non-solder mask defined or NSMD), and pad sizes having a realistic failure opportunity for package/PWB (80%-100 package pad), and so forth were standardized in order to minimize their effects on the test results.

The thermal cycle (TC) test ranges, test profile, and the number of thermal cycles (NTC) to be reported were also standardized. These include the reference cycle in the range of 0° to 100° C (TC1) and a severe military cycle condition of −55° to 125°C (TC4). Three out of five total TC conditions are identical to the test conditions recommended by JEDEC 22 Method A104, Revision A. The NTC varied from a minimum value of 200 cycles to a reference value of 6,000 cycles.

8.5.6.3. Test Results Figure 8.15 shows optical photomicrographs of a CCGA assembly after 200 B condition (−55°/100°C) thermal cycles. No intermittent failure has yet been observed, but the solder joints show significant damage, especially the solder joints for the corner columns. As stated previously, these test vehicles were assembled with an 8 mil stencil rather than thicker mini-stencil used in subsequent characterizations. Reductions in solder volume and graininess appearance are other features of these solder joints. This is clearly evident from the optical photomicrographs at 200 cycles. Figure 8.16 shows scanning electron microscopy of the solder and cross-section. Significant cracking—both at the interposer interface and the board solder interconnection—is apparent.

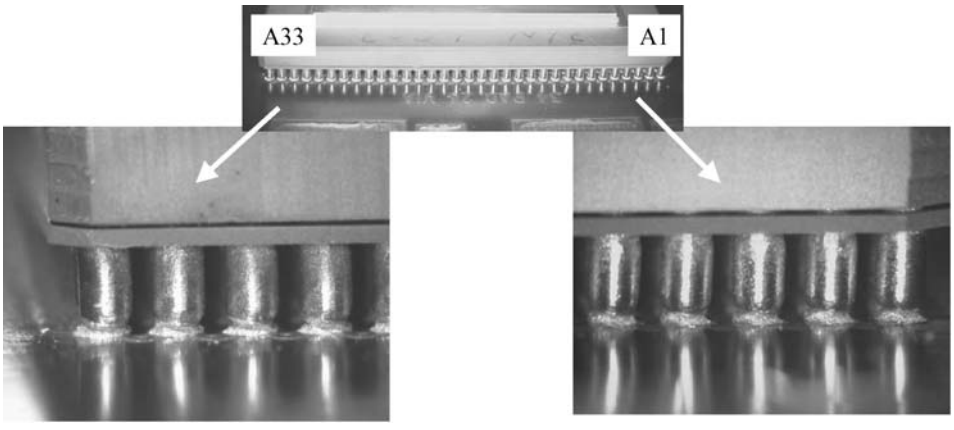


FIGURE 8.15. Optical photomicrographs of CCGA built with 8 mil thick stencil after 200 cycles (−55°/100°C). Note graininess and solder volume reduction as well as solder column shifts at corners.

Figure 8.17 shows optical photomicrographs of both PBGAs and CCGAs after 2,937 (−50°/75°C) and 478 (−55°/125°C) thermal cycles. There is no apparent degradation of the solder joints in the case of both the PBGAs and the CCGAs for the E condition where the maximum temperature was only 75°C. Even though most of the CCGA assemblies had failed by this cycle; however, the board solder joint interconnects appear to be similar to the pristine condition shown previously in Figure 8.14. The failures were from the package side within the package and interposer where it is not visually apparent and cannot be easily inspected. The solder joints at the board interface showed insufficient solder with graininess for the case where the maximum temperature was 125°C. The feature changes with cycling for the C condition are similar to the F condition with a 100°C maximum temperature. The PBGA balls exposed to a similar maximum temperature condition did not show this reduction in solder volume.

Figure 8.18 compares optical to SEM photomicrographs of a PBGA ball after exposure to 1819 cycles (−50°/75°C) and another one subjected to 588 cycles in the range of −55°/125°C. Both assemblies had their corners staked. There were no significant microstructural changes for the former E condition, whereas a small microcrack was initiated in the solder joint at the package interface for the C condition (−55°/125°C). The latter photo also clearly shows the grain growth due to exposure to elevated temperature. Similar optical photomicrographs for CCGA assemblies with corner staking after the same number of cycles and conditions (1,819 cycles/−55°/75°C and 588 cycles/−55°/125°C) are shown in Figure 8.19. Note that even though both assemblies had corner staking, the failure mechanisms were different. One failed away from staking, whereas the other (−55°/125°C) failed within the staking adhesives at the interposer solder interconnection interfaces.

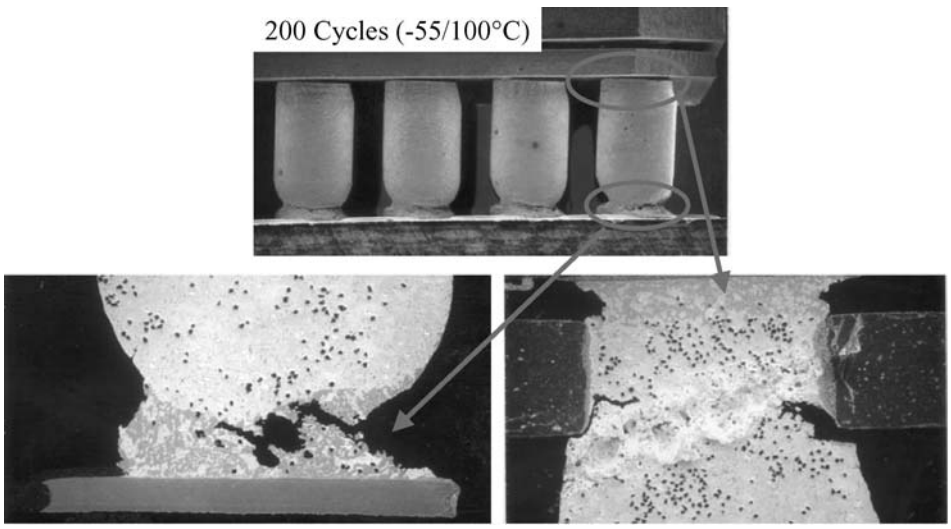


FIGURE 8.16. SEM photomicrographs of CCGA built with 8 mil thick stencil after 200 cycles (−55°/100°C). Note cracking from board solder joint and interposer interface at package side.

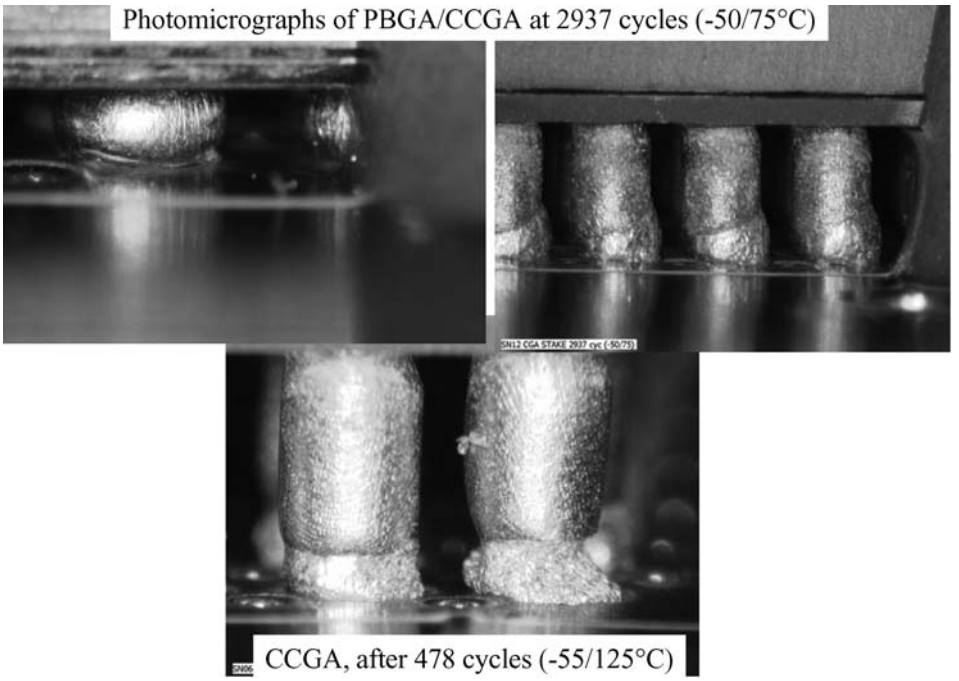


FIGURE 8.17. Optical photomicrographs of PBGA and CCGA under two different thermal cycle conditions. Note graininess and solder volume reduction for CCGA exposed at 125°C. CCGAs are built with a 10.5 mil stencil thickness.

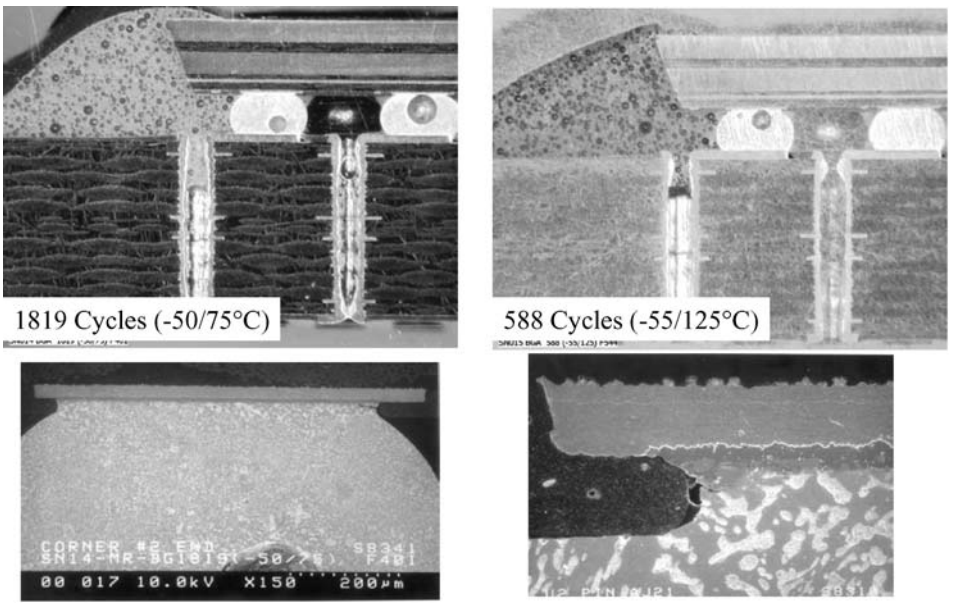


FIGURE 8.18. SEM photomicrographs before and after cross-section for PBGA package after 588 cycles (−55°/125°C).

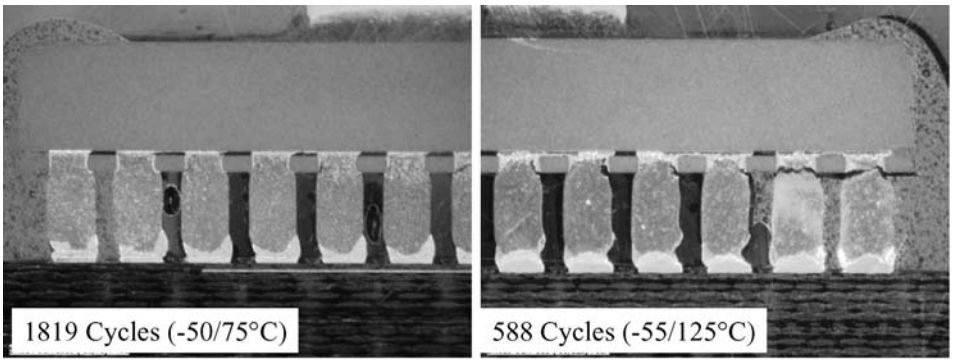


FIGURE 8.19. Photomicrographs of column failure at interposer with corner staked over two thermal cycle ranges.

8.6. SUMMARY

The key issues for BGAs, CSPs, and ceramic packages and their assembly are as follows:

- Extremely limited flight heritage data are available for these styles of packages because area array technologies have only recently been considered for use in NASA systems.
- Plastic BGAs and CSPs withstand fewer 2nd level (assembly) thermal cycles than their leaded counterparts, e.g., QFPs and TSOPs. In general, CSPs have lower life cycles than plastic BGAs when mounted on polymeric boards.
- Ceramic BGAs have 5 to 10 times lower assembly reliability than plastic packages when assembled on polymeric boards. Ceramic column arrays have 2 to 3 times better reliability than their BGA versions.
- New ceramic column packages have been developed to improve 2nd level assembly reliability by using higher CTE ceramic materials and using improved attachment processes such as interposers. Projected improvements may be optimistic and may not be correlated with the test results based on limited data published on this subject.
- When in doubt about 2nd level reliability of area array packages or when the application requires withstanding a large number of thermal cycles, it is advisable to perform testing using dummy daisy chain packages.
- After a decade of progress in this field, there is still no good non-destructive inspection technique for detecting interconnection cracks. In addition, even if workmanship defects are detected, the individual defects cannot be reworked. Therefore, strict process control is the key to successful implementation of area array technologies. This requires good manufacturing procedures and well-established discipline on the manufacturing floor.
- Most area array packages are built for commercial applications, and many issues with COTS plastic and ceramic BGA and CSP packages are similar to those encountered with conventional COTS microcircuits.
- It has been recently shown that 560 I/O plastic package assemblies did not show failures to 2,000 cycles, whereas 560 I/O CCGA assemblies showed the first failure at slightly above 1,000 cycles when they were subjected to the −50°/75°C cycle. The solder ball attachment for the plastic package version, however, did not meet

the package level die burn-in requirements generally performed at 125°C for space applications.

- Recent investigation indicates that CCGA solder joints at the board interface showed signs of graininess and reduction in volume when they were exposed to thermal cycling with a maximum temperature of either 100°C or 125°C. These changes were not observed for the PBGA solder joints or when CCGA assemblies were exposed to a 75°C maximum temperature.
- The effect of corner staking on assembly reliability has only been recently characterized. The PBGA assemblies with the corner stake adhesive showed no additional degradation or changes in failure mechanism compare to those without corner staking when subjected to thermal cycling to 2,000 cycles.
- The CCGA assembly with the corner stake; however, showed changes in failure mechanism. For the −55°/125°C cycle condition, failures were at the interposer interfaces of those columns that were covered on by the adhesive. For the −50°/75°C cycle, failures were still at the interposer. However, they were away from those covered by adhesive.

8.7. LIST OF ACRONYMS AND SYMBOLS

Au	Gold
BGA	Ball Grid Array
CBGA	Ceramic Ball Grid Array
CCA	Circuit Card Assembly
CCGA	Ceramic Column Grid Array
CGA	Column Grid Array
COTS	Commercial-off-the-shelf
CSP	Chip Scale (Size) Package, a.k.a. Fine Pitch BGA (FPBGA)
CTE	Coefficient of Thermal Expansion
CTF	Cycles to failure
ENIG	Electroless Nickel Immersion Gold
FCBGA	Flip Chip Ball Grid Array
FEA	Finite Element Analysis
FEM	Finite Element Method
FPBGA	Fine Pitch BGA, a.k.a. Chip Scale Package (CSP)
FR-4	Epoxy/fiberglass laminate material for PWB construction
HASL	Hot Air Solder Level
I/O	Input/Output
IPC	Currently the acronym has no meaning. It now refers to an international trade association of the interconnection electronic industries. Formerly, it meant Institute of Printed Circuits.
KGD	Known Good Die
LCCC	Leadless Ceramic Chip Carrier
MER	Mars Exploration Rover
Mo	Molybdenum
NEPP	NASA Electronic Parts and Packaging
Ni	Nickel

NSMD	Non-Solder Mask Defined
NTC	Number of Thermal Cycles
OSP	Organic Solderability Preservative
Pb	Lead
PBGA	Plastic Ball Grid Array
PTH	Plated-through Hole
PWA	Printed Wiring Assembly
PWB	Printed Wiring Board
QFP	Quad Flat Pack
SMC	Surface Mount Component
SMT	Surface Mount Technology
Sn	Tin
TC	Thermal Cycle
TCE	Also CTE, Thermal Coefficient of Expansion
TSOP	Thin Small Outline Package

ACKNOWLEDGMENTS

The research presented in this chapter was conducted at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not constitute or imply its endorsement by the United States Government or the Jet Propulsion Laboratory, California Institute of Technology.

The author extends his appreciation to the JPL-led consortia team members and many others at JPL who contributed to the progress of these activities during many years of research in this area. Special thanks to Dr. Namsoo Kim at Boeing and Emmanuel Siméus and Steve Stegura at Raytheon. Acknowledgment also goes to Rick Fry and his team at CMC electronics for the design of one of the board, Also, Atul Mehta, Ana Rosa Arreola, Ken Evans, and Dr. Kirk Bonner at JPL are thanked for their support in test vehicle assembly, thermal cycling, failure analysis, and review.

The author extends his appreciation to the project manager, Eric Schwartzbaum, and his team, Tom Jedrey, Dr. Hadi Mojaradi, Ann Devereaux, and Dorothy Stosic, for partial funding support and their technical contributions. The continuous support of Phillip Barela, mission assurance manager, is also appreciated. Sincere thanks to program mangers at NASA Electronic Parts and Packaging Program (NEPP) including Michael Sampson, Dr. Charles Barnes, and Phillip Zulueta for their continuous support and encouragement.

REFERENCES

1. R. Ghaffarian, N. Kim, D. Rose, B. Hunter, K. Devitt, and T. Long, Rapid qualification of CSP assemblies by increase of ramp rates and cycling temperature ranges, The Proceedings of Surface Mount International, Chicago, Sept. 30–Oct. 4, 2001.
2. R. Ghaffarian, G. Nelson, M. Cooper, D. Lam, S. Strudler, A. Umdekar, K. Selk, B. Bjorndahl, and R. Duprey, Thermal cycling test results of CSP and RF package assemblies, The Proceedings of Surface Mount International, Chicago, Sept. 25–28, 2000.
3. J. Fjelstad, R. Ghaffarian, and Y.G. Kim, Chip Scale Packaging for Modern Electronics, Electrochemical Publications, 2002.

4. R. Ghaffarian, Chip scale package assembly reliability, in K. Puttlitz and P. Totta, Eds., Area Array Interconnect Handbook, Kluwer Academic Publishers, 2002, Chapter 23.
5. R. Ghaffarian, BGA assembly reliability, Chapter 20, in K. Gilleo, Ed., Area Array Packaging Handbook, McGraw-Hill Publisher. [date?]
6. R. Ghaffarian, Characterization and failure analyses of lead-free solder alloy defects, Chapter 16, in D. Shangguan, Ed., Lead-Free Solder Interconnect Reliability 2006, to be published by ASM International.
7. IPC-9701, Performance test methods and qualification requirements for surface mount solder attachments, Published by IPC, Association Connecting Electronics Industries.
8. A. Mawer and L. Luquette, Interconnect reliability of ball grid array and direct chip attach, IRPS 1997 Tutorial, 1997.
9. http://www.amkor.com.
10. http://www.altera.com.
11. http://www.xilinx.com.
12. A.C. Shiah and X. Zhou, A low cost reliability assessment for double-sided mirror-imaged flip chip BGA assemblies, The Proceedings of Pan Pacific Conference, Surface Mount Technology Association, 2002.
13. J.W. Evans, J.Y. Evans, R. Ghaffarian, A. Mawer, K. Lee, and C. Shin, Monte Carlo simulation of BGA failure distributions for virtual qualification, ASME, Hawaii, 1991.
14. K. Newman, M. Freda, H. Ito, N. Yama, and E. Nakanishi, Enhancements in 175 FPBGA board-level solder joint reliability through package constructions, The Proceedings of the 6th Pan Pacific Microelectronics Symposium, Kauai, HI, Jan. 2001.
15. R.N. Master and O.T. Ong, Ceramic grid array technologies for ACPI applications, The Proceedings of Surface Mount International, Chicago, Sept. 25–28, 2000.
16. M. Farooq, L. Goldmann, G. Martin, C. Goldsmith, and C. Bergeron, Thermo-mechanical fatigue reliability of Pb-free ceramic ball grid arrays: experimental data and lifetime prediction mounting, The Proceeding of IEEE Electronic Components and Technology, 2003, pp. 827–833.
17. Z. Burnette, et al., Underfilled BGAs for ceramic BGA packages and board-level reliability, The Proceeding of IEEE Electronic Components and Technology, 2000, pp. 1221–1226.
18. R.N. Master, T.P. Dobear, M.S. Cole, and G.B. Martin, Ceramic ball grid array for AMD K6 microprocessors applications, Proc. Components and Technology Conference, Seattle, Washington, 1998.
19. S.Y. Teng and M. Brillhart, Reliability assessment of a hight CTE CBGA for high availability systems, Proc. Components and Technology Conference, 2002.
20. M. Interrante, et al., Lead-free package interconnection for ceramic grid arrays, IEEE/CPMT/SEMI Int's Electronics Manufacturing Technology Symposium, 2003.
21. R. Ghaffarian, Comparison of X-ray inspection systems for BGA/CCGA quality assurance and crack detection, IPC SMEMA Council APEX Confererence, 2003.

9

Metallurgical Factors Behind the Reliability of High-Density Lead-Free Interconnections

Toni T. Mattila, Tomi T. Laurila, and Jorma K. Kivilahti

Department of Electrical and Communications Engineering, Helsinki University of Technology, P.O. Box 3000, Otakaari 7 A, 02015 TKK, Finland

9.1. INTRODUCTION

Product reliability is an important factor especially in portable electronics, because these increasingly powerful and more complex electronic equipment experience different kinds of electrical, thermal, mechanical, and thermo-mechanical strains and stresses in their service environments. The importance of solder interconnection reliability is increased mainly due to two reasons: Firstly, higher interconnection densities, e.g., in small-scale Ball Grid Array, Chip Scale Packaged or Flip Chip components, are related to decreasing solder interconnection volumes (see Figure 9.1). Decreased size of solder interconnections has brought the components closer to the printed wiring boards (PWB) and therefore stresses experienced by these micro-interconnections are considerably increased. Furthermore, due to the small solder volumes there is a risk that too large fractions of solder interconnections will transform into brittle intermetallic compounds [1,2]. Secondly, the employment of lead-free solders, components under bump or lead metalizations, and PWB protective coatings add to the complexity of the interconnection metallurgies. The number of different material combinations increases markedly as the traditionally used SnPb-base solders and protective coatings are replaced with different lead-free alternatives. Solders are replaced with alloys such as Sn3.8Ag0.7Cu, Sn3.5Ag3.0Bi, Sn3.5Ag, Sn0.7Cu, Sn58Bi, or Sn10Zn [3–10]. PWB coatings are replaced with Organic Solderability Preservatives (OSP), matte Sn, electrochemical Ni(P) with a thin flash Au on top (Ni(P)|Au), Ni|Pd|Au, Ag or Bi, for instance [11,12]. The most common choices for the component metalization seems to be matte Sn, Ag|Pd, Ni|Au, Ag, Bi, or Ni(V)|Cu [12].

Because microstructures ultimately control reliability of soldered interconnections, reliability of each material combination is likely to be different. Therefore, it is of primary importance to investigate systematically metallurgical reactions in the effective joining region and resulting microstructures within the solder interconnections as well as to study their impacts on reliability with test samples assembled as in volume surface mount pro-

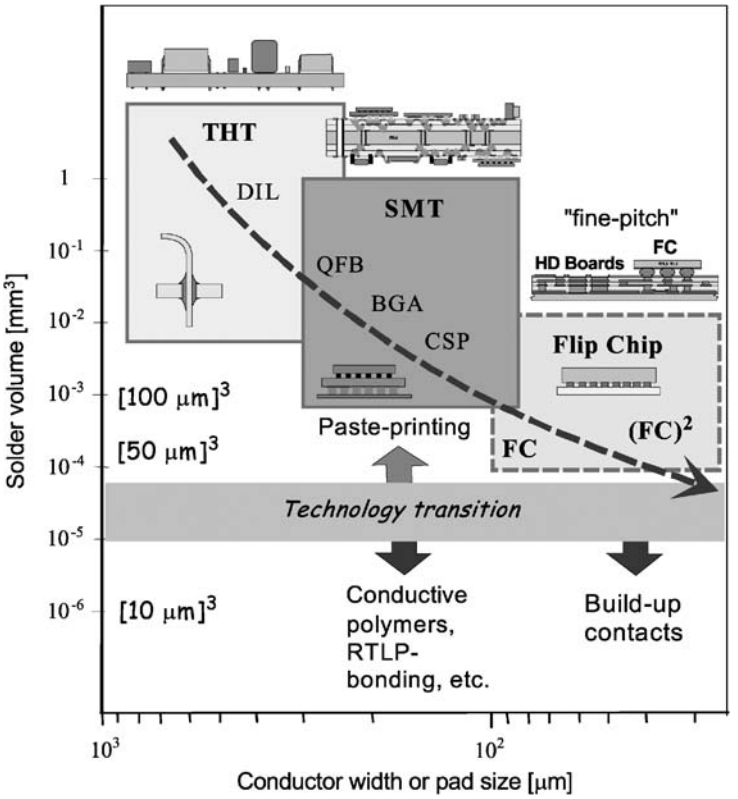


FIGURE 9.1. Impact of miniaturization on electronics production.

duction. The effective joining region can be defined as the region where concentrations of the components differ from those of the original contact materials [13,14]. Lead-free solder interconnections can contain more complex intermaterial layers such as phosphides, which weaken solder interconnections. Therefore it is anticipated that mechanical failures will be encountered more frequently in the future, especially if the fine-pitch components with metallurgically incompatible component metalizations or PWB coatings are introduced into electronic products. Moreover, the microstructures formed during soldering are not stable and will evolve during the operation of products. Hence, in order to ensure the best possible reliability of electronic assemblies against various loading conditions, much better understanding of interconnections microstructures and their evolution, including interfacial reaction products, is needed.

Portable electronic equipment experience during their use many different kinds of loading conditions, in which mechanical shocks and thermomechanical loadings are perhaps the most critical ones. As we shall see later on in this chapter, different loading conditions will evoke different failure mechanisms leading to dissimilar failure modes. Under thermomechanical loading nucleation and propagation of cracks is controlled by the microstructures formed during soldering and their recrystallization behavior during use [15–19]. Mechanical shock impacts, on the other hand, are known to produce entirely different kinds of failure modes [20,21]. Cracks in the newly soldered interconnections do not propagate through the bulk solder of the interconnections, but mainly in the brittle

intermetallic compound layers formed between solder and contact metalizations [20–23]. However, portable products are seldom dropped soon after they are assembled; most likely their components have been exposed to elevated temperatures or experienced thermomechanical loadings, and therefore the microstructures of the solder interconnections have also evolved. Dense continuous networks of grain boundaries may have already been produced by recrystallization and so the above-mentioned two failure modes are mixed when the products are dropped.

What has been stated above increases also the importance of reliability testing. This, in turn, causes significant expenses to companies owing to the increased testing time as well as adds to the costs following from longer time-to-market cycles. Therefore, the employment of proper test procedures for different applications is becoming increasingly important. To ensure the feasibility of the test results, a better understanding of the failure mechanisms occurring under different loading conditions and their correlation with the field environments must be obtained. Therefore, the employment of different simulation and statistical methods together with detailed microstructural studies are needed.

In this chapter we will concentrate on the evolution of microstructures during accelerated reliability tests, since they eventually determine the failure mechanisms that control reliability of electronic products. The focus is in identifying the factors driving the microstructural evolution in lead-free interconnections and the effects of different testing conditions, in particular those of mechanical shock loading and thermal cycling, because they are considered the most relevant tests for portable products. However, we will begin with a presentation describing the iterative approach adopted in the Laboratory of Electronics Production Technology at the Helsinki University of Technology to emphasize that reliability of solder interconnections is not only a metallurgical issue but we have to incorporate electrical, mechanical, thermal, and statistical aspects into our scope as well.

9.2. APPROACHES AND METHODS

The iterative approach for studying the reliability of soldered assemblies consists of four major steps: (i) design and manufacturing of test assemblies, (ii) reliability testing and simulation of the devices under test, (iii) statistical analysis of the reliability test results, (iv) non-destructive inspection and detailed (destructive) microstructural characterization of failed interconnections. The role of simulation and modeling tools in this process cannot be overemphasized, since most of the major steps involve to some extent either electrical, thermal, mechanical, statistical, or materials modeling.

9.2.1. The Four Steps of The Iterative Approach

The approach illustrated in Figure 9.2 establishes an iterative loop in which the designs are constantly improved on the basis of simulation and experimental results. In this subchapter some of the steps involved in the reliability studies are introduced. It should be noted, however, that the extent of design and simulation tools needed depends on the type of test chosen. For example, in the thermomechanical and mechanical shock tests discussed later in this chapter electrical simulation has a minor role, whereas thermal, mechanical, and microstructural simulations play a more central part. Electrical and thermal simulations, on the other hand, are vital tools in studies on power cycling. Because different kinds of simulation tools are used in different stages of the iteration loop, modeling and simulation are discussed collectively in Section 9.2.2.

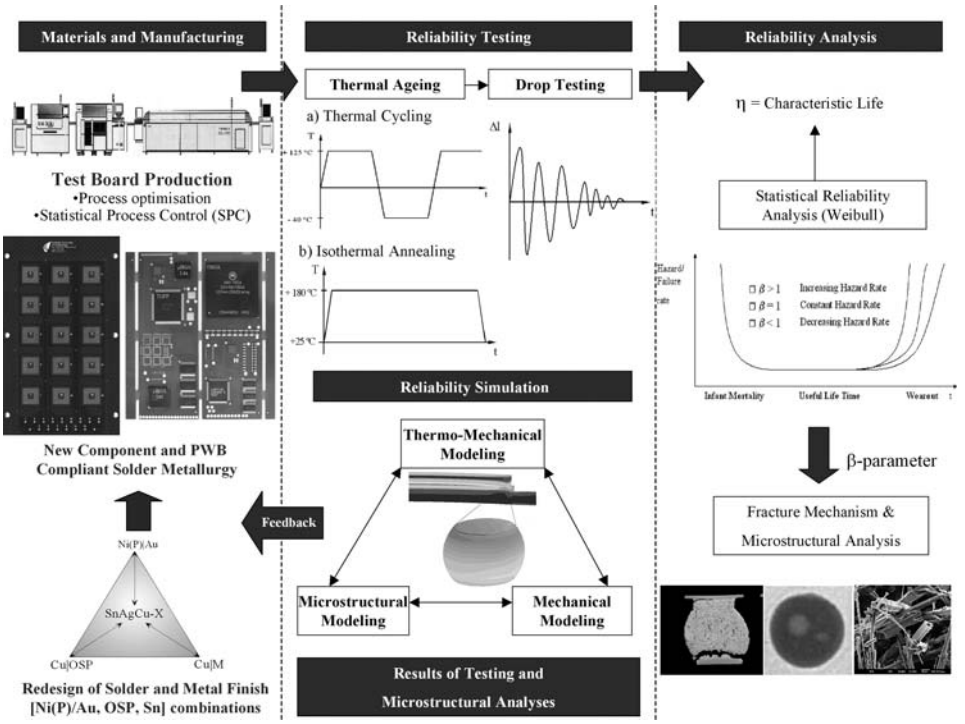


FIGURE 9.2. A schematic presentation of the reliability research procedure [1].

9.2.1.1. Design and Manufacturing of Test Assemblies After the electrical design of test boards, their assembly is carried out on a production-scale reflow soldering and/or wave soldering lines. It is, naturally, most important that the test assemblies are free of production related defects and therefore all the process steps—solder paste printing, component placement and, soldering—are continuously inspected and assured that they operate within their (statistically determined) control limits. However, even electrically conductive and visually acceptable interconnections can be unreliable. Inspection alone does not guarantee good reliability. Different parameters of the soldering profile, such as heating rate, activation time and temperature, time above liquidus, peak temperature, and cooling rate, all affect the quality of solder interconnections. Processing parameters of soldering determine the microstructures formed during soldering. These structures are the onset of the microstructural evolution that takes place either during the subsequent accelerated testing or service of devices. Hence, at this stage the combined thermal and thermomechanical simulations are utilized to get a better understanding and control of the solidification during reflow soldering. An example of such a simulation is presented in the next section.

9.2.1.2. Reliability Testing Reliability testing follows the after-reflow inspection of the test boards. Thermal cycling tests are used to study the effect of thermomechanical stresses generated by heat dissipating elements or changes in ambient temperature on the reliability of electronic equipment. Strains and stresses are produced by differences in coefficient of thermal expansion of dissimilar materials. Thermal cycling tests are typically carried out in air-to-air or fluid-to-fluid test chambers depending on the required rate of temperature change. The profile of the test is determined by values of upper and lower temperatures, dwell time, rate of change between the temperatures, and number of cycles (see Figure 9.3).

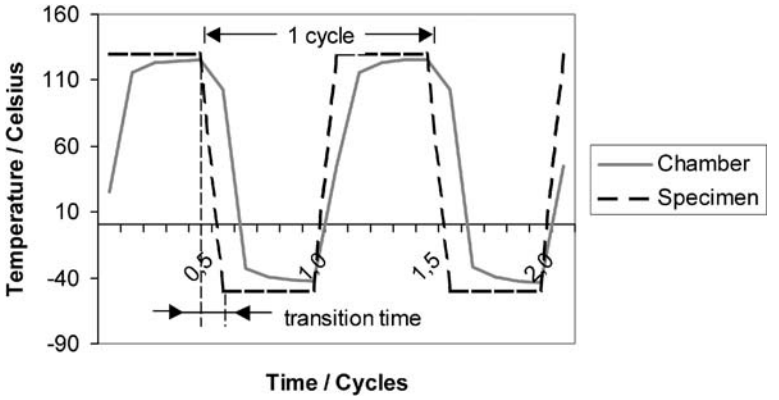


FIGURE 9.3. Typical temperature profile used in thermal cycling tests.

Since even the minimum temperatures of typical thermal cycling profiles remains above the 0.3–0.4 of the absolute melting range of almost all lead-free solder alloys, dwell time at temperature extremes allow creep processes to transform elastic strain into plastic strain. But because deformation during thermal cycling is largely plastic a change in the dwell time does not affect the severity of the test considerably. It should, however, be long enough for the temperature of the specimen to stabilize.

Because portable electronic products are more likely to be dropped than affected by changes in thermal conditions, the emphasis of the reliability research has moved, during the recent years, from thermomechanical testing to mechanical shock testing. Reliability testing of solder interconnections on printed wiring boards during high impact drop loading is studied utilizing different kinds of drop testers specially designed for this purpose. Different organizations are preparing standardized tests for mechanical shock loading of portable electronic equipment. JEDEC has recently published its own JESD22-B111 standard for handheld electronic products [24].

Drop test apparatuses are generally composed of a mechanism to drop the board repeatedly in a specified orientation and a high-speed data acquisition system to record the deacceleration condition of the test, strains on the component boards, and the drops-to-failure. The test board is typically attached to a fixture from four corners or the edges of the board. The fixture is mounted on a sledge that is dropped down on a rigid surface from a specified height in a controlled manner with the help of guiding rails. Dropping can be performed in different orientations, but it is often performed horizontally, components facing downwards. This is because the most detrimental factor for the assembly caused by dropping is not the mechanical shock itself, but the subsequent bending and vibration of the board [20,21]. Placing the component boards horizontally achieves maximum flexure of the board and onsets the natural vibrating motion. Bending causes displacement between the board and the components resulting in component, interconnection, or board failures.

The shape of the optimal deceleration pulse according to the JESD22-B111 standard is a half-sine with 0.5 ms width and G level of 1500 (see Figure 9.4), or 0.3 ms width and 2900 Gs [24]. The shape of the pulse is not only a function of the drop height but depends also on the characteristics of the strike surface: drop height determines the maximum deacceleration and the strike surface the pulse width. Real time components daisy chain measurement for failure identification requires a high-speed data acquisition system.

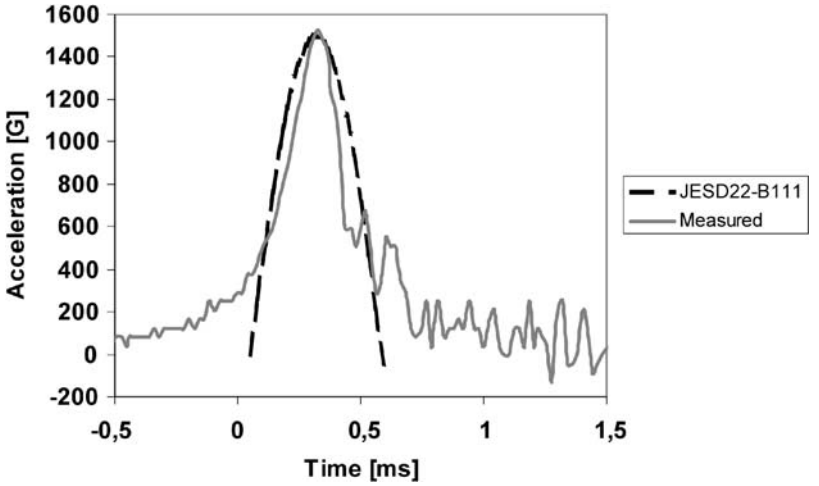


FIGURE 9.4. Strain according to the JESD22-B111 standard and that measured in the longitudinal direction of the test board [see board layout in Figure 9.18(b)].

The system should be able to detect any discontinuity of daisy chain resistance lasting for few microseconds or even less. Deacceleration of the test vehicle and strains on the component boards are also measured. Strain measurements are performed by very light strain gages attached in different parts of the test assembly in different directions. The test environment is highly accelerated because the supporting effect of the product covers and other such components in the drop impact of a functional product is neglected in such board level tests as the JESD22-B111. Furthermore, because the behavior of the test board is strongly dependent on the test board construction, dimensions, and materials, the drop test performance should be studied utilizing a standardized PWB construction. Only one type of component is used at a time. The purpose of the tests is to evaluate the reliability performance of the most common surface mounted components used in handheld electronic products and therefore component types such as ball grid arrays, land grid arrays, (wafer-level) chip scale packages, or small outline packages are used. Standardized drop tests are used mainly to compare the reliability performance of different material combinations, component structures and other product design related issues.

High deformation rate inherent in drop tests increases the strength of the solder interconnections. As the result, the brittle intermetallic layers between the solder and the metalizations become more prone to failure as the stresses in these layers increase compared to the stresses in thermal cycling tests for instance. This issue will be discussed in more detail in the following case studies.

9.2.1.3. Statistical Analysis of Reliability Test Results The amount of numerical data gathered during reliability tests is usually extensive. Therefore the use of statistical tools is indispensable. Statistical hypothesis testing is needed for making decisions whether there are statistically significant differences in reliability between certain types of assemblies. Reliability of the solder interconnections is often studied also by making use of the statistical Weibull reliability analysis. The purpose of the Weibull method is to characterize the failure distribution and to make inferences about the failure mode in operation.

In order to statistically compare and test the affects of more than one factor on the reliability, experiments have to be designed in the right manner. The factorial experiments allow examination of several factors as well as their interactions, and to determine whether the observed differences in the response i.e., the value of the measured reliability characteristic (times-to-failure or drops-to-failure) are statistically significant [25–28]. The choice of statistical method for testing hypotheses is dependent on how its failure distribution conforms to normal distribution. The test for normality can be carried out by the Shapiro-Wilk Test for instance [29]. If the data acquired conforms to normal distribution the results from the experiment can be analyzed with the Analysis of Variance (ANOVA). If the time-to-failure data fails to conform to normal distribution nonparametric methods must be used to carry out the statistical testing. Wilcoxon Rank-Sum Test procedure is perhaps the most widely used test for such purposes [26]. Tests are typically carried out at less that 5% risk level, which means that if the resulting p-value of the test is smaller that stated, it is a good indication to reject the null hypothesis and conclude that the two means are different.

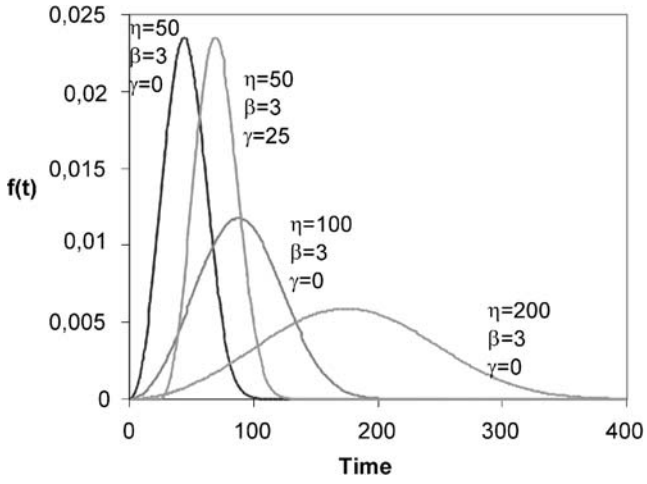
Large numbers of different statistical models are available for modeling time-to-failure data. The Weibull distribution is one of the most widely used lifetime distributions in reliability engineering due to its versatility. It relates the reliability data to a failure mechanism and it can also be used with relatively small sample sizes. Depending on the values of the parameters, it can be used to model a variety of life behaviors. There are two different forms of the Weibull distributions, the two-parameter and three-parameter distribution. The two-parameter cumulative distribution is characterized by the characteristic life (η) and the shape parameter (β). The third parameter, γ, is called the failure free life. The choice of the distribution depends on the fit of the test data to the distribution in question. The three-parameter Weibull distribution can sometimes give a better fit to the data. When the γ is included in the distribution, it takes the form:

$$F(t) = 1 - \exp\left[-\left(\frac{t-\gamma}{\eta}\right)^{\beta}\right].$$

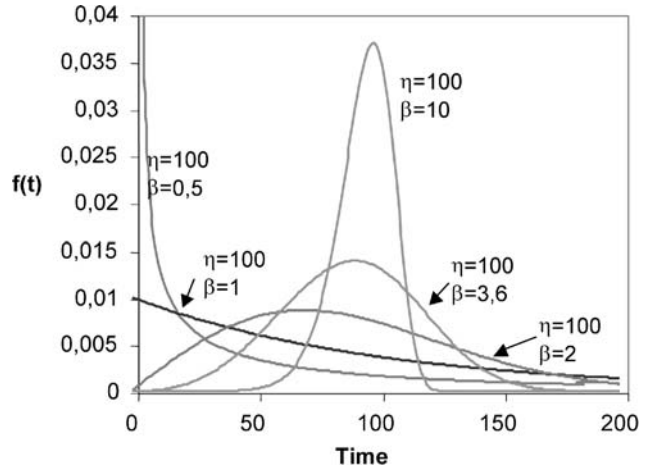
With different parameters, the function takes a variety of shapes as shown in Figure 9.5(a). A change in the parameter γ changes the time scale without changing the shape of the distribution. When η is increased while keeping β constant [see Figure 9.5(a)], the probability density function stretches out and decreases in height because the area under the density function is a constant value of one. The effect of the β on the probability distribution is illustrated in Figure 9.5(b).

The value of β represents a certain failure rate and failures can be classified by its value to the three life-stages of the bathtub curve [30]. The relationship between the Weibull shape parameter and the bathtub curve is presented in Figure 9.6 [see also Figure 9.5(b)]. When the β value is less than unity (decreasing failure rate), the plot represents early or "infant mortality" failures. If β equals one (constant failure rate) the plot represents the intrinsic failures during the product lifespan. Values greater than one (increasing failure rate) represent the wear-out failures.

There are many methods available for the Weibull parameter estimation such as the probability paper plotting, the least squares estimation, and maximum likelihood estimation. The least squares method is a mathematical version of the probability plotting and therefore provides more objective parameter estimates. The method of least squares is often chosen instead of the maximum likelihood estimation due to its relative simplicity. The parameters are estimated with the least squares regression in the following manner:



(a)



(b)

FIGURE 9.5. Weibull density functions with different values of: (a) η and γ (β constant), (b) β (η and γ constants).

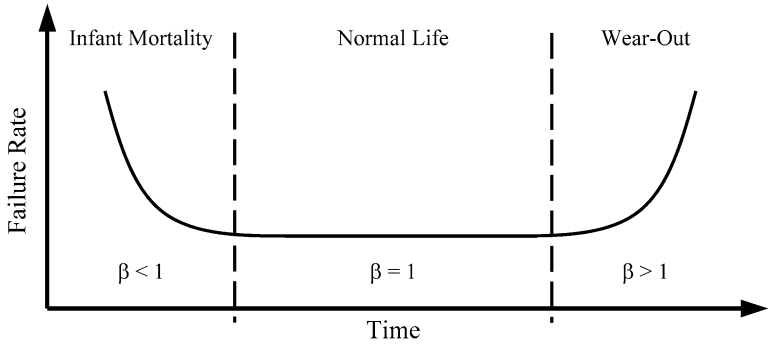


FIGURE 9.6. Relationship between the Weibull shape parameter β and the bathtub curve.

a straight line is fitted to a scatter plot ($y = \log\ln[1/(1 - F(t))] = \beta \log_{10}(t) - \beta \log_{10}(\eta)$ is linear as a function of $\log_{10}(t)$). The slope of the regression equals to β while the intercept on the y-axis equals to $-\beta \log(\eta)$, where η can be calculated. The η can also be read from the diagram at the 63.2% cumulative failure rate.

9.2.1.4. Failure Analysis and Microstructural Characterization The next step is the comprehensive failure analyses of the tested samples. Non-destructive inspection of the test samples, which provides initial information about the failure modes, is carried out first. This is usually done with the help of x-ray inspection or scanning acoustic microscopy. Detailed destructive analyses of cross-sectional samples needed for studying microstructures are carried out with optical and scanning electron microscopy. When even higher resolutions are needed the transmission microscopy will be used. The microscopes are equipped with either energy or wave dispersive spectrometers for obtaining local chemical analyses from the samples.

9.2.2. The Role of Different Simulation Tools in Reliability Engineering

Simulation tools are of great importance in the iterative approach because they help us to rationalize the test results as well as to obtain better understanding about the test conditions. For example the finite element calculations (presented in more detail in Ref. [20]) on the effect of via-in-pads structure showed that the stresses on the component side of the interconnections with the via-in-pad structure was one fifth higher as compared to those without the vias. Thus the micro-via makes the PWB side more rigid and thereby increases the stresses in upper parts of the interconnections. In the following section some central aspects of thermal, mechanical and physical-chemical simulation will be introduced.

9.2.2.1. Thermal Simulation of Reflow Soldered Components As an example of the use of simulation in the manufacturing stage, we will briefly consider a case discussed in more details elsewhere [31,32]. The goal of the work was to study the solidification of solder interconnections of the chip scale packaged components in a commercial forced-convection reflow oven with the help of thermal simulation tools. The simulation work was carried out in three consecutive steps: (i) thermal modeling of reflow oven, (ii) thermal modeling of the component and, (iii) thermal modeling of solder interconnections.

The first step was performed with the help of the Computational Fluid Dynamics technique and the other two steps were based on the Finite Element Method. Before the simulation, temperature profiles and oven temperature were measured by using multichannel concurrent data logger. The measured results were adopted to optimize the thermal model of the reflow oven. Results from the first step were used as boundary conditions for the model in the next step. Thermal properties of SnAgCu solder were presented with the help of thermodynamic analysis on equilibrium solidification procedure. Two fundamental assumptions were made: solidification of solder was treated as a nearly equilibrium process and that interconnection microstructures are homogeneous at the scale of the calculation mesh. Both experimental data and the results from the component level calculations indicated that the component is cooled faster than the board. However, the interconnection model revealed that temperature gradient over interconnection was not likely to be large because of high conductivity of solder. Noticeable temperature gradient inside interconnection only occurred at the final stage of solidification, when the isothermal eutectic reaction took place. It was also suggested that inner interconnections were subjected to more uniform temperature field and slower solidification than outer interconnections. This provides important information for further analysis on the mechanism of primary Sn growth.

9.2.2.2. Mechanical Simulation The Finite Element Method is used to evaluate the stress states that samples experience during testing and thereby to aid in the interpretation of the test results. The aim of stress analysis is to calculate where in the test board layout the most critical components are located, which interconnections are most prone to fail, and how the stresses are distributed in the component boards as well as in the interconnection areas. The calculations are typically carried out by utilizing the sub-modeling technique. It means that first the displacements of the whole board are calculated with a rather rough model of the whole assembly and then the results are used as boundary conditions to a model with more details included and only parts of the board modeled. Typically three geometric accuracy levels are used: (1) board level model, (2) component level model and, (3) interconnection level model.

Mechanical simulation tools are valuable especially in analyzing results from drop tests. As discussed above, the rapid loading during the drop makes the board bend and vibrate that ultimately makes the assemblies to fail. The bending of the component boards can be regarded as a sum of different vibration modes, which are natural deformation shapes of the structure. The board bends to each one of these natural modes with certain natural

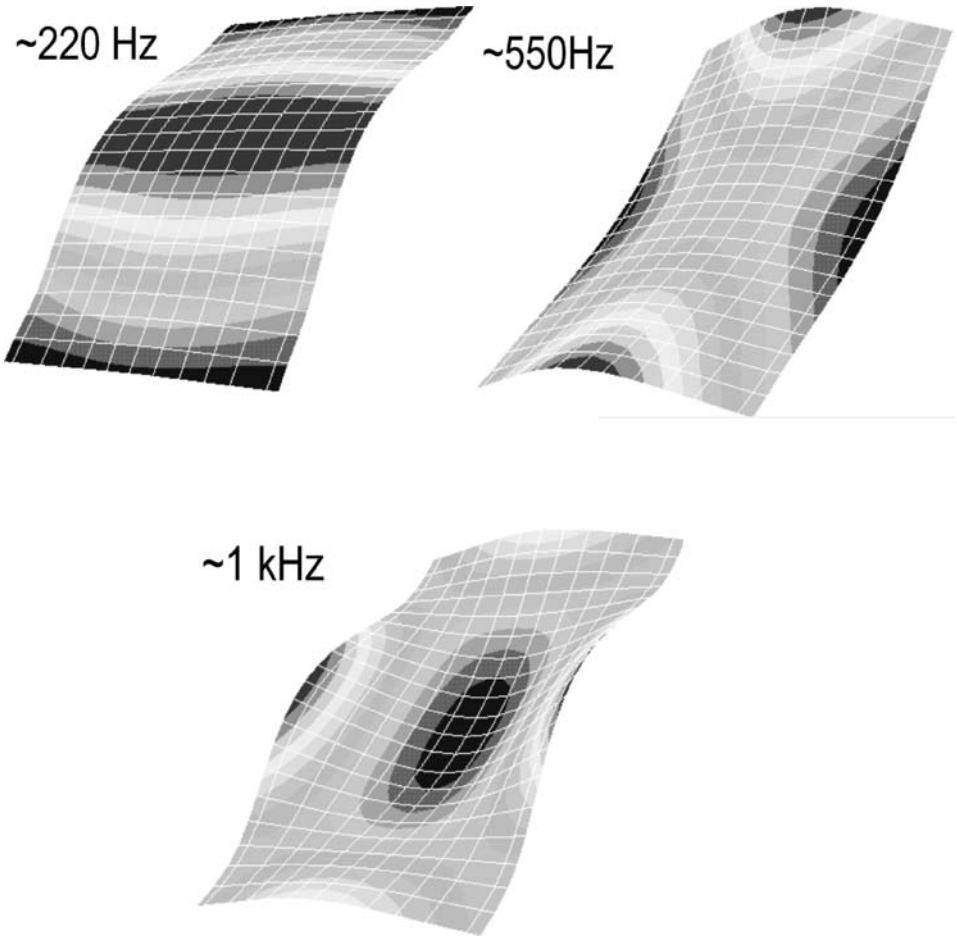


FIGURE 9.7. Three eigenforms of the drop test board (JESD22-B111) and their eigenfrequencies.

frequency that is characteristic to the mode. As the result the location of the maximum stresses changes rapidly and therefore the stress states may be rather difficult to comprehend intuitively.

In free vibration—as in the drop test—the rule of thumb is that the lower the frequency of the natural mode the greater the impact on the failure. The natural modes having higher frequencies have smaller amplitudes and they are damped away quicker. Figure 9.7 shows three of the most important forms of the JESD22-B111 test assembly. Figure 9.8 shows the longitudinal strain on the centre of the JESD22-B111 board (board layout shown in Figure 9.18). The strain is rather clearly a sinuous function of time. The first natural mode is responsible for the vibration where the maximum and minimum values are about 2.3 ms apart, while the others cause the smaller and faster oscillations in the strain history.

9.2.2.3. Simulation of Interconnection Materials As noted earlier, manufacturing reliable lead-free electronic products becomes even more challenging when solder interconnection volumes are decreased while the number of reactive metals is increased [1,2,33–35]. In addition, the microstructures formed during soldering are under continuous microstructural evolution during the use of the devices. Thermodynamic and kinetic modeling tools can help in determination of the potentially reliable solder, component metalization, and PWB protective coating combinations and thereby limit the amount of experimentations needed. Together with careful microstructural investigations these simulation tools can speed up the R&D work and testing of new products considerably.

Thermodynamics of materials provides fundamental information on the stabilities of phases (i.e., basically microstructures), the driving forces for chemical reactions and diffusion processes occurring in solder interconnections during processing, testing and in long-term use of electronic devices. Further, the thermodynamics provides us the phase diagrams that contain information also on metastable equilibria—usually not available in experimentally determined (stable) equilibrium diagrams. The phase equilibria in solder or solder-substrate systems—as in any system—are computed by summing up first all

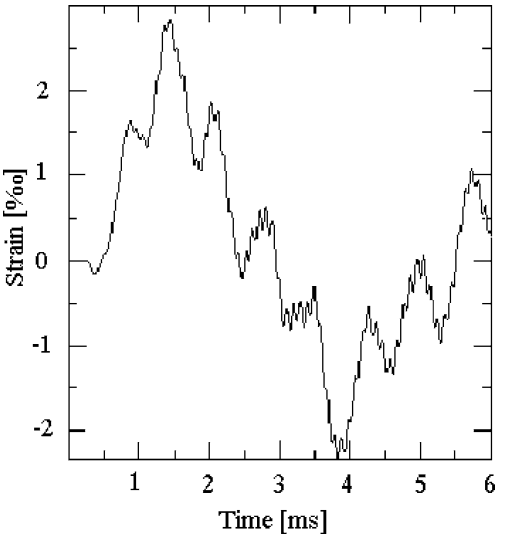


FIGURE 9.8. Longitudinal strain on the centre of the drop test board [JESD22-B111; see board layout in Figure 9.18(b)].

the Gibbs (free) energies of individual phases (i.e., solutions and compounds) and then minimizing—according to the second law of thermodynamics—at constant temperature and pressure the total Gibbs energy of the n-component system. Readers interested in the thermodynamic or diffusion kinetic modeling procedures and the calculation of phase diagrams are referred to vast amount of available literature, for example the review articles and books [36–38] to begin with.

Even though the complete phase equilibrium is practically hardly ever met in solder interconnections, the stable or metastable local equilibrium is, however, generally attained at interfaces between dissimilar materials (or phases) in contact with each other. Since the equilibrium is attained only at the interfaces there are activity gradients in the adjoining phases even though the chemical potential (or activity) of a component has the same value at the interface. These gradients determine the diffusion of components in various phases of an interconnection region. By making use of the fundamental condition that no atom can diffuse intrinsically against its own activity gradient as well as of the mass balance requirement it is possible to rule out impossible reaction product sequences [36].

It should be emphasized that phase diagrams do not contain any information on size, shape or distribution of the phases in a material system; calculated diagrams have to be clarified experimentally by employing different methods of microscopy. Furthermore, it is not possible to calculate ternary or multicomponent phase diagrams solely on the basis of the data from binary systems, since they do not include information about ternary (or higher order) interactions or ternary compounds which are not connected to any of the binary systems. Finally, it should be noted that even though thermodynamics provides the basis for analyzing reactions between different materials one cannot predict the time frame of the reactions on the basis of the phase diagrams. This is why diffusion kinetics must be included in the analysis.

9.3. INTERCONNECTION MICROSTRUCTURES AND THEIR EVOLUTION

It is important to know as much as possible about microstructures because they affect the failure mechanisms in operation. The initial microstructures of the solder interconnections are generated during solidification at the cooling stage of the soldering process. This structure establishes the starting point for the microstructural evolution that takes place during the field service of electronic devices.

9.3.1. Solidification

The majority of the lead-free solders are Sn-rich alloys with few major and minor additional elements. Therefore, the solidification behavior of solder interconnections is dominated by Sn. At the beginning of solidification, primary grains are formed and their morphology strongly affects the solidified microstructure. For example, in the case of SnAgCu solders, the primary crystals can be either β-Sn, Cu_6Sn_5 or Ag_3Sn, depending on the composition. During solidification, the primary crystals start forming wherever they find suitable places for nucleation. Even though solidification most likely starts from either the component or the PWB side interfaces, any oxide layer on top of molten solder interconnections or impurity particles in the interconnections may also act as suitable places for heterogeneous nucleation.

Usually, when a solder alloy solidifies, a cellular or dendrite structure is formed depending on the growth conditions. The most important factors affecting the solidified microstructure are solidification properties of the growing phases, temperature distribution

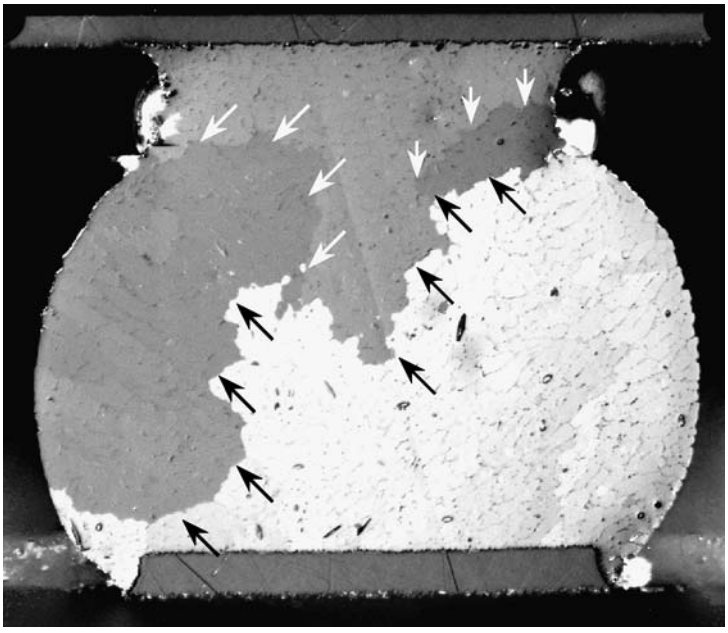
during solidification, and solute redistribution between liquid and solid during the cooling of an alloy. Readers interested in thorough treatment of solidification are directed to the literature [39,40]. The as-solidified microstructure in the Sn0.5Ag0.5Cu alloys used in our experiments presented in the two case examples below is a cellular structure of tin. It is noticeable that the interconnections seem to consist of only few colonies with different orientation (high angle boundaries between the colonies) when investigated with polarized light [15–19]. The areas themselves are composed of cells with angle boundary between them (see Figure 9.9).

Another interesting point is the behavior of minor elements during solidification. Since the last droplets of liquid that solidify are present at the high angle boundaries between the large colonies, this is also the site where most of the impurities should be located. In fact, in our investigations gold dissolved from PWB surface finish has been observed to enrich at the high angle boundaries as small needle-like $AuSn_4$ intermetallic particles [41]. What has been stated above indicates clearly that mechanical behavior of solder interconnections is most probably quite different from that of a "normal" polycrystalline material. For example, the grain boundary cracking should not occur in the as-solidified structure due to the absent of high angle boundaries (other than those between colonies). Therefore, when stress is applied to interconnections having this kind of microstructure, it undergoes microstructural evolution before fractures can propagate. A more detailed discussion on this issue is presented below in the context of the two examples.

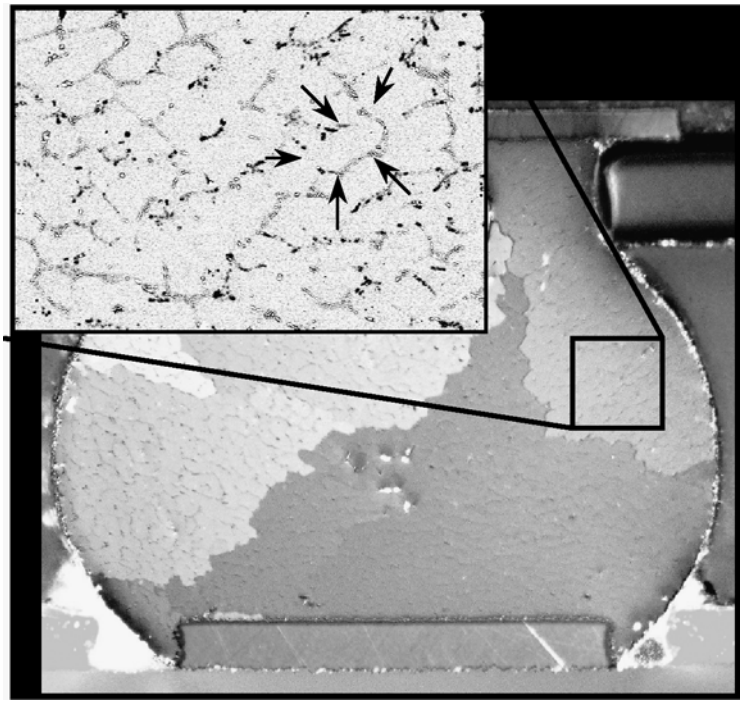
9.3.2. Solidification Structure and the Effect of Contact Metalization Dissolution

Under the reflow conditions typically used in lead-free soldering, solidification structure is generally cellular, where the small Cu_6Sn_5 and Ag_3Sn phases are dispersed between large primary Sn grains, as already discussed. If a protective Au metalization is used some small needle-like $AuSn_4$ can also be found inside the solder matrix at the high angle boundaries. An example of microstructure formed in the interconnections soldered with the Sn0.5Ag0.5Cu alloy on electrochemical Ni(P) with a thin flash Au on top (denoted Ni(P)|Au in the following) is shown in Figure 9.10. Both the Cu_6Sn_5 and the Ag_3Sn particles are uniformly distributed around the relatively large Sn cells. Figure 9.11 shows a micrograph taken from a sample soldered with the same alloy but this time on the boards with Organic Solderability Preservative (OSP) on the Cu pads (noted Cu|OSP). The resulting microstructure seems to be different even though the same solder alloy was used: relative to Ni(P)|Au, interconnections formed on the Cu|OSP contain more and larger Cu_6Sn_5 intermetallic particles dispersed inside the solder.

Why is the resulting microstructure different? We must take into consideration what happens during soldering. It is well known that PWB coatings and component metalizations in contact with the molten solder dissolve into the melt and thus the solder is alloyed further with the dissolving coatings and metalization. Too great dissolution can potentially degrade the performance of solder interconnections due to the subsequent impact on microstructures. Let us consider the differences in protective coating solidification between the Ni(P)|Au and Cu|OSP protective coatings and the implications in the microstructures. The thin layer of Au dissolves instantly and completely into the molten solder and the Ni metalization starts dissolving next. In the case of the Cu|OSP interconnections, the OSP partially evaporates and the rest dissolves into the solder flux during soldering and the Cu pad that starts dissolving into the solder alloy. The dissolution rate of Cu in Sn0.5Ag0.5Cu (wt%) is about 0.07 μm/s [15,42,43]. Based on this, the amount of Cu dissolution at the entire area of the soldering pad during the typical 40–45 second time above



(a)



(b)

FIGURE 9.9. Optical micrograph from cross-section of the interconnection taken with polarized light: (a) colonial boundaries indicated (high angle boundaries), (b) cell boundaries indicated (small angle boundaries).

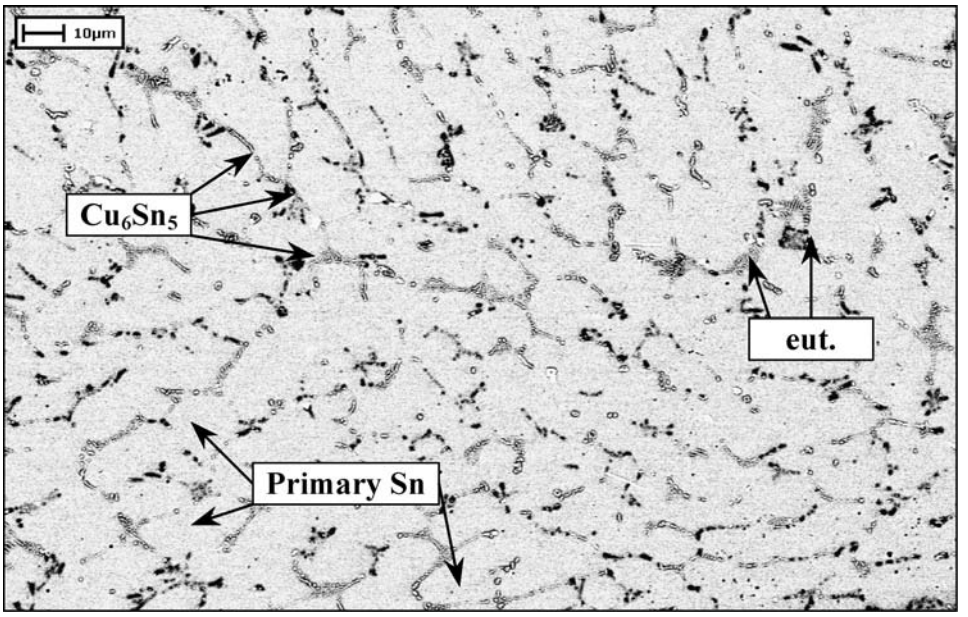


FIGURE 9.10. Microstructure of interconnections on Ni(P)|Au-coated soldering pads, where the nominal composition is Sn0.5Ag0.3Cu.

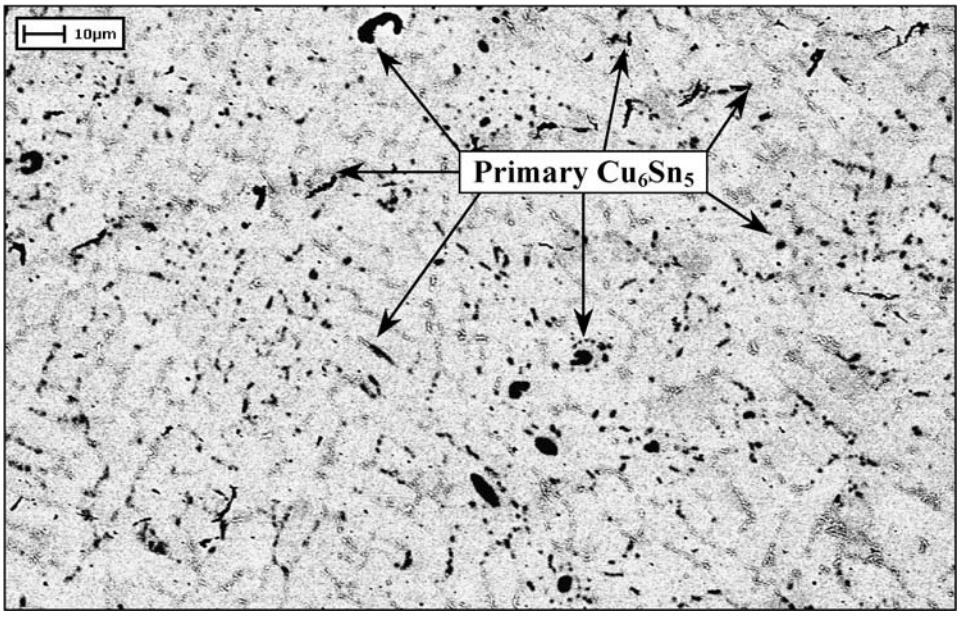


FIGURE 9.11. Microstructure of the interconnections on OSP-coated soldering pads, where the nominal composition is Sn0.5Ag1.0Cu.

217°C is enough to lift Cu concentration in the soldered interconnections close to 1 wt%, event when taking the amount of Cu bonded into the inter metallic layers on both sides of the interconnections into account. On the other hand, the dissolution rate of Ni is about 50 times smaller than that of Cu and thus, the dissolution of Ni to the solder is insignificant. All Ni that is dissolved at the interface is expected to be bonded to the $(Cu,Ni)_6Sn_5$ layer. Taking into account the amount of Cu bonded to the inter metallic layers on both sides of the interconnections, the nominal composition of the interconnections soldered on the Ni(P)|Au coated pads will result in about Sn0.5Ag0.3Cu where as the final composition on the interconnection on Cu was about Sn0.5Ag1.0Cu.

An important consequence of higher Cu content is that the solidification process is different between the interconnections soldered on Ni and those soldered on Cu. When considering solidification, it is very useful to first examine the solidification of the solder interconnections with the help of equilibrium phase diagrams. It should be noted, however, that the equilibrium diagrams do not contain information about either the distribution or the morphology of the phases, as already discussed. Figures 9.13 and 9.14 present the phase fraction diagrams, where the amount of different phases in relative number of moles can be presented as a function of temperature. The interconnections soldered on Ni(P)|Au PWB have the Sn0.5Ag0.3Cu composition, whereas the interconnections soldered on Cu|OSP have the Sn0.5Ag1.0Cu.

As can bee seen from Figure 9.12, the solidification of the liquid interconnections soldered on Ni(P)|Au boards starts with the formation of primary Sn phase when the interconnections are cooled down from the peak reflow temperature to below the liquidus temperature of 229°C. The Cu_6Sn_5 phase does not nucleate until below 222°C, where the composition of the liquid reaches the eutectic valley. Figure 9.13 presents the phase fraction diagram of the liquid interconnections soldered on Cu|OSP boards. In this case the solidification begins with the formation of primary Cu_6Sn_5 below 229°C. However, the nominal composition of the liquid soon meets the curve of two-time saturation, af-

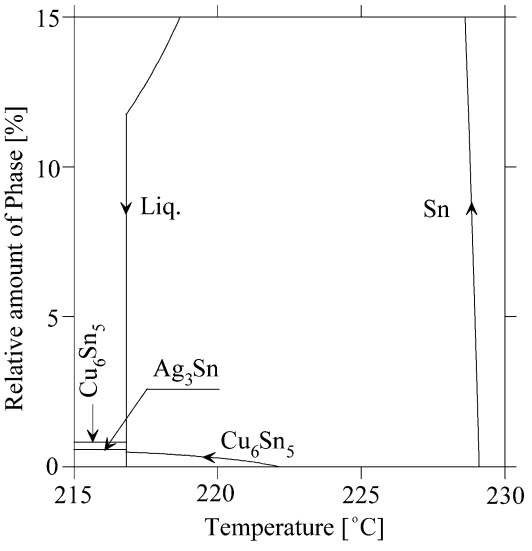


FIGURE 9.12. Phase fraction diagram of a system with nominal composition on the interconnection on the Ni(P)|Au.

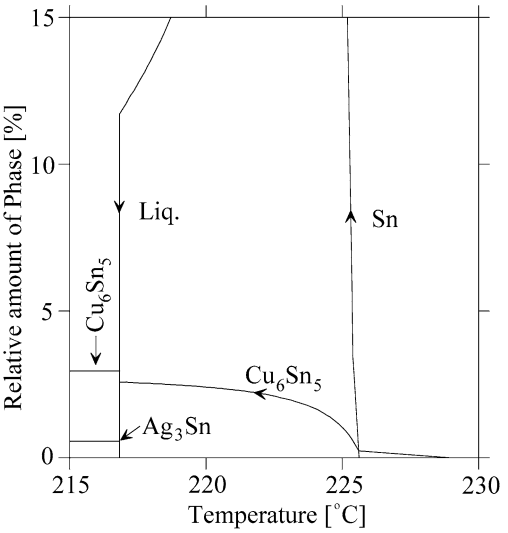


FIGURE 9.13. Phase fraction diagram of a system with nominal composition on the interconnection on the CuIOSP.

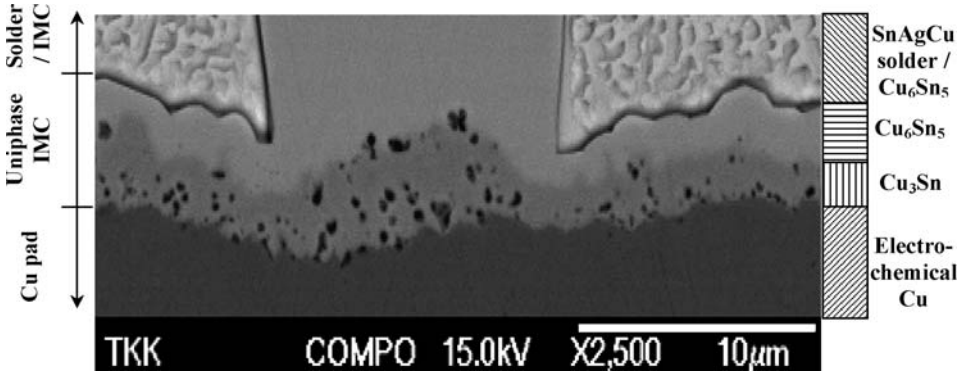


FIGURE 9.14. Electrolytic Cu/Sn diffusion couple annealed at 125°C for 1000 hours.

ter which the solidification of the interconnections proceeds by the binary eutectic reaction $L \rightarrow (Sn)eut + (Cu_6Sn_5)eut$. Below the four-phase invariant temperature, there is more than three times as much Cu_6Sn_5 in the CuIOSP interconnections as in those on the Ni(P)|Au substrate and this difference is clearly visible between the microstructures in Figures 9.10 and 9.11. It should be noted, however, that the cooling rates used in reflow processes are normally much faster than the rate assumed in equilibrium considerations. Higher cooling rates will eventually evoke marked under-cooling and more refined microstructures. Hence, in practice the solidification process always departs somewhat from that of equilibrium solidification.

9.3.3. Interfacial Reactions Products

In electronic products all the common base materials, coatings and metalizations, form intermetallic compounds (IMC) with Sn at the solder|conductor interface, and therefore these compounds must be treated as major elements of solder interconnections. A thin and continuous IMC layer(s) at the solder|conductor interfaces is an essential requirement for good wetting and bonding, and it produces distinct improvements in mechanical properties of interconnections. However, due to their inherent brittle nature too thick IMC layers may degrade the reliability of the solder interconnections. Thus, knowledge of the solder|conductor interaction and phase evolution in the solder interconnections is important not only to understand the reliability issues of the solder interconnections but also to the optimization of the soldering process from the metallurgical standpoint. The importance of knowing the properties of the IMCs in solder interconnections is clearly demonstrated in the reliability tests carried out under fast deformation rates.

Even though there are some characteristic differences between the systems met in soldering applications, the intermetallic reaction layers are formed, in principle, in three consecutive steps: the dissolution, chemical reaction, and solidification, although, the relative importance of each stage varies between the systems depending on the solubility of conductor metal in Sn [33,44]. The general sequence of events during a soldering operation can be described as follows. Immediately after the flux has removed the oxides and permits metallurgical contact of solder with the conductor metal, the contacted metal starts to dissolve into the molten solder. Initially the rate of dissolution is very high, particularly if the solder is not alloyed with the metal in question and therefore very high concentrations of solute elements can be realized locally. After a short period of time, the layer of molten solder adjacent to the contacted metal becomes supersaturated with the dissolved metal throughout the interface. At the local (metastable) equilibrium solubility, the solid IMC starts to form in this part of the interconnection. The formation of the IMC takes metal solutes out of the saturated liquid solder and causes some further dissolution of the contacted metal, especially if the intermetallic layer is not uniform on top of the substrate. Generally, after this stage the intermetallic reaction structure consists of two parts. Next to the base metal there is a relatively thin "uniphase" layer and on top of that sometimes quite thick irregular two-phase (or solder + IMC) layer. The thickness ratio of the two parts varies strongly between different systems. What is of particularly interest is that the thickness of the two-phase layer (solder + IMC) seems to increase with increasing equilibrium (stable and metastable) solubilities. During storage or in use of the assemblies, the IMCs generated during soldering grow further in thickness or increase in number, especially if the operational temperatures are well above the room temperature. Therefore, both solid|liquid and solid|solid systems must be studied to have better understanding of the reliability of soldered assemblies.

It should be noted that the local equilibrium conditions in the solder interconnections will change locally owing to the consumption of one or more of the components. This may then change the phases that can exist in local equilibrium accordingly and result in new interconnection microstructures. For a detailed discussion about the formation of IMC's in both solid|solid and solid|liquid reaction couples readers are pointed to recent review article [44].

9.3.3.1. Compounds between Cu Conductor Pads and Sn Based Solders In general, at peak temperatures typical for lead-free reflow soldering processes, i.e., around 240–250°C, Cu_6Sn_5 is the first phase to form at the liquid-tin–copper-conductor interface. The first

stage of the reaction is the dissolution of Cu to liquid solder, until the solder becomes supersaturated with Cu more or less uniformly at the Cu|liquid interface. This saturation limit (metastable solubility) can be determined from the assessed thermodynamic data as shown elsewhere [33,44].

The metastable solubility indicates the largest possible amount of Cu that can dissolve in the liquid without precipitation. The metastable solubility is important, because it essentially determines the dissolution rate of metal to liquid solder. The metastable solubility is typically 2–3 times higher than the stable one in metallic systems. When Cu comes into contact with molten Sn it starts to dissolve rapidly. Initially, the dissolution is a non-equilibrium process and locally very high concentrations of Cu can be realized in the very vicinity of the Cu–liquid interface. However, the composition of the liquid at the interface tends to decrease instantly to the metastable solubility, because extra Cu atoms are depositing back to the Cu surface. Nevertheless, since there is large driving force for the chemical reaction between Cu and Sn atoms, the metastable composition Cu_6Sn_5 crystallites can form very fast by the heterogeneous nucleation and growth at the Cu–liquid interface. In addition to more or less uniform Cu_6Sn_5 layer (uniphase) the two-phase layer (Cu_6Sn_5 + Sn) can form next to the uniphase layer, most likely enhanced by the local constitutional supercooling of liquid. Thermodynamically there should also be a layer of Cu_3Sn between Cu and Cu_6Sn_5. This layer has been experimentally observed to form in many investigations, however, the thickness of the layer appears to be much smaller than that of Cu_6Sn_5 layer and the formation requires rather long contact times.

9.3.3.2. Evolution of the Sn-Cu Intermetallic Compound Layers During Use Because Cu is not in equilibrium with Cu_6Sn_5, reaction in this intermetallic zone will continue through solid state diffusion to form the layer of Cu_3Sn between the Cu pad and the Cu_6Sn_5. The layer of Cu_3Sn generated during soldering is very thin compared with the thickness of Cu_6Sn_5 phase, but the thickness of both these layers increases during the solid state annealing. The growth rate of the Cu_6Sn_5 phase in solid state is faster than that of the Cu_3Sn phase over the temperature range of 60°C and 200°C but the Cu_3Sn will grow partially at the expense of the Cu_6Sn_5 phase. The diffusion rate of Cu in Cu_3Sn is known to be as much as three times higher than that of Sn [45]. This is why Kirkendall voids have been reported to take place in Cu|Sn reaction couple during solid state annealing [46–49]. The authors of this work have also detected these voids but found out that their amount depends on the type of copper foil. In some cases voids form rather uniform plane inside the Cu_3Sn layer or at the Cu–Cu_3Sn interface where as in other cases, such as that shown in Figure 9.14, where the Sn|electrolytic-Cu reaction couple has been annealed at 125°C for 1000 hours, it is not possible to determine such a plane. It should be noted that the quality of copper used in the diffusion couple experiments is very important. When using high purity Cu, only very small sporadic voids can be observed, whereas when using electroplated/electroless deposited copper the voids are easily seen, as shown in Figure 9.14. More discussion on the above observation can be found in Ref. [44].

9.3.3.3. Other Metalization Systems In addition to Cu also other metals, such as Ni, Au and Ag, are used as printed wiring board and component metalization. They also react with Sn to form intermetallic compounds. The reactions in these systems are shortly discussed next. In general, at temperatures around 250°C Ni_3Sn_4 is the first phase to form at the liquid tin/nickel conductor interface. The first stage of the reaction is the dissolution of nickel into the liquid solder, until solder is supersaturated with nickel. Similar arguments regarding the initial periods of dissolution are valid here as already discussed in the case of

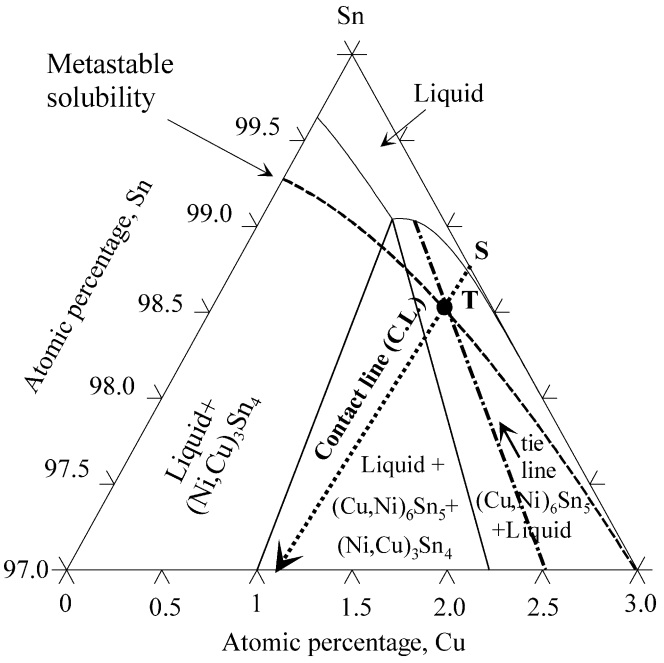


FIGURE 9.15. Enlarged corner of the Sn-Cu-Ni isotherm at 235°C.

Cu. After solder has been supersaturated with Ni more or less uniformly at the Ni–liquid interface, Ni_3Sn_4 nucleates at the interface and starts to grow. Dissolution rate of nickel to a liquid tin is much lower than for example that of copper [42,43]. This is reflected in the thickness of the nickel intermetallic formed during soldering that is generally much thinner than copper intermetallic compounds. Especially the (Ni_3Sn_4 + Sn) two-phase layer is usually absent or very thin.

What has been stated above is the general picture of events when Sn-Pb (or Sn-based solders without copper) are used. However, when using lead-free solders, which include small amounts of Cu, the situation changes and the first phase to form is Cu_6Sn_5[or more precisely $(Cu,Ni)_6Sn_5$]. The formation of $(Cu,Ni)_6Sn_5$ on the Ni metalization can be briefly explained with the help of Figure 9.15, which shows the enlarged Sn-rich corner of the ternary Cu-Ni-Sn isotherm at 235°C. The arrow that starts from S is the contact line from the nominal composition of the SnCu solder to the Ni-corner. The evaluated ternary metastable solubility is shown with the dotted line. During the short soldering period, the composition of the molten solder changes along the contact line from the original solder composition (S) towards Ni as shown in Figure 9.15, because Ni is dissolved into the solder where the Cu/Sn ratio stays unchanged. This is owing to the fact that dissolution of Ni into liquid solder is much faster than diffusion of Cu or Sn into Ni that has to take place via solid-state mechanism. Contact line crosses the evaluated metastable liquidus inside the two-phase region (point T). This crossing point determines the absolute maximum amount of Ni in the solder. Since T is situated inside $(Cu,Ni)_6Sn_5$ + liquid two-phase region, the first phase to form is $(Cu,Ni)_6Sn_5$. The metastable solubility is reached very quickly and therefore the formation of the $(Cu,Ni)_6Sn_5$ on the Ni layer takes place rapidly. The Ni content in the first $(Cu,Ni)_6Sn_5$ crystals formed is determined by the tie line passing through the point T at the $(Cu,Ni)_6Sn_5$ end. The composition of the liquid in local equilibrium

with the $(Cu,Ni)_6Sn_5$ crystals can be obtained from the other end of the tie line. It can be seen from Figure 9.15 that if Cu content of the solder is decreased, the crossing point (T) starts to move to the left along the metastable solubility line (at constant temperature). When Cu concentration in the solder has decreased to about 1 at% (Figure 9.15) the crossing point is inside the $(Ni,Cu)_3Sn_4 + (Cu,Ni)_6Sn_5 + Sn$ three phase region, which means that both IMC phases will form during soldering. Finally, when Cu concentration is decreased to about 0.7 at% the connection line crosses the metastable solubility line inside the $(Ni,Cu)_3Sn_4 + Sn$ two-phase region, $(Ni,Cu)_3Sn_4$ is the first intermetallic compound to form during the soldering process. Thus, depending on the Cu content of the solder $(Ni,Cu)_3Sn_4$, $(Cu,Ni)_6Sn_5$ or both IMCs can form on top of the Ni metallization.

In a solid-state reaction only Ni_3Sn_4 out of the three stable phases grow between Ni and Sn end elements. If the other two stable IMCs grow their thickness are so small that they cannot be detected within the resolution limits of SEM even after prolonged annealing [45]. The formation of fast growing metastable $NiSn_3$ has also been detected at least in one investigation [50]. The temperature region where the metastable phase grows is quite restricted and additional elements present in Sn (for example Pb) suppress its growth effectively.

During soldering Au is the fastest metal to dissolve to Sn-rich alloys. Nevertheless, the same arguments concerning the dissolution process of Au-Sn as discussed in the case of Cu-Sn, also apply here. After solder has been supersaturated with Au more or less uniformly at the Au|liquid interface, the IMCs form out of the supersaturated melt. Results from the solderability experiments with the wetting balance show that in this system the two-phase layer (Au-Sn-IMC + Sn) tends to be very thick with respect to the uniphase IMC layer. This is expected to be related to high metastable solubility of Au in liquid Sn, which in turn indicates high dissolution rate. Thus, if the Au-layer is thin it is dissolved completely and $AuSn_4$ is found as randomly distributed needle-like phases inside the solder matrix after cooling. The dissolution rate of Ag is nearly as high as that of Au [42,43] and therefore during soldering tin will dissolve silver (if applied in a form of a plating) substrate rapidly. Similar arguments concerning the dissolution process with respect to time as discussed previously apply here again. After solder has been supersaturated with Ag more or less uniformly at the Ag|liquid interface, IMCs form by solidification out of the supersaturated melt. The intermetallic phase that has been observed to form is the orthorhombic Ag_3Sn [42,43,51]. If the silver substrate is thick enough the intermetallic forms a continuous layer on top of the original surface. Also in the Ag-Sn system the two-phase layer (Ag-Sn IMC + Sn) tends to be very thick with respect to the uniphase IMC layer. As silver (like gold) is usually used as a surface finish in electronics (i.e., in small quantities) it quickly dissolves from the original interface and forms the Ag_3Sn intermetallic into the bulk solder. The morphology of the Ag_3Sn resembles little bit that of $AuSn_4$. Ag-intermetallic compounds are in the form of relatively large flakes (Figure 9.16) and can therefore cause severe problems with relatively low concentration of the compound.

9.3.4. Deformation Structures (Due to Slip and Twinning)

The familiar stress–strain diagram found practically in all the textbooks on mechanics represents the mechanical properties of metals. If the stress applied is below the yield stress (σ_y) the deformation is said to be elastic. This means that the strain induced is completely removable upon release of stress. At stress levels equal or higher than the yield stress, the metal deforms plastically. This means that deformation is not recoverable upon release

of stress. Any stress above the yield stress is referred as flow stress. Solder materials are generally quite soft (under normal deformation rates) and therefore one is usually interested in plastic deformation. It should be noted that in the stress–strain diagrams metal is usually pictured to be under uniaxial loading. In reality loading is multiaxial in most of the cases. This means that there are several tensile and shear stress components acting simultaneously on the body. In addition, it should be noted that the tensile behavior of material can be highly strain-rate dependent. Especially soft materials, like solders, can behave as much stronger materials under high strain rates than they do under slow strain rates. This issue will be further discussed in the Section 9.4.2.

Plastic deformation of metals occurs by four primary mechanisms: (1) Slip by dislocations, (2) twinning, (3) grain boundary sliding and (4) diffusional creep. The importance of each mechanism depends among other things on stress state, strain rate, temperature and microstructure. The slip mechanism can be defined as: "The parallel movement of two adjacent crystal regions relative to each other across some plane (or planes)" [52]. Slip occurs along some specific (usually close packed) plane in definite slip direction by dislocation movement. The combination of slip plane and slip direction defines the slip system along which dislocation glide occurs. The amount and type of slip systems depends on the crystal structure of the metal in question. In tin, the slip systems are (110) [001], (100) [001], $(10\bar{1})$ [101], and (121) [101] [53]. If a metal crystal posses an insufficient number of independent slip systems, temperature is very low or the strain rate is very high, twin modes may be activated in some metals to provide additional deformation mechanism. Twinning can be defined as follows "A deformation twin is a region of a crystalline body which had undergone a homogeneous shape deformation in such a way that the resulting structure is identical with that of the parent, but oriented differently" [54]. Examples of such twins can be seen for example after drop testing (Section 9.4.2) where the deformation rate has been so high that slip has not had time to occur in large scale.

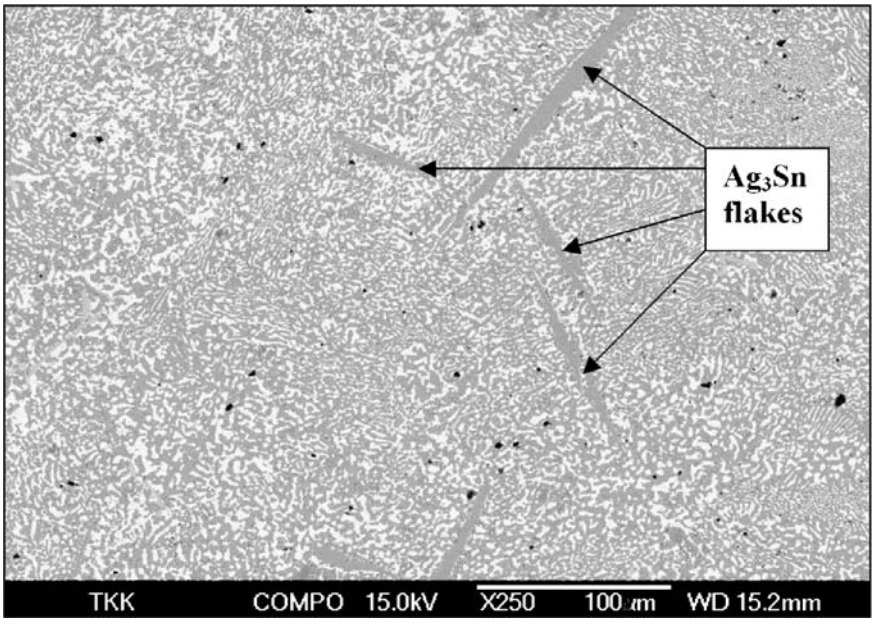


FIGURE 9.16. Solidification structure of SnAgPb (∼3.5 at% Ag) solder showing the large Ag_3Sn flakes.

9.3.5. Recovery, Recrystallization and Grain Growth

During plastic deformation at a sufficiently low temperature, hardening of a metal or alloy occurs. This strain hardening is a result of a net increase in the number of dislocations and other defects in the material. If the metal is subsequently annealed, typical mechanical properties, such as yield strength and hardness, gradually return to values before the deformation. If this process takes place without changes in the grain structure, the phenomenon is called recovery. At higher temperatures or with longer annealing times, most metals will undergo a discontinuous change in grain structure, known as recrystallization. In this process new stress-free crystals grow within the deformed structure and they grow until the original grains are consumed. This primary recrystallization is followed by uniform grain growth, or by highly selective grain growth (secondary recrystallization). The driving force for the recrystallization is the increase in internal energy caused by plastic deformation. Experiments have shown that during plastic deformation only about 1–15% of energy is stored into the structure and rest is dissipated irreversibly as heat [55]. Due to the nature of the recrystallization process a minimum deformation is necessary before it can take place. The kinetics of recrystallization is dependent on a large number of variables, the most important of which are the amount of deformation, the alloying and impurity elements, stacking fault energy, and the original grain size. Naturally the rate is also dependent on time and temperature.

9.4. TWO CASE STUDIES ON RELIABILITY TESTING

In the following two case studies the component used is a lead-free Sn0.2Ag0.4Cu (wt%)-bumped chip scale packaged (CSP) component with 144 bumps (500 μm in diameter) and bump pitch of 800 μm (denoted CSP144 in the following). The heights of the bumps are 480 μm and the under-bump metallurgy consists of ∼0.6–0.8 μm-thick electrochemical Ni, on top of which there originally was a thin gold layer before bumping. The structure of the component and materials used are shown in Figure 9.17.

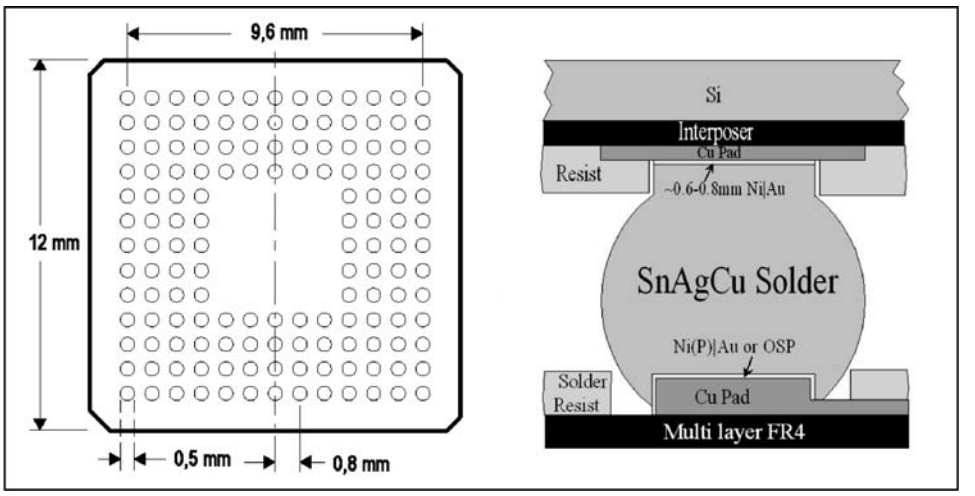


FIGURE 9.17. Structure and dimensions of the CSP144 component.

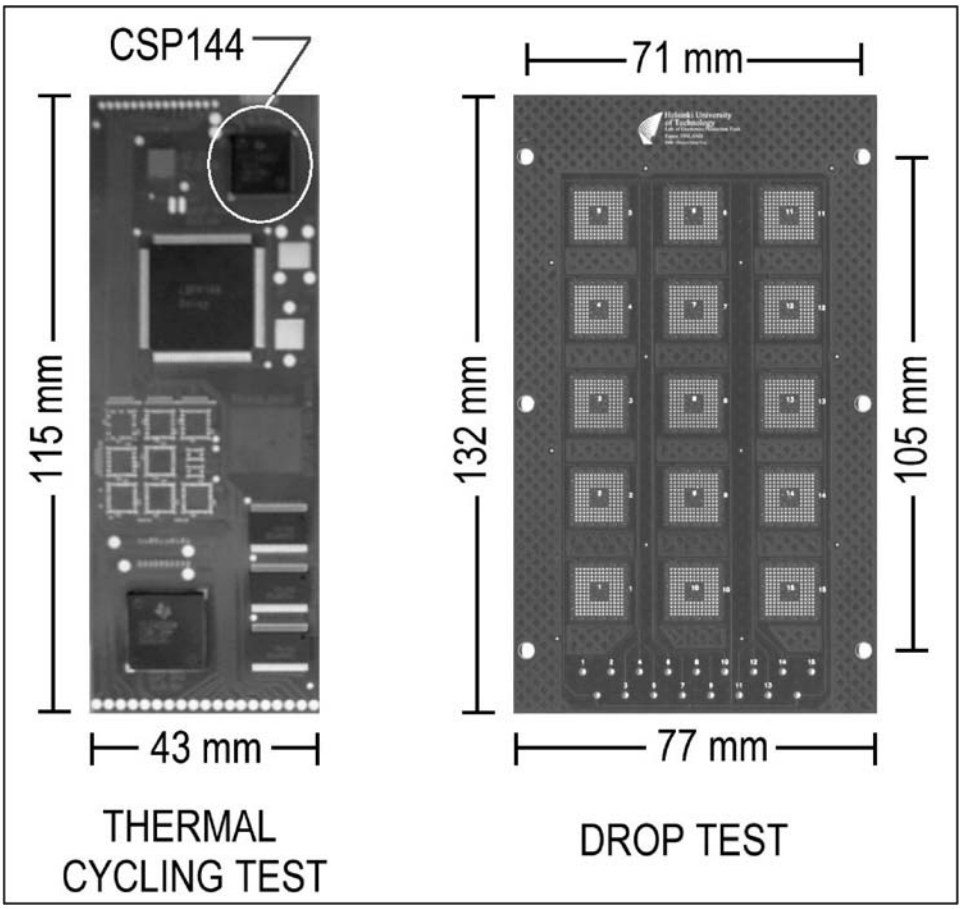


FIGURE 9.18. Layout of the test boards used in (left) thermal cycling tests, and (right) drop tests.

The protective coatings on the printed wiring boards (PWB) are the same in both case examples but the board layouts are different. The high density FR4 boards were manufactured with either electroless Ni plating containing about 9 wt% (16 at%) P with flash Au finish on top (denoted Ni(P)|Au) or organic solderability preservative (denoted Cu|OSP) protective coating on the Cu soldering pads.

The layout and dimensions of both test boards are shown in Figure 9.18. The dimensions of the boards used in thermal cycling tests were 43 mm $\times$ 115 mm $\times$ 1.6 mm. The board was a double-sided four-layer FR4 board. The test assemblies were thermally cycled (Weiss TS 130) according to the IEC 68-2-14N standard ($+125°C/-45°C$, with 15 min dwell time) until all components had failed. The criterion for failure was defined as 20% increase in the initial resistance from that after the reflow.

The high-density circuit boards for the drop tests were designed according to the JESD22-B111 drop test standard. The PWB was a double-sided $(1+6+1)$ stack-up multilayer FR4 board with six inner layers in addition to the topmost resin coated copper layers. The middle component lies in the geometrical centre of the board. The pad and conductor patterning on the board is the same on the two sides except that one side of the board has micro-vias in all of the soldering pads and the other does not. Although the board is double-

sided, components are mounted on one side only. The weight of a fully furnished assembly is 28.5 ± 0.3 g. The test assemblies were drop tested according to the JESD22-B111 standard (see Section 9.2.1.2). A failure was recorded when the resistance through the daisy chain network exceeded the 1.5 kΩ threshold resistance for 200 nanoseconds three times in a sequence of five drops. The test boards were mounted on support rods with screws at the four corners of the board (see Figure 9.18). The components were facing downwards during the test. The drop height was set to 82 cm in order to achieve the required peak deceleration of 1500 G for the duration of 0.5 ms (half-sine pulse).

9.4.1. Case 1: Reliability of Lead-Free CSPs in Thermal cycling

In this case study, the effect of solder pastes and printed wiring board protective coatings on the reliability of the CSP144 interconnections under fatigue stressing will be discussed. The boards were assembled using three different Pb-free no-clean solder pastes. The solder paste compositions (wt%) were Sn4.0Ag0.5Cu, Sn3.8Ag0.7Cu, and Sn3.5Ag0.75Cu, which will be later referred as P_1, P_2 and P_3 respectively. After assembly, the test boards were inspected and subjected to thermal shock testing (IEC 68-2-14N standard: +125°C/−45°C, dwell 15 min/15 min, up to 3000 cycles).

The experimental design included a large number of test structures assembled in a full-scale production line to enable comprehensive statistical analysis. The reliability test procedure was constructed as a full factorial design so that the significance on each factor could be tested with the Analysis of Variance. The type of lead-free *solder paste* and the *PWB coating* were the main variables studied. Results from the statistical analysis carried out with Analysis of Variance showed that no statistically significant differences were found between the CSP144 assemblies soldered with the different solder pastes (risk level $< 0.1\%$). This is due to the fact that the composition of the bump is dominant; most of the solder material composing the interconnections originates from the component bump and only about ten percent, by volume, from the solder paste. The compositions of the solders and the bump material, as well as the nominal interconnection compositions after the reflow, are presented in Figure 9.19, which presents the Sn-rich corner of the SnAgCu phase diagram with isothermal lines representing the liquidus temperatures. Letter B in the diagram depicts the original composition the component bump. N represents the nominal composition on the interconnections on Ni(P)|Au after the reflow and O represents Cu|OSP, respectively (see also Section 9.3.2 of this chapter). Owing to the very small differences in the interconnections compositions between the assemblies soldered with the different solder pastes, the effect of different pastes was ignored in the following analyses.

The type of PWB coating material, on the other hand, was highly significant (significant at risk level $< 0.05\%$; ANOVA): under thermo-mechanical loading the interconnections on the Ni(P)|Au were more reliable than those on the Cu|OSP. The Weibull plot drawn from the thermal cycling results is presented in Figure 9.20. The characteristic life times (η) for the Ni(P)|Au and the Cu|OSP were 1937 and 1485 cycles, respectively. The Weibull distribution shape parameter (β) for the Ni(P)|Au and Cu|OSP are 3.47 and 4.72 respectively. The difference in the beta values is also significant at less than 5% risk level.

Because the only difference between the two groups of samples is the coating material on the soldering pads, the root cause for the different reliability performance must be related to that. As mentioned earlier in Section 9.3.2, an important consequence of higher

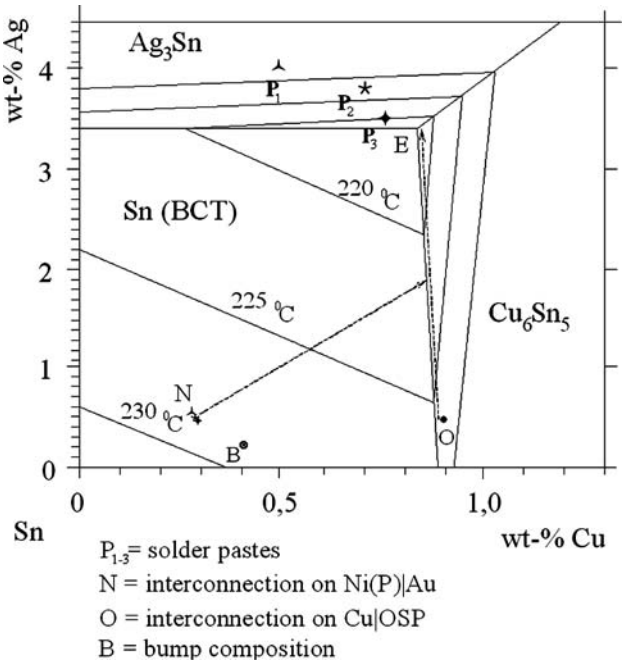


FIGURE 9.19. Compositions of different materials and the solidification path.

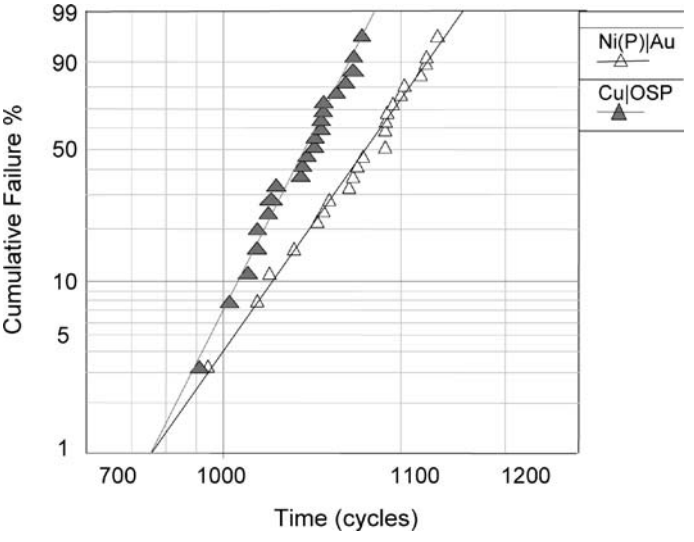


FIGURE 9.20. Weibull reliability plots of the CSP144s on different coatings.

Cu content (due to dissolution of the soldering pad) is that the solidification process and the resulting microstructure is different in these two types of interconnections: relative to Ni(P)|Au, interconnections on the Cu|OSP boards contain more and larger Cu_6Sn_5 intermetallic particles dispersed in the bulk solder. Otherwise the microstructures are very similar: The components under bump metallurgy consist of ~0.6–0.8 μm thick electro-

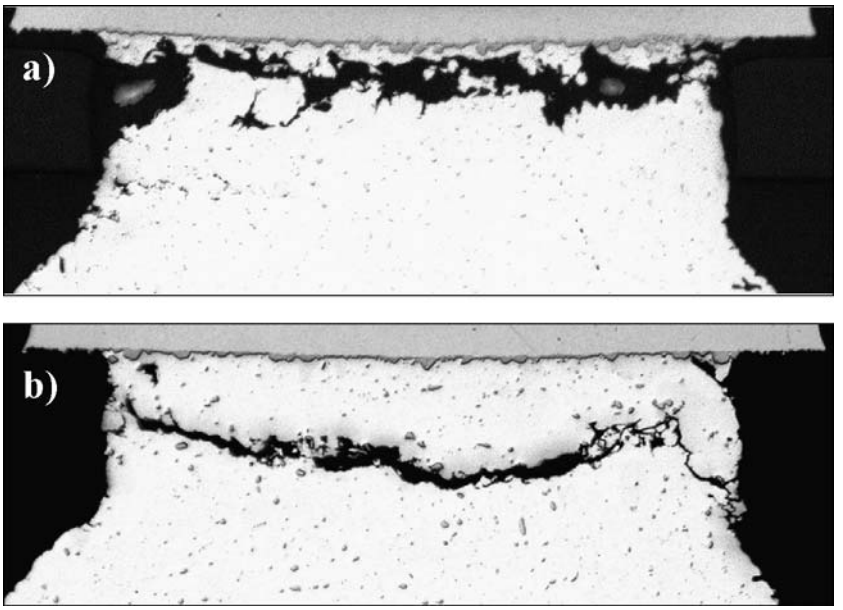


FIGURE 9.21. Crack paths in the interconnection on Ni(P)|Au (a) and Cu|OSP (b).

chemical Ni on the top of which there has been a thin gold layer. The original gold layer had dissolved completely into the bump alloy during bumping. In the bumping stage reflow, the first phase to form is the $(Cu,Ni)_6Sn_5$ instead of the Ni_3Sn_4 as already discussed. The detailed analysis of the reaction can be found elsewhere [56,57]. The component side interconnection interface is thus the same in both the cases. However, the intermetallic layer on the PWB side is different: Cu_6Sn_5 in the Cu|OSP interconnections and $(Cu,Ni)_6Sn_5$ in the Ni(P)|Au interconnections. The detailed fractographic studies showed that no other failure modes were operational than cracks in solder interconnections (see Figure 9.21: (a) interconnections on the Ni(P)|Au, (b) interconnections on the Cu|OSP). Therefore to be more exact, the root cause for the different reliability performance must be related to the differences in the bulk solder and how the microstructures evolve under thermo-mechanical loading.

Micrographs in Figure 9.22 show how the microstructures evolve due to the deformation-induced recrystallization. Polarized light is useful in evaluating microstructures because the reflection is dependent on the grain orientation (asymmetric crystals) of the surface and therefore areas with different orientation are seen in different colors. The topmost micrograph shows the initial structure of a SnAgCu CSP interconnections on Cu|OSP after reflow soldering. The interconnections after reflow consist of relatively few colonies inside of which a cellular solidification structure is visible. The boundaries between the contrasting areas, as seen in Figure 9.9(a), are boundaries of uniformly oriented colonies of small Sn cell. The eutectic structure ($Sn + Cu_6Sn_5 + Ag_3Sn$ phases) is embedded in between the tin cells. $AuSn_4$ is typically seen at the boundaries between the colonies. Micrographs in the middle in Figure 9.22 exemplify how the microstructures evolve during thermomechanical loading. Recrystallization is apparent in the entire "neck region" of the interconnection even after 1000 thermal cycles, while the rest of the interconnection seems to be mostly unaffected. It seems that microstructural deformation

of the most highly stressed areas leads to localized recrystallization of the interconnections.

In fact, the reliability performance on the two types of assemblies is related to microstructural changes that take place during the thermomechanical loading. The difference in the performance between the interconnections on Ni(P)|Au and Cu|OSP reflects the differences in bulk microstructures of the solder interconnections. Because the microstructure of the Cu|OSP interconnections contains numerous relatively large Cu_6Sn_5 primary crystals, the progress of recrystallization is more rapid. These primary particles enhance the onset of recrystallization in the Cu|OSP interconnections. These non-coherent high-angle boundaries between large Cu_6Sn_5 crystals and solder matrix provide advantageous nucleation sites for recrystallization [58,59] and therefore the rate is faster in the Cu|OSP interconnections.

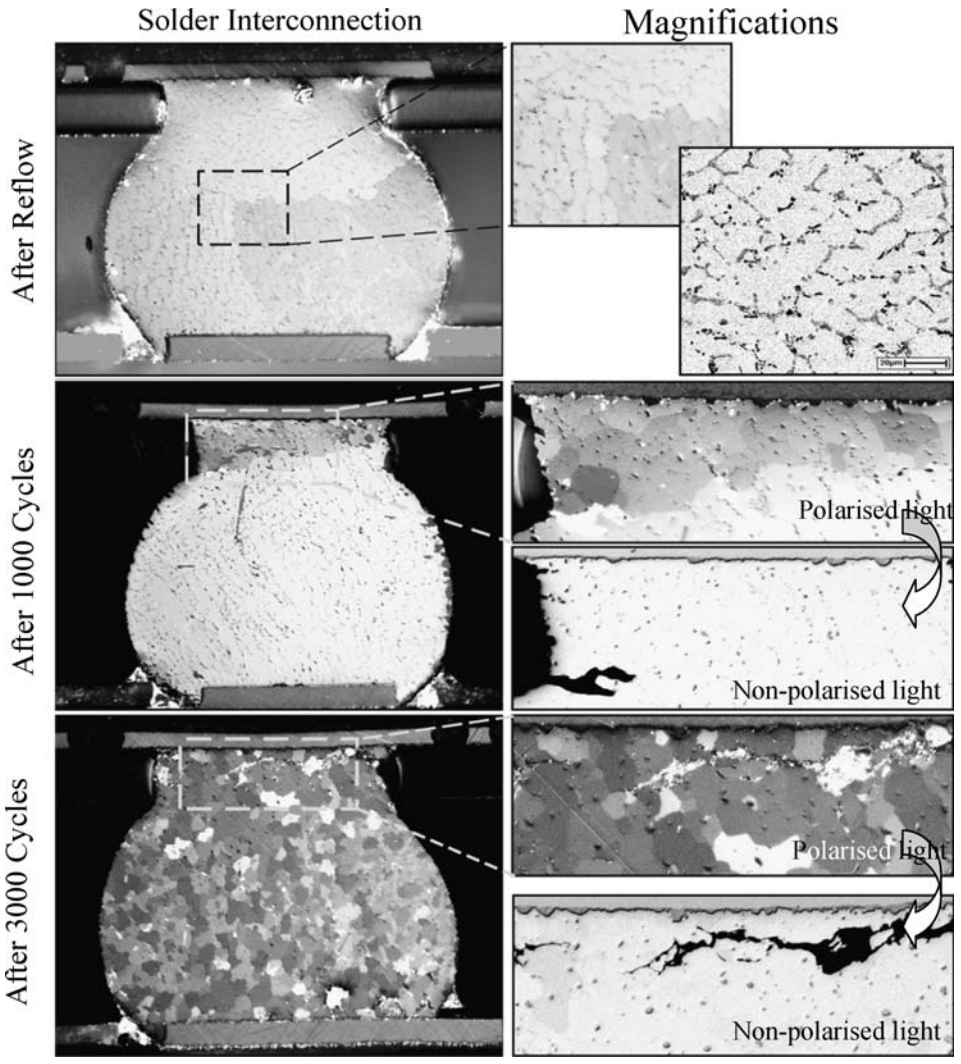


FIGURE 9.22. Evolution of microstructures in solder interconnections during thermal cycling.

The formation of high-angle boundaries between the recrystallized grains favors the nucleation and propagation of intergranular cracks in the boundaries between the recrystallized grains. The cracks do not nucleate only at the corners of the interconnections, but also at the boundary between the recrystallized grains and the non-recrystallized part [15,17–19]. Therefore, the formation of grain boundaries is a prerequisite for the cracks to propagate through the solder interconnections. Because the onset and progress of recrystallization of the interconnections on Cu|OSP is faster, cracks can also nucleate and propagate more rapidly and therefore interconnections on Cu|OSP also fail earlier.

9.4.2. Case 2: Reliability of Lead-Free CSPs in Drop Testing

Portable products are more prone to being dropped than affected by the changes in thermal conditions. Therefore over the past few years, the emphasis of the reliability research has moved from performance of assemblies under thermomechanical loading more towards mechanical shock loading, as has already been discussed. The following case study discusses some of the central issues concerning the reliability of solder interconnections under fast deformation rates.

The reliability of the same lead-free CSP144 component, as in the case study 1, is investigated under mechanical shock loading. The test boards were assembled using the Sn3.8Ag0.7Cu solder paste. After the post-reflow inspection the assemblies were drop tested according the JESD22-B111 standard. The failure mechanisms were studied from cross sections made with the standard metallographic methods. Cross sections were analyzed with the optical, scanning electron, and transmission electron microscopy. Figure 9.23 presents Weibull plot of the Ni(P)|Au and Cu|OSP finished assemblies.

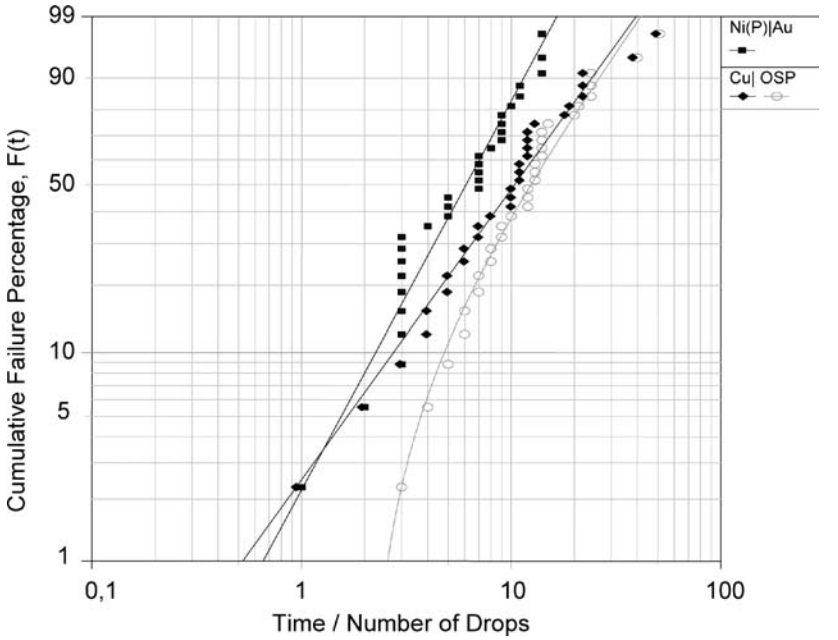


FIGURE 9.23. Weibull reliability plots for the CSP144s on different protective coatings.

The characteristic life times (η) were 7 and 13 drops, and the shape parameters (β) were 1.8 and 1.4 for Ni(P)|Au and Cu|OSP assemblies, respectively. The failure free life time in the case of Cu|OSP is 2 drops. The decision to use either the two or three parameter form of the Weibull distribution was based on the goodness of fit test. The reliability performance between the components soldered on the Ni(P)|Au and the Cu|OSP coated boards was statistically significant at less than 0.01% risk level. The significance between the beta parameters was also statistically significant.

The β values of the drop-tested samples are considerably smaller than those typically encountered in thermal cycling tests and, accordingly, the failure modes and mechanisms were expected to be different. The primary failure modes investigated at the failure analysis stage, were indeed different from those observed in the thermally cycled interconnections. In addition, the mechanisms were also different between the interconnections on the Ni(P)|Au and the Cu|OSP. Interconnections on the Cu|OSP failed from the component side interface, where cracks propagated inside the $(Cu,Ni)_6Sn_5$ intermetallic compound (see Figure 9.24), where as those on the Ni(P)|Au failed from the PWB side interface, where cracks propagated between the $(Cu,Ni)_6Sn_5$ and the Ni(P) metalization (see Figure 9.25). Cracks in the Cu|OSP interconnections typically nucleate at the corner of the interconnections, safe distance away from the intermetallic compound (IMC) layers in the bulk solder, but jump very quickly into the IMC layer, which obviously provides a favorable path for the crack to propagate due to the brittle nature of the compound. The fractures in the Ni(P)|Au interconnections propagate very close to the nickel metalizations underneath the $(Cu,Ni)_6Sn_5$ intermetallic layer as shown in Figure 9.25. The failure modes of the present drop tested samples are very different from those observed after thermal cycling. The failure mode determined earlier in thermally cycled samples of the same material combinations was always an intergranular fracture in the bulk solder.

What makes the crack propagate under drop test conditions inside the IMC rather than in the bulk solder, as was typical for thermally cycled samples? The drop tests are carried out at room temperature ($\sim$295 K), which is relatively high ($0.6T_m$) compared to the melting point of the solder ($\sim$500 K). Therefore, the plastic behavior of the solder is strongly strain rate dependent. As shown in Figure 9.26, the solder becomes remarkably stronger as the strain rate increases from that used in thermal cyclic tests ($\sim 10^{-3}\%/s$) to that used in drop tests ($\sim 10^{4}\%/s$). Thus, in drop tests, where the deformation rate is very high, the solder interconnections are much stronger than those in thermal cycling tests. Subsequently the magnitudes and distributions of the stresses in the solder interconnections are different under thermal cycling and drop test conditions. Finite element calculations showed that as the strain rate increases not only the stresses in solder interconnections increase but also they become more concentrated on the component side of the interconnections [20,21]. Due to the much higher flow stress of the solder interconnections in the drop tests, the intermetallic compound layers will experience significantly higher stresses than those in thermal cycling. The same calculations showed that stresses at the solder|pad interphase on the PWB side are less than half of that on the component side. The tensile strength of the solder increases above the fracture strength of the IMC and this ultimately makes the fractures propagate inside the IMC layers, instead of the bulk solder. In thermal cycling, where the strain rates are relatively low, the cyclic thermomechanical loading of the interconnections generates plastic deformation, which ultimately leads to propagation of fatigue fracture through the solder interconnections.

No recrystallization was observed in the drop-tested samples, even after several months of storage at room temperature. This is because during drop testing the strength of

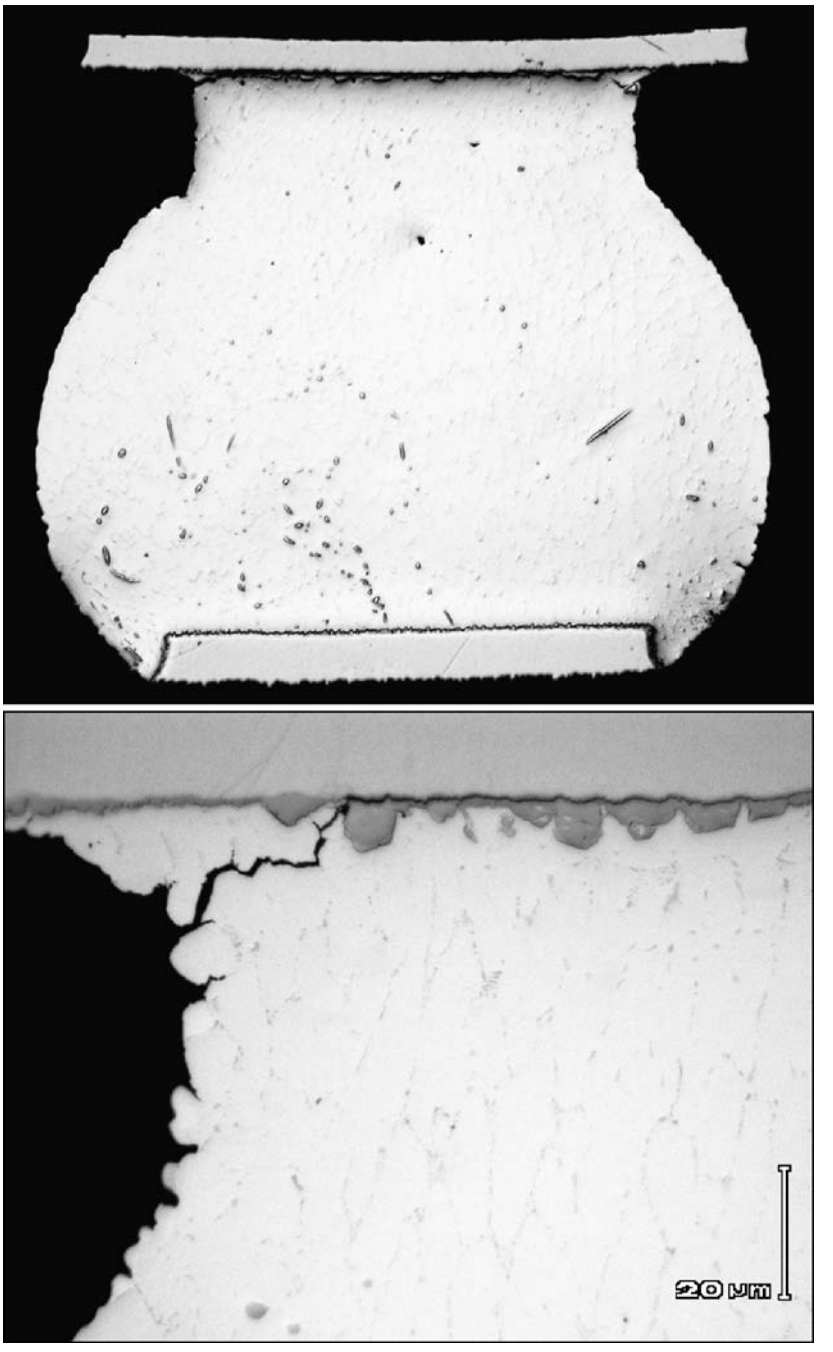


FIGURE 9.24. Component side fracture through the $(Cu,Ni)_6Sn_5$.

the solder interconnections increases and the solder does not markedly deform plastically. As the strain rate is increased twinning mechanism is activated. Twins are typically observed in the regions of the interconnections where stresses are highest [see Figure 9.27(a)].

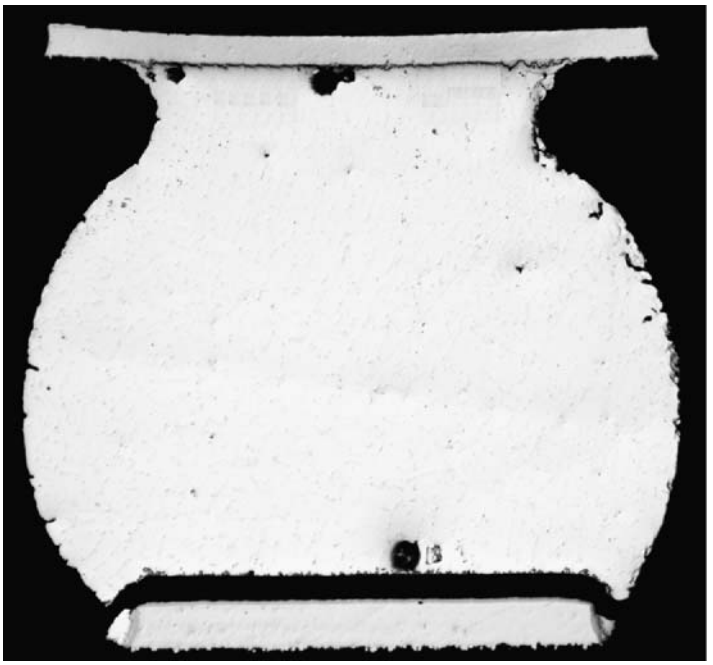


FIGURE 9.25. Failure on the PWB side below the $(Cu,Ni)_6Sn_5$ ("Black Pad" $\Leftrightarrow$ Ni(P)|Au assemblies).

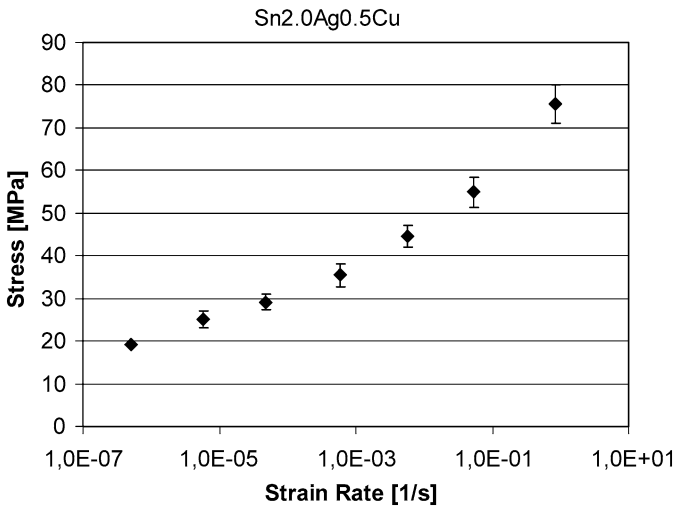


FIGURE 9.26. Flow stress vs. strain rate with standard deviations (Sn2.0Ag0.5Cu at room temperature).

Twins sometimes observed crossing the colonial boundaries [see Figure 9.27(b)] support the conclusion that the boundaries between the colonies are indeed high angle boundaries, as pointed out earlier.

Despite the result that the stresses at the PWB side interface are much smaller than at the component side, the primary failure mode for the interconnections soldered on

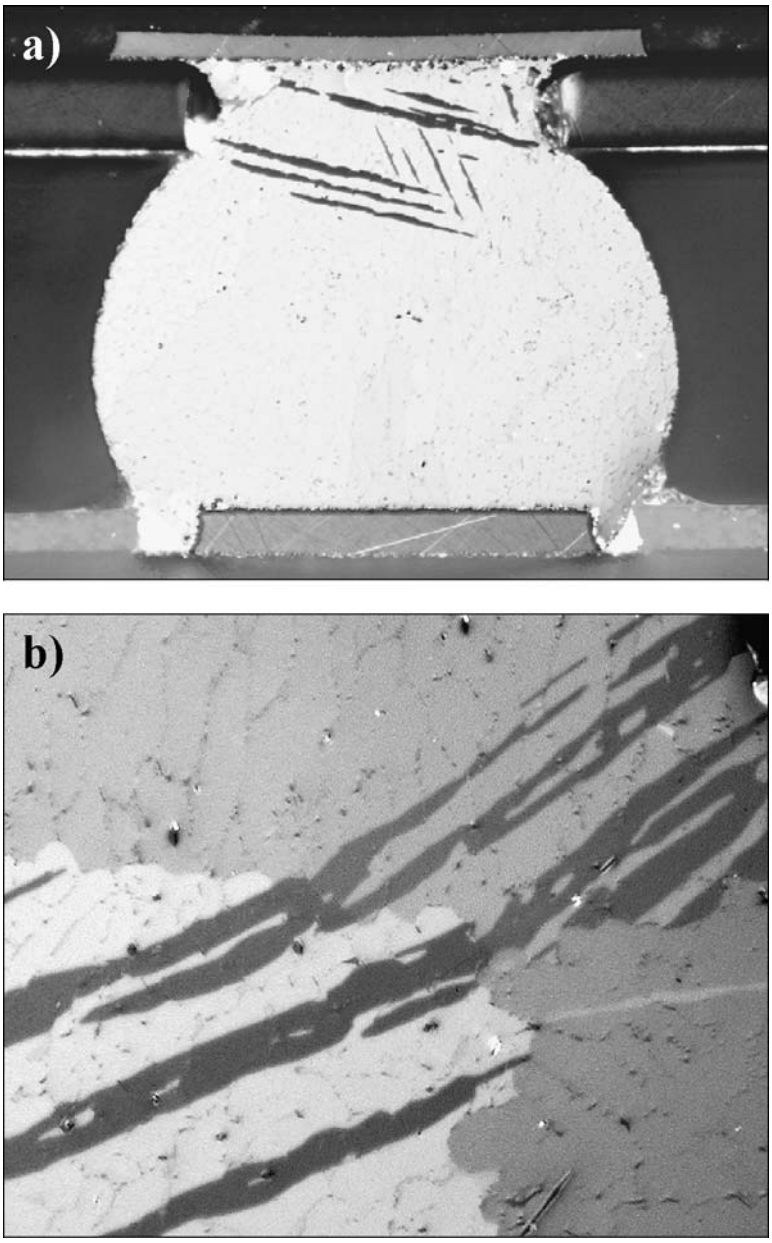


FIGURE 9.27. Structure of the bulk solder after drop tests: mechanical twins. (a) Twins are located at the most highly stresses parts of the solder interconnections. (b) Twins reflect the orientation difference between the large colonies of Sn [see Figure 9.9(a)].

Ni(P)|Au was a fracture on the PWB underneath the $(Cu,Ni)_6Sn_5$ intermetallic layer. This emphasizes the weak nature of this interface. The fracture typically propagates completely through the solder interconnections at a single or very few impacts. The fracture path is always very smooth and straight, compared with the Cu|OSP interconnections fractures

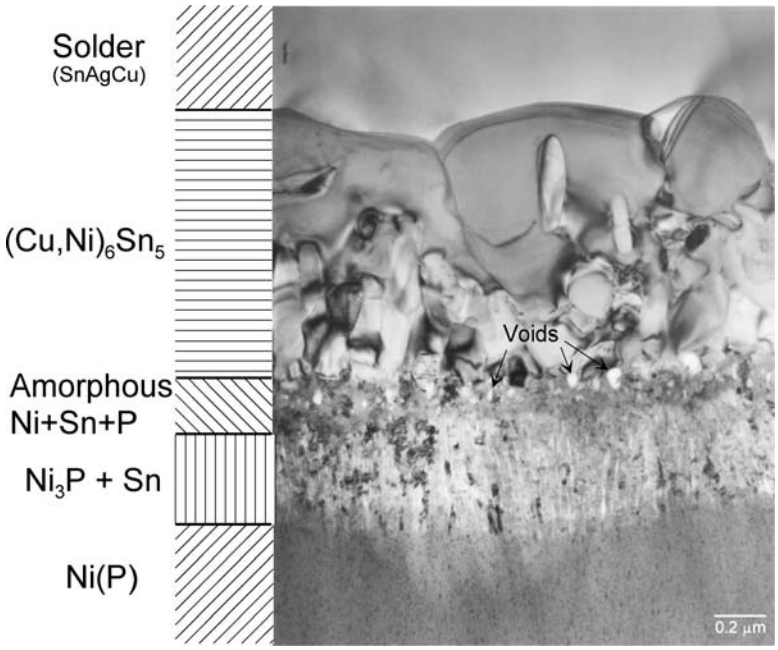


FIGURE 9.28. Transmission electron micrograph from the interfacial reaction layer in the interconnections on the Ni(P)|Au metalization.

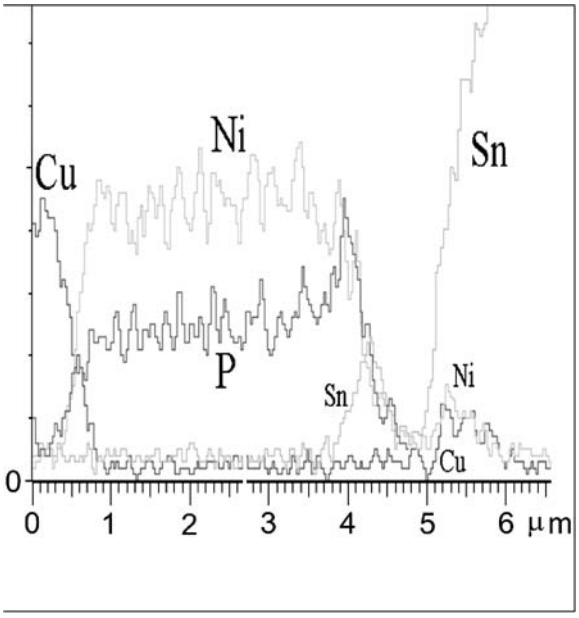


FIGURE 9.29. An energy dispersive X-ray line scan perpendicular to the fracture path.

discussed above. The PWB side intermetallic layer becomes a complex multilayered structure due to the phosphorus trapped in the Ni metalization during chemical deposition. The phase formed next to the Ni(P) metalization is a two phase layer composed of Ni_3P and Sn. This is a columnar layer where Sn can be found in between the Ni_3P columns. The crystalline nickel-phosphate contains more P than the initial electroless Ni(P). On top of that, a thin nanocrystalline layer containing Ni, Sn, and P has been identified. The formation of this complex reaction structure has been discussed in more detail elsewhere [60]. Micrograph in Figure 9.28, taken with transmission electron microscope, shows the structure of the intermetallic layer.

Figure 9.29 shows an EDS line analysis perpendicular to the fracture path. The fracture is located relative to the graphs where all the concentrations decrease considerably (between about 4.5–5μm on the abscissa). To the left, where the amount of Ni and P increase, are the PWB solder pad metalization and to the right, where the Sn concentration increases steeply, is the solder interconnection. Relative to the fracture path layers contain Ni, Sn and P on the board side of the fracture and Cu, Ni and Sn on the solder side. This indicates that the fracture propagates somewhere between the $(Cu,Ni)_6Sn_5$ and the Ni pad metalizations most probably in the nanocrystalline ternary phosphide layer. This type of fracture is very common for the Ni(P)|Au. Since the stresses experienced at the PWB side are only about half of the stresses at the component side, the fact that failure in Ni(P)|Au cases still occurs at the PWB interface shows how strongly the phosphorus influences the reliability.

9.5. SUMMARY

The ongoing trend towards ever smaller electronic products force to larger scales of integration and to the use of smaller and finer pitch components, such as (wafer-level) chip scale packages and flip chips. Because of the small-scale interconnections components become closer to the printed wiring boards and subsequently the strains and stresses experienced by solder interconnections are increased. These miniaturized interconnections must be able to withstand sudden mechanical and thermomechanical shock loadings, local heating of power components, and varying chemical environments.

Lead-free solder interconnections contain more complex intermaterial layers, which weaken the bonding of the solder filler to boards' and components' terminations. Because the microstructures of solder interconnections ultimately control the reliability performance of soldered assemblies a more fundamental understanding of their formation and evolution is needed to ensure the best possible reliability. The above-described development increases also the importance of testing. Because testing is time consuming and expensive, all solutions that can control and limit required testing time are valuable. Simulation tools can be used to reduce unnecessary testing. However, because simulation always requires experimental work to verify the results, the above-mentioned approach must be combined with carefully executed experimental investigations.

Hence, understanding why the different tests yield different failure mechanisms and ultimately different reliability performance is of utmost importance. This can only be achieved by knowing how the stress states produced, how the materials respond to different types of loading, and how the microstructures of solder interconnections affect the failure mechanisms. Therefore the emphasis in this chapter is on microstructures of solder interconnections: solidification, interfacial reactions and evolution of microstructure

are central themes. An approach to study reliability of electronic is introduced. The approach presented consists of simulation, reliability testing, statistical analysis of the test results, and experimental failure analysis. Furthermore, two cases of accelerated testing (thermal cycling and drop tests) are analyzed in the form of case examples. The reasons for the marked differences in failure mechanisms are rationalized with the help of material scientific considerations. Failure mechanism under thermomechanical loading was found out to be entirely different from that under drop testing. Under thermomechanical loading nucleation and propagation of cracks is controlled by the microstructures formed during soldering and their recrystallization behavior during use. Grain boundaries created by recrystallization enable cracks to propagate intergranularly in bulk solder. On the other hand, mechanical shock impacts caused entirely different kinds of failure modes. Cracks in the newly soldered interconnections did not propagate through the bulk solder of the interconnections, but mainly in the brittle intermetallic compound layers between the bulk solder and contact metalizations. This is primarily due to the fact that under very fast loading the ultimate tensile strength of Sn-rich solders is strongly increased because of strain-rate hardening, and therefore the stresses in the solder interconnections grow very rapidly above the fracture strength of the intermetallic compound layers leading to intermetallic fracture.

ACKNOWLEDGMENTS

The authors would like to thank Mr. Pekka Marjamäki for the finite element calculations presented in this work and Dr. Hao Yu for the useful discussions about thermal simulation.

REFERENCES

1. J.K. Kivilahti, Impact of lead (Pb)-free materials on manufacturing and reliability of portable electronics, The Proceedings of IMAPS Nordic Conference, Finland (keynote), 21–24-September 2003.
2. J.K. Kivilahti, Modelling new materials for microelectronics packaging, IEEE Transactions on Components, Packaging and Manufacturing Technology B, 18(2), pp. 326–333 (1995).
3. I.E. Anderson, Tin-silver-copper: A lead free solder for broad applications, The Proceedings of the NEPCON West'96, Anaheim, California, March 25–28, Vol. 2, 1996, pp. 882–885.
4. R. Ninomiya, K. Miyake, and J. Matsunaga, Microstructure and mechanical properties of new lead free solder, Proceedings of ASME INTERPack'97, June 15–19, 1997, Kohala Coast, Island of Hawaii, 1997, pp. 1329–1332.
5. J.S. Hwang, A strong lead-free candidate: the Sn/Ag/Cu/Bi system, Surface Mount Technology, 14(8), pp. 20–22 (2000).
6. P.T. Vianco and D.J. Frear, Issues in the replacement of lead-bearing solders, Journal of Metals, (July), pp. 14–18 (1993).
7. H. Mavoori, J. Chin, S. Vaynman, B. Moran, L. Keer, and M. Fine, Creep, stress relaxation and plastic deformation in Sn-Ag and Sn-Zn eutectic solders, Journal of Electronic Materials, 26(7), pp. 783–790 (1997).
8. Z. Mei and H. Holder, Thermal fatigue failure mechanism of 58Bi-42Sn solder joints, Journal of Electronic Packaging, Transactions of the ASME, 118(6), pp. 62–66 (1996).
9. D. Frear, J. Jang, J. Lin, and C. Zhang, A metallurgical study of Pb-free solders for flip chip interconnects, Journal of Metals, 53(6), pp. 28–38 (2001).
10. K.-L. Lin and H.-M. Hsu, Sn-Zn-Al-Pb-free solder—an inherent barrier solder for Cu contact, Journal of Electronic Materials, 30(9), pp. 1068–1072 (2001).
11. T.T. Mattila and J.K. Kivilahti, Impact of the PWB coatings on the reliability of Pb-free CSP interconnections, The Proceedings of the IMAPS Nordic Conference, Stockholm, 2002.

12. R. Schetty, Lead-free finishes for printed circuit boards and components, in K. Puttliz and K. Stalter, Eds., Handbook of Lead-Free Solder Technology for Microelectronic Assemblies, Marcel Dekker, New York, 2004, p. 431.
13. K.J. Rönkä, F.J.J. van Loo, and J.K. Kivilahti, The local nominal composition—useful concept for microjoining and interconnection applications, Scripta Materialia, 37(10), pp. 1575–1581 (1997).
14. K.J. Rönkä, F.J.J. van Loo, and J.K. Kivilahti, A diffusion-kinetic model for predicting solder/conductor interactions in high density interconnections, Metallurgical and Materials Transactions A, 29A, pp. 2951–2956 (1998).
15. T.T. Mattila, V. Vuorinen, and J.K. Kivilahti, Impact of printed wiring board coatings on the reliability of lead-free chip scale package interconnections, Journal of Materials Research, 19(11), pp. 3214–3223 (2004).
16. D.W. Henderson, J.J. Woods, T.A. Gosselin, J. Bartelo, D.E. King, T.M. Korhonen, M.A. Korhonen, L.P. Lehman, E.J. Cotts, S.K. Kang, P. Lauro, D. Shih, C. Goldsmith, and K. Puttliz, The microstructure of Sn in near eutectic Sn-Ag-Cu alloy solder joints and its role in thermomechanical fatigue, Journal of Materials Research, 19(6), pp. 1608–1612 (2004).
17. S. Terashima, K. Takahama, M. Nozaki, and M. Tanaka, Recrystallization of Sn grains due to thermal strain in Sn-1.2Ag-0.5Cu-0.05Ni solder, Materials Transactions, 45(4), pp. 1383–1390 (2004).
18. S. Terashima and M. Tanaka, Thermal fatigue properties of Sn-1.2Ag-0.5Cu-xNi flip chip interconnects, Materials Transactions, Special Issue on Lead-Free Soldering in Electronics, 45(3), pp. 681–688 (2004).
19. P. Lauro, S.K. Kang, W.K. Choi, and D. Shih, Effect of mechanical deformation and annealing on the microstucture and hardness of Pb-free solders, Journal of Electronic Materials, 32(12), pp. 1432–1440 (2003).
20. T.T. Mattila, P. Marjamäki, and J.K. Kivilahti, Reliability of CSP interconnections under mechanical shock loading conditions, IEEE Transactions on Components and Packaging Technologies (in print).
21. T.T. Mattila and J.K. Kivilahti, Failure mechanisms of CSP interconnections under fast deformation rates, Journal of Electronic Materials, 34(7), pp. 969–976 (2005).
22. T.O. Reinikainen, P. Marjamäki, and J.K. Kivilahti, Deformation characteristics and microstructural evolution of SnAgCu solder joints, The Proceedings of the 6th IEEE EuroSim Conference, Berlin, Germany, 18th–20th of April, 2005, pp. 91–98.
23. K.C. Ong, V.B. Tan, C.T. Lim, E.H. Wong, and X.W. Zhang, Dynamic materials testing and modelling of solder interconnects, The Proceedings of the 54th Electronic Components and Technology Conference, 2004, pp. 1075–1079.
24. JESD22-B111, Board level drop test method of components for handheld electronic products, JEDEC Solid State Technology Association, 2003, p. 16.
25. D.C. Montgomery, Design and Analysis of Experiments, 5th edition, John Wiley & Sons Inc., New York, 2001, p. 672.
26. J.S. Milton and J.C. Arnold, Introduction to Probability and Statistics, 3rd ed., McGraw-Hill, New York, 1995, p. 811.
27. A. Mitra, Fundamentals of Quality Control and Improvement, Prentice Hall, New Jersey, 1998, p. 752.
28. J.S. Hunter, Design and analysis of experiments, in J.M. Juran and F.M. Gryna, Eds., Juran's Quality Control Handbook, 4th edn, McGraw-Hill, New York, 1988.
29. Engineering Statistics Handbook, NIST/SEMATECH e-Handbook of Statistical Methods, http://www.itl.nist.gov/div898/handbook/, 20.1.2005.
30. P.D.T. O'Connor, Practical Reliability Engineering, John Wiley & Sons, Chichester, 1998, p. 431.
31. H. Yu and J.K. Kivilahti, Thermal modelling of reflow process, Soldering and Surface Mount Technology, 14(1), pp. 38–44 (2002).
32. H. Yu, T.T. Mattila, and J.K. Kivilahti, Thermal simulation of the solidification of lead-free solder interconnections, IEEE Transactions on Components and Packaging technologies (in print).
33. J.K. Kivilahti and K. Kulojärvi, A new reliability aspect of high density interconnections, in R.K. Mahidhara, Ed., The Proc. of Design and Reliability of Solders and Solder Interconnections, TMS Annual Meeting, Orlando, USA, February 9–13, 1997, pp. 377–384.
34. P. Savolainen and J.K. Kivilahti, Feasibility of lead-free solder alloys as filler materials for Z-axis adhesives, Soldering and Surface Mount Technology, 5(20), pp. 10–12 (1995).
35. J.K. Kivilahti, The chemical modelling of electronic materials and interconnections, Journal of Metals, 54(12), pp. 52–57 (2002).
36. F.J.J. van Loo, Multiphase diffusion in binary and ternary solid-state systems, Progress in Solid State Chemistry, 20(1), pp. 47–99 (1990).
37. U.R. Kattner, The thermodynamic modeling of multicomponent phase equilibria, Journal of Metals, 49(12), pp. 14–19 (1997).

38. M. Hillert, Phase Equilibria, Phase Diagrams and Phase Transformations: Their Thermodynamic Basis, Cambridge Univ. Press, 1998.
39. W.A. Tiller, The Science of Crystallization, Cambridge University Press, Cambridge, UK, 1991.
40. W. Kurz and D.J. Fisher, Fundamentals of Solidification, Trans. Tech. Publications, 1989.
41. T.J. Koivisto, Master's Thesis, Helsinki University of Technology, Laboratory of Electronic Production Technology, 2004.
42. W.G. Bader, Dissolution of Au, Ag, Pd, Pt, Cu and Ni in a molten tin-lead solder, Welding Journal: Research Supplement, 48(12), pp. 551–557 (1969).
43. W.G. Bader, Dissolution and formation on intermetallics in the soldering process, Proceedings of the Conference on Physical Metallurgy and Metal Joining, St. Louis, MO, Warrendale, USA, Oct. 16–17, 1980.
44. T. Laurila, V. Vuorinen, and J.K. Kivilahti, Interfacial reactions between lead-free solders and common base materials, Materials Science and Engineering R, 49(1–2), pp. 1–60 (2005).
45. M. Oh, Growth Kinetics of Intermetallic Phases in the Cu-Sn Binary and the Cu-Ni-Sn Ternary Systems at Low Temperatures, Doctoral Dissertation, Lehigh University, 1994.
46. T.-C. Chiu, K. Zeng, R. Stierman, D. Edwards, and K. Ano, Effect of thermal aging on board level drop reliability for Pb-free BGA, The Proceedings of the Electronic Components and Technology Conference, 2004. ECTC '04, June 1–4, Vol. 2, 2004, pp. 1256–1262.
47. M. Amagai, Y. Toyoda, T. Ohnishi, and S. Akita, High drop test reliability: lead-free solders, the proceedings of the electronic components and technology conference, 2004. ECTC '04, June 1–4, Vol. 2, 2004, pp. 1304–1309.
48. M. Umemoto, K. Tanida, Y. Nemoto, M. Hoshino, K. Kojima, Y. Shirai, and K. Takahashi, High-performance vertical interconnection for high-density 3d chip stacking package, The Proceedings of the Electronic Components and Technology Conference, 2004. ECTC '04, June 1–4, Vol. 2, 2004, pp. 616–623.
49. A. Paul, The kirkendall effect in solid state diffusion, Doctoral Deissertation, Technical University of Eindhoven, 2004.
50. J. Haimovich, Intermetallic compound growth in tin and tin-lead platings over nickel and its effects on solderability, Welding Journal: Research Supplement, 68(3), pp. 102–111 (1989).
51. C. Thwaites, Solderability of coatings for printed circuits, Institute of Metal Finishing Transactions, 43, pp. 143–152 (1965).
52. J.D. Verhoeven, Fundamentals of Physical Metallurgy, John Wiley & Sons, New York, 1975, p. 567.
53. C.S. Barrett and T.B. Massalski, Structure of Metals, McGraw-Hill, 1960, p. 654.
54. B.A. Bilby and A.G. Crocker, The theory of the crystallography of deformation twinning, Proceedings of the Royal of London, Series A, 288(1413), pp. 240–255 (1965).
55. A.L. Titchener and M.B. Bever, The stored energy of cold work, Progress in Metal Physics, 7, pp. 247–338 (1958).
56. T. Laurila, V. Vuorinen, and J.K. Kivilahti, Analyses of interfacial reactions at different levels of interconnection, Material Science in Semiconductor Process, 7(4–6), pp. 307–311 (2004).
57. T. Laurila, V. Vuorinen, T.T. Mattila, and J.K. Kivilahti, Analysis of the redeposition of $AuSn_4$ on Ni/Au contact pads when using SnPbAg, SnAg, and SnAgCu solders, Journal of Electronic Materials, 34(1), pp. 103–111 (2005).
58. W.C. Leslie, T.J. Michalak, and F.W. Aul, The annealing of cold-worked iron, in C.W. Spencer and F. E. Werner, Eds., Iron and its Dilute Solid Solutions, Interscience Puhlishers, New York, 1963.
59. R.W. Cahn, Recovery and recrystallization, in R.W. Cahn, Ed., Physical Metallurgy, North-Holland Publishing Company, Amsterdam, 1965, pp. 925–987.
60. V. Vuorinen, T. Laurila, H. Yu, and J.K. Kivilahti, Phase formation between lead-free SnAgCu solder and Ni(P)/Au finished on PWB, Journal of Applied Physics, 99(2), pp. 3530–3536 (2006).

10

Metallurgy, Processing and Reliability of Lead-Free Solder Joint Interconnections

Jin Liang[a], Nader Dariavach[a], and Dongkai Shangguan[b]

[a]*EMC Corp., Hopkinton, MA 01748, USA*
[b]*FLEXTRONICS, San Jose, CA 95131, USA*

10.1. INTRODUCTION

Soldering is one of the most important manufacturing processes in the electronics industry. Reliable, high quality, long-lasting electronic products demand uncompromised integrity of each individual solder interconnection inside electronic packages and in PCB/component assemblies. The drive to lead-free soldering due to global legislations and market forces has a fundamental effect on material selection and manufacturing processes in the industry. The current viable Pb-free solders are based on the near-eutectic composition of the Sn-Ag-Cu ternary system with or without additions of other alloying elements. These Pb-free alloys have higher melting temperatures, different mechanical and physical properties from the current Sn-Pb eutectic solders, and could have significant implications in terms of soldering processes, compatibility, and end product reliability.

In this chapter, the metallurgy, processing and reliability of lead-free solder joint interconnections is discussed. Intermetallic formation kinetics and morphology, as well as microstructures of typical lead-free solder joints, are presented, followed by a discussion on the solder wetting behavior with different board and component metallic finishes and general process compatibility with the current technologies of packaging, design, and processing. Furthermore, this chapter reviews effects of mechanical loading and thermal conditions on the time-dependent non-linear deformation and fatigue behavior of lead-free solder alloys. An attempt is made to compare the reliability of Pb-free solder joint interconnections to the Sn-Pb eutectic solder joints in different applications and loading conditions. A guideline is provided at the end of this chapter for better utilization of the unique beneficial attributes of Pb-free alloys, and for prevention of potential processing and reliability issues.

10.2. PHYSICAL METALLURGY OF LEAD-FREE SOLDER ALLOYS

Use of lead-bearing solders, such as the Sn-Pb eutectic solder, can be dated back to early human history. In the last 50 years, the Sn-Pb solders have found extensive usage in electronic and semiconductor industries for large volume production for printed circuit board (PCB)/component assemblies with highly automated processes and process controls. In principle, solders are used for joining purposes because they possess the following useful characteristics: a liquidus temperature lower than melting points of the materials to be joined; molten solders wet or spread on the substrate metallic or metalization surfaces and form sound metallic bonds without significant erosion of the surfaces to be joined [1]. Strengths of the final solder interconnections are determined by the solder chemical composition, processing conditions, and particularly by the metallurgical reactions of the molten solder with the metallic surfaces to be joined. The interfacial reactions, such as wetting or spreading between molten solder and metallic surfaces, depend on many factors, such as intrinsic chemical affinity, surface cleanness, thermodynamics and kinetics of intermetallic formation and growth.

10.2.1. Tin-Lead Solders

The Sn-Pb binary system has a eutectic reaction around 183°C with a composition of 63Sn-37Pb (wt%). The eutectic reaction is taken as the following form [2,3]:

$$\mathrm{L} \rightarrow \beta \text{ Sn (solid solution with Pb, tetragonal lattice)} + \alpha \text{ Pb (solid solution with Sn, BCC lattice).}$$

The solubility at the eutectic temperature is 19% of Sn in Pb, and 2.5% of Pb in Sn. The Sn solubility in Pb decreases significantly to less than 2% at room temperature, while Pb solubility in Sn is reduced to literally zero. Thus, during the solidification or aging at room temperature, the secondary Pb and Sn will precipitate from the original Sn-Pb eutectic structures. Also, Sn-Pb solders with compositions other than eutectic will also have primary lead phase (for Pb-rich solders) or primary Tin-rich phase (for Tin-rich solders) forms as dendrites, with subsequent precipitations of saturated Pb or Tin phases within these primary phases.

The Sn-Pb solder alloys offer the following advantages as compared with other solders:

(1) Superior wetting and spreading characteristics on metallic substrates, such as, Au, Pt, Pd, Cu, Ni, Ag, and other metals and alloys, with minimal substrate erosion.
(2) Satisfactory metallic bonding strength, ductility, stiffness, and fatigue resistance.
(3) Ready application as soldering preservation coatings on PCBs and on component leads by electro-plating or dipping.
(4) Relatively inexpensive to produce and use.

However, lead (Pb), as a pure element or an additive and alloying element, is also toxic to human beings. For this reason it has been or will be banned for use in many industries and applications. Even with its superior manufacturing process attributes in the electronics industry, the European Union has passed laws to ban the usage of Pb in electronic products starting from July 1, 2006 [4,5].

10.2.2. Lead-Free Solder Alloys

For a smooth transition to lead-free soldering, ideally, the lead-free alloys would have melting points around 180°C, close to that of the Sn-Pb eutectic alloy, and are of their constituent eutectic compositions. However, there are very few tin-based Pb-free solder alloys that satisfy the above criteria. Sn-Bi (42Sn-58Bi) eutectic and Sn-In (48Sn-52In) eutectic have relatively low melting points, 138°C and 118°C, respectively [6,7]. Due to relative poor high temperature mechanical strength, lack of ductility (for Sn-Bi alloys), and limited resource, these alloys are finding only very limited applications in the industry. Sn-Zn or Sn-Zn-Bi alloys have melting points close to the Pb-Sn eutectic alloys, and have been used with some success, particularly in East Asia for some consumer products [8]. However, with Zinc being prone to oxidation in the high temperature soldering processes, and to corrosion (possible conductive corrosion by-products), these alloys are unlikely to be used for volume production as a general Pb-free solution. The currently most promising Pb-free solder candidates are based on the Sn-Ag-Cu ternary system, which has a eutectic composition around Sn-3.8Ag-0.7Cu, melting temperature around 217°C, about 34°C above that of the Sn-Pb eutectic alloy (183°C) [9].

The potential Pb-free candidates are listed in Table 10.1 and shown in Figure 10.1. Both eutectic composition and non-eutectic alloys are shown. For increased fluidity of molten solders, non-eutectic alloy's paste range (the temperature range from solidus to liquidus points) should be kept as small as possible.

The binary system phase diagrams for Sn-Ag and Sn-Cu are shown in Figure 10.2 and Figure 10.3, respectively [10,11]. Unlike Sn-Pb eutectic alloy, the Sn-Cu and Sn-Ag alloys form eutectic reactions with their intermetallic compounds (η' Cu_6Sn_5 for the Sn-Cu binary system and γ Ag_3Sn for the Sn-Ag system) instead of their solid solutions like in the case of the Sn-Pb binary system. Under a nearly thermodynamic equilibrium solidification condition, tin will solidify as nearly pure β phase without any significant solid solution of either Ag or Cu, co-existing with η' Cu_6Sn_5 for the Sn-Cu system or γ Ag_3Sn for the Sn-Ag system. However, under most industrial solidification conditions, the eutectic reactions will be off-equilibrium, thus beta tin could contain solid solute atoms of Ag and/or Cu.

TABLE 10.1.
Tin-based Pb-free solder alloys [10].

Alloy system	Composition (wt%)	Melting points or range (°C)
Sn-Bi	Sn-58Bi	138 (e)
Sn-In	Sn-52In	118 (e)
	Sn-50In	118–125
Sn-Zn	Sn-9Zn	198.5 (e)
Sn-Bi-Zn	Sn-8Zn-3Bi	189–199
Sn-Cu	Sn-0.7Cu	227 (e)
Sn-Ag	Sn-3.5Ag	221 (e)
	Sn-2Ag	221–226
Sn-Ag-Bi	Sn-3.5Ag-3Bi	206–213
	Sn-7.5Bi-2Ag	207–212
Sn-Ag-Cu	Sn-3.8Ag-0.7Cu	217(e)
	Sn-3.0Ag-0.5Cu	218
Sn-Ag-Cu-Sb	Sn-2Ag-0.8Cu-0.5Sb	216–222
Sn-Sb	Sn-5Sb	232–240
Sn-Au	Au-20Sn	280

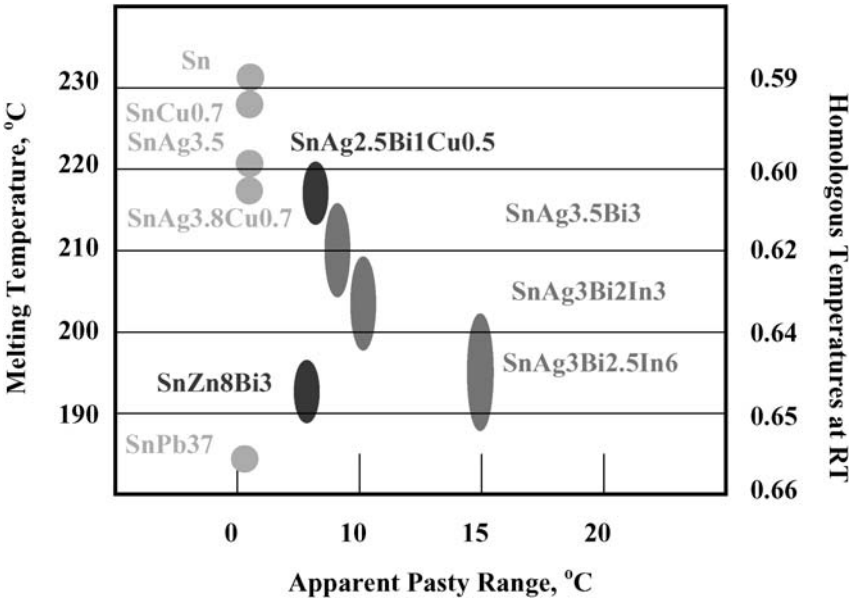


FIGURE 10.1. Pb-free solder candidates to replace the Sn-Pb eutectic solder, their melting temperatures, solidification paste ranges, and homologous temperatures at ambient.

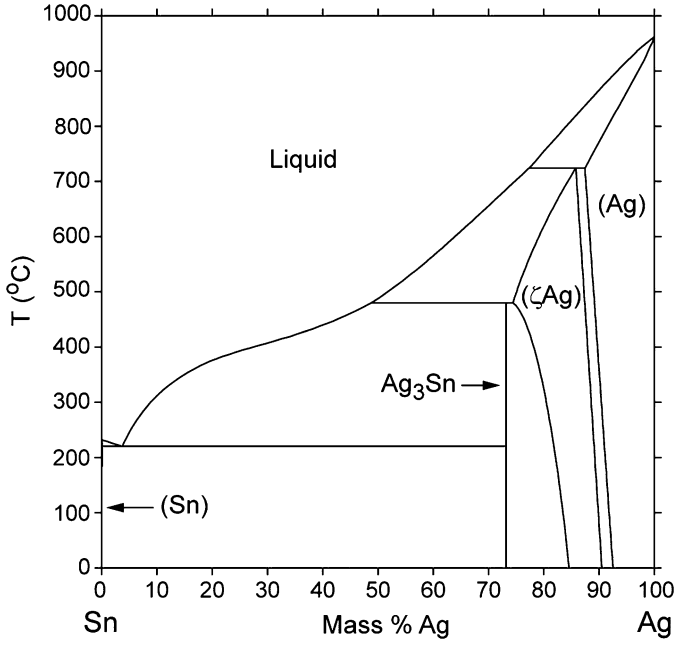


FIGURE 10.2. Sn-Ag binary phase diagram [10].

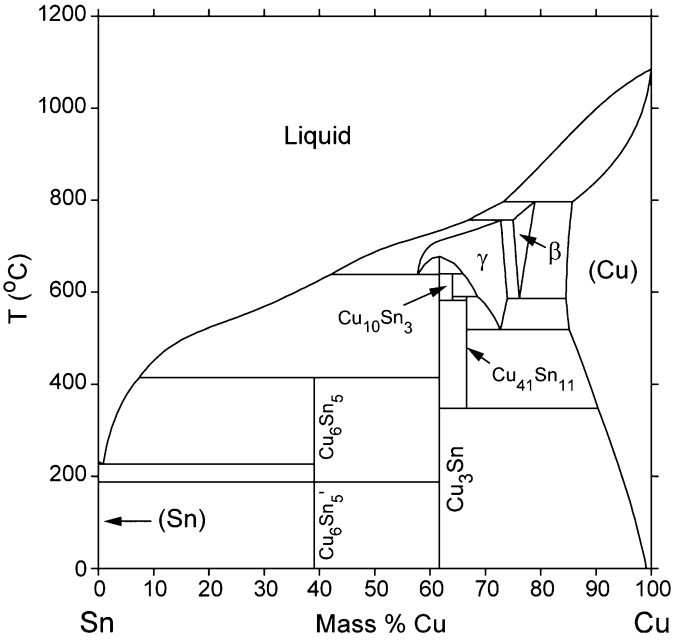


FIGURE 10.3. Sn-Cu binary phase diagram [11].

Subsequent precipitations of stable or metastable intermetallic phases are possible at room temperature and during solidification or aging at temperatures below melting points.

There is still no general agreement about the exact eutectic composition for the Sn-Ag-Cu ternary system; the most reported eutectic compositions are Sn-3.8Ag-0.7Cu and Sn-3.6Ag-0.9Cu (wt%). Currently, the most widely recommended and studied Pb-free alloys are SAC387 (Sn-3.8Ag-0.7Cu) and SAC305 (Sn-3.0Ag-0.5Cu), along with SAC369 (Sn-3.9Ag-0.6Cu) and SAC405 (Sn-4.0Ag-0.5Cu), and a handful of quaternary systems, such as Sn-Ag-Cu-Bi and Sn-Ag-Cu-Fe [12].

The ternary eutectic reaction of the Sn-Ag-Cu system can be expressed as [13]:

$$\mathrm{L} \rightarrow \mathrm{Ag_3Sn} + \mathrm{Cu_6Sn_5} + \beta\ (\mathrm{Sn}).$$

Depending on the exact chemical compositions and solidification conditions (cooling rates, undercooling temperature ranges, etc.), Sn-Ag-Cu alloys may experience primary reactions prior to the ternary eutectic reaction. For example, it was reported that for SAC387, the solidification reaction sequence was found to be [13]:

$$\mathrm{L} \rightarrow \mathrm{L1} + \mathrm{Ag_3Sn} \text{ at } 221.9^{\circ}\mathrm{C},$$

$$\mathrm{L1} \rightarrow \mathrm{L2} + \beta\ (\mathrm{Sn}) \text{ at } 218.7^{\circ}\mathrm{C},$$

$$\mathrm{L2} \rightarrow \mathrm{Ag_3Sn} + \mathrm{Cu_6Sn_5} + \beta\ (\mathrm{Sn}) \text{ at } 217^{\circ}\mathrm{C}.$$

However, should the Ag concentration be less than the eutectic composition, the first reaction would not take place. The possible exiting phases in the Sn-Ag-Cu ternary system are

TABLE 10.2.
Phases, crystal structures in Sn-Ag-Cu system [10].

Phase	Common names	Spacegroup	Chemical composition
Liquid	L	n/a	$(Ag,Cu,Sn)_1$
FCC	(Ag), (Cu)	Fm-3m	$(Ag,Cu,Sn)_1(Va)_1$
BCC	(beta Cu), beta	Im-3m	$(Cu,Sn)_1(Va)_3$
HCP	(zeta Ag), (epsilon Pb)	$P6_3/mmc$	$(Ag,Sn)_1(Va)_{0.5}$
BCT	(Sn), (beta Sn)	$I4_1/amd$	$(Ag,Cu,Sn)_1$
Ag_3Sn	epsilon	Pmmn	$(Ag)_{0.75}(Sn)_{0.25}$
$Cu_3Sn.h$	gamma	Fm-3m	$(Cu,Sn)_{0.75}(Cu,Sn)_{0.25}$
$Cu_{41}Sn_{11}$	delta	F-43m	$Cu_{0.788}Sn_{0.212}$
$Cu_{10}Sn_3$	zeta	$P6_3$	$Cu_{0.769}Sn_{0.231}$
Cu_3Sn	epsilon	Cmcm	$Cu_{0.75}Sn_{0.25}$
Cu_6Sn_5	eta, $Cu_6Sn_5.h$	$P6_3/mmc$	$Cu_{0.545}Sn_{0.455}$
Cu_6Sn_5'	eta′, $Cu_6Sn_5.l$	...	$Cu_{0.545}Sn_{0.455}$

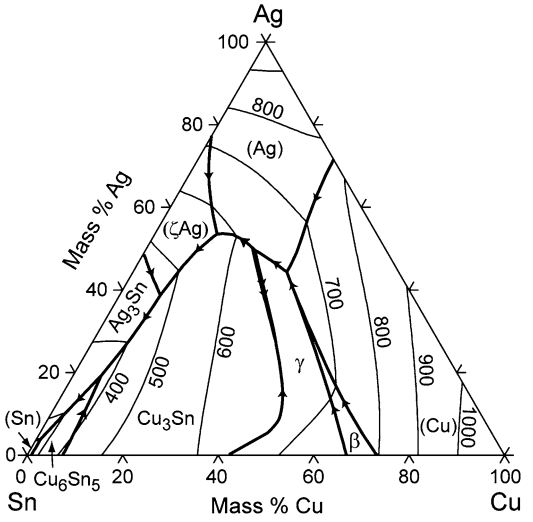


FIGURE 10.4. A projection of Sn-Ag-Cu phase diagram showing the ternary eutectic reaction.

listed in Table 10.2 [14]. The projected ternary phase diagram is shown in Figure 10.4. Figure 10.5 shows the detailed phase-temperature relationships at the Tin-rich corner where most of the current Pb-free solder alloys are based.

The typical Sn-Pb and -Sn-Ag-Cu solidification structures and microstructures are shown in Figure 10.6, with the pictures taken *in-situ* with an ESEM with heating stage [15]. The as-solidified structures for Sn-Ag-Cu (SAC378) and Sn-Pb eutectic have sharp difference. The Sn-Pb eutectic solder solidifies as dendrites and fine eutectic two phase structures. For SAC387, the structure is of fine plate shape without obvious ternary eutectic structure seen with the unpolished surface at low magnification. Also, the SAC alloy tends to have more solidification shrinkage and looks dull compared to the Sn-Pb eutectic solder, which is much smoother and shinier. The microstructures of actual Sn-Ag-Cu solder joints for a chip capacitor and a BGA solder ball are shown in Figure 10.7 and Figure 10.8. High magnification optical micro-photos for Sn-Ag-Cu solder on Ag (Figure 10.9) and OSP (Fig-

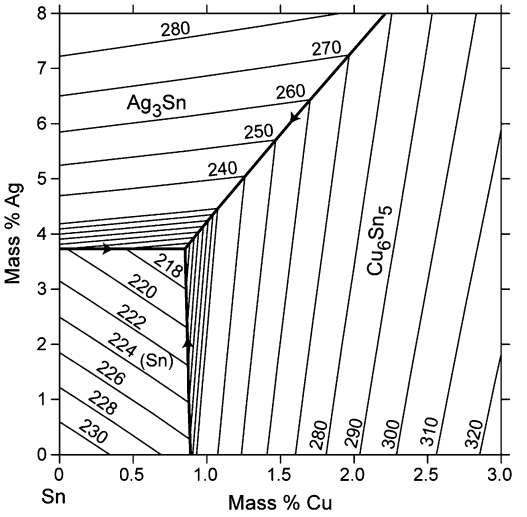
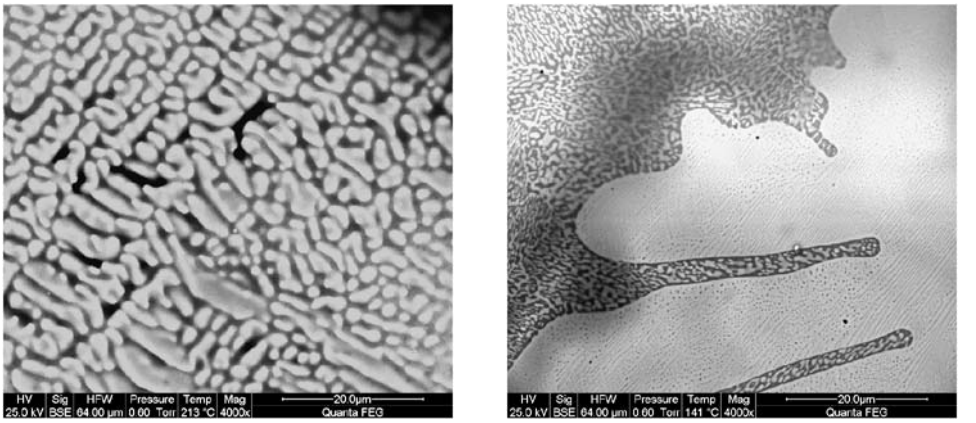


FIGURE 10.5. Sn-Ag-Cu phase diagram at the Sn-rich corner [14].



(a) (b)

FIGURE 10.6. Typical as-solidified structures of solders, (a) SAC387 Pb-free; (b) Sn-Pb eutectic.

ure 10.10) finish boards show clearly secondary intermetallic phases present in the SAC378 alloy, and a bilayer of intermetallic compounds at the solder-Cu substrate interface.

10.2.3. Interfacial Reaction: Wetting and Spreading

The wetting of a molten solder on metallic surfaces is a rather complex phenomenon. The soldering technology has generally evolved in an empirical manner. Factors, such as the conditions of the metallic surfaces (i.e., the nature of oxides or other films, surface roughness), temperature distribution during the soldering, as well as the interfacial metallurgical reactions, and flux chemical reaction with the metallic surface, all play important roles in determining the final solder wetting, spreading and the solder joint shape, as well as the mechanical strength of the joints in general [1].

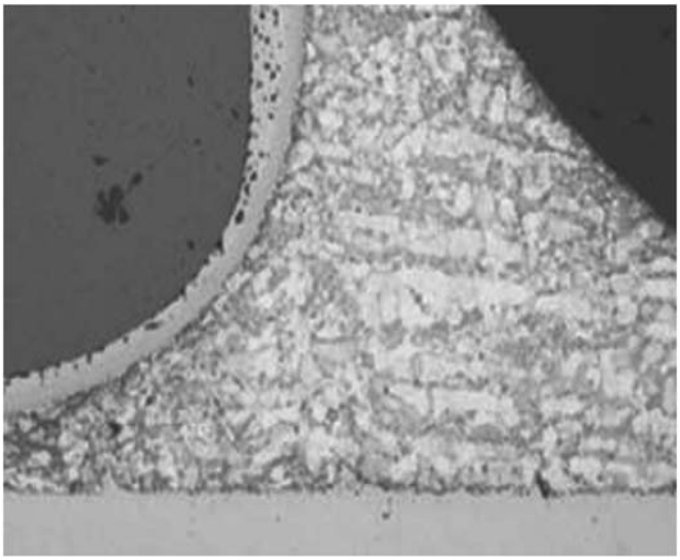


FIGURE 10.7. Microstructure of Sn-Ag-Cu solder joint for a chip capacitor.

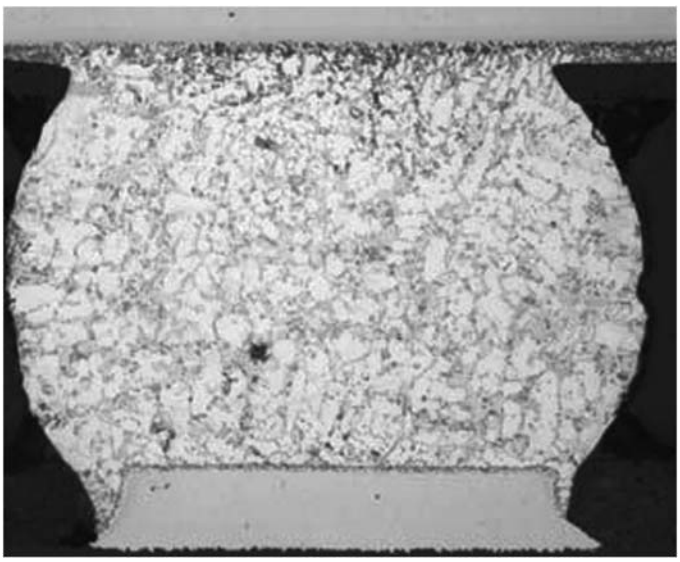


FIGURE 10.8. Microstructure of Sn-Ag-Cu solder for a BGA solder ball.

Classic wetting theory has been established on simple systems like water or oil on a non-reaction surface, such as glass, at relatively low temperatures. The physico-absorption-dominated wetting driving force is the capillary reaction to reduce the total surface energy in the liquid-solid-vapor system of interest. The restraining force for the wetting or spreading is viscosity. At balance, the three surface tensions between the liquid/solid substrate

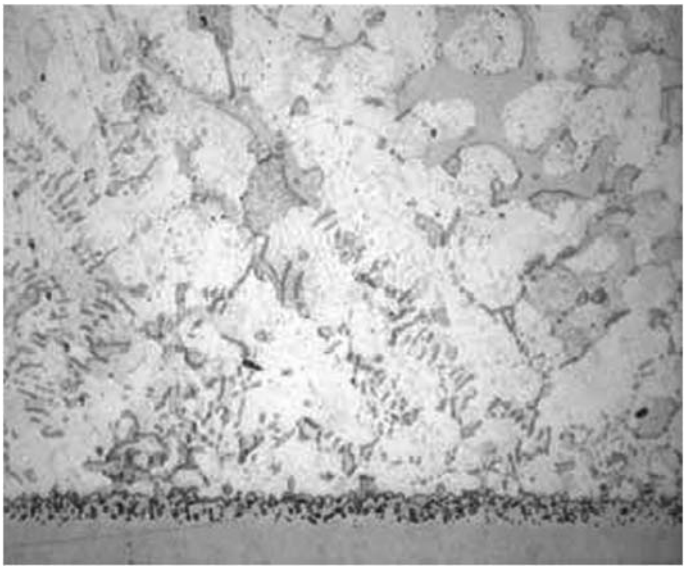


FIGURE 10.9. Microstructure of Sn-Ag-Cu solder on Ag finish board.

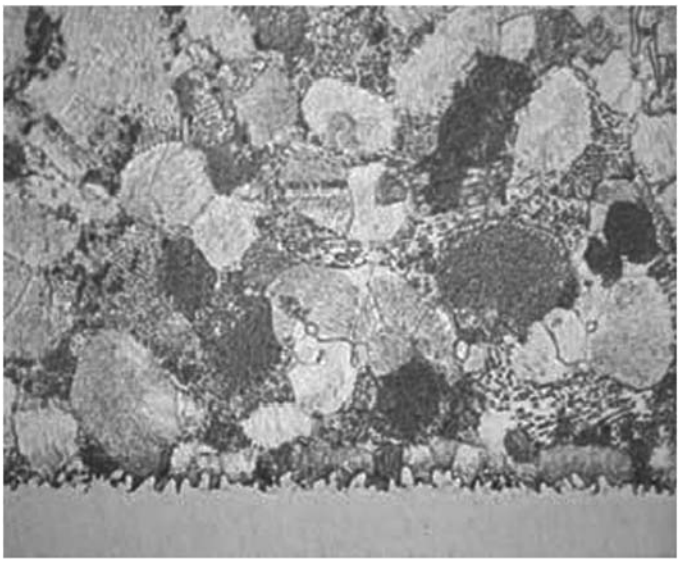


FIGURE 10.10. Microstructure of Sn-Ag-Cu solder on OSP finish board.

(γ_{LS}), the liquid/vapor (γ_{LV}), and the solid substrate/vapor (γ_{SV}), together determine the final wetting angle θ, which can be expressed as follows [1]:

$$\cos\theta = \frac{\gamma_{SV} - \gamma_{LS}}{\gamma_{LV}}. \tag{10.1}$$

A perfect wetting means that the wetting angle θ approaches zero. For physical wetting, the efficient way to decrease the wetting angle is to increase the surface energy of solid substrate/vapor (γ_{SV}) by cleaning the solid surface with removal of absorbed species which tend to reduce the surface energy of the solid surface. The surface energy between the liquid/solid substrate (γ_{LS}), is constant for a given non-reaction solid–liquid system, but very much temperature dependent. As temperature rises, γ_{LS} decreases rapidly, thus promoting wetting and spreading of the liquid phase. The liquid/vapor surface energy (γ_{LV}) is also constant under a fixed temperature and pressure condition for a particularly liquid–vapor system. It changes with pressure and atmosphere chemical composition. Lower pressure or vacuum will reduce (γ_{LV}), thus promoting the wetting and spreading.

The soldering process almost invariably involves many physical-chemical reactions at the interfacial and liquid–solid–vapor junction area. The surface tension energies between the liquid/solid substrate (γ_{LS}) and between the solid/vapor substrate (γ_{SV}) are changed rapidly both by flux chemical reactions, dissolution of parent metals, and intermetallic formations at the junction area. The pure physical wetting phenomenon takes place rather quickly (from 100 ns to 10 ms), while wetting or spreading of molten solder takes a much longer time (usually from 0.5 second to up to a few of minutes), indicating the spreading rate and wetting angle in the soldering process are very much dependent on the complex local chemical reaction kinetics and thermodynamics [1], not a purely physical wetting process.

The surface tension energies between the liquid/solid substrate (γ_{LS}) in a solderable system may be significantly different before and after interfacial reactions. The formation of an intermetallic layer would reduce the total energy balance at the liquid-solid interface. Assuming the Gibbs free energy per unit area for forming an intermetallic layer between the molten solder and the metallic substrate is ΔG_r (which is a negative number), the surface energy between the liquid/solid substrate can be expressed as:

$$\gamma_{LS} = \gamma_{LS}^0 + \Delta G_r, \tag{10.2}$$

where γ_{LS}^0 is the surface tension energy between liquid and solid before any interfacial reaction takes place. Since ΔG_r is a negative number, the liquid/solid surface tension energy γ_{LS} is reduced by the formation of the intermetallic layer, thus reducing the wetting angle according to Equation (10.1). Furthermore, it has been found that the absolute ΔG_r value is much larger (by two orders of magnitude) than the initial surface energy for molten solders between the liquid/solid substrate, as reported by Yost and Roming [16] and Wang and Conrad [17]. Therefore, it is clear that anything that affects proper intermetallic formation will cause significant change to the wetting rate and wetting angle in the soldering processes. This explains why molten solders rarely wet or spread on some metallic (such as aluminum and its alloys) and nonmetallic (such as ceramic and plastic) surfaces which do not form an intermetallic layer or the intermetallic formation kinetics is too slow to promote wetting and spreading of the molten solders.

As discussed above, the wetting and spreading of molten solders depend on many factors. Should a clean surface be maintained, typical wetting angles for any liquid solder and metallic substrates are intrinsic wettability compatibility measurements, which are dependent on chemical compositions of the liquid solder and solid substrates. Table 10.3 gives some typical wetting angles for Sn-Pb eutectic solder, and some Sn-based Pb-free alloys on different metallic substrates or metalization surfaces at the typical reflow temperatures for these alloys. What is clear from this table is that the intrinsic wetting angles of

TABLE 10.3.
Typical wetting angles for Sn-based Pb-free alloys [16].

	Reflowed alloy pellet (Sn+)					
Substrate	0.5Cu	3.5Ag	3.8Ag0.7Cu	3.5Ag0.7Cu	3.8Ag0.7Cu0.5Sb	37Pb
Cu	42	43	43	41	43	12
Ag	19	26	24	30	33	13
Sn37Pb	19	19	22	20	22	5
Sn0.7Cu	15	11	18	11	10	17
Au over Ni	9	6	10	14	5	4

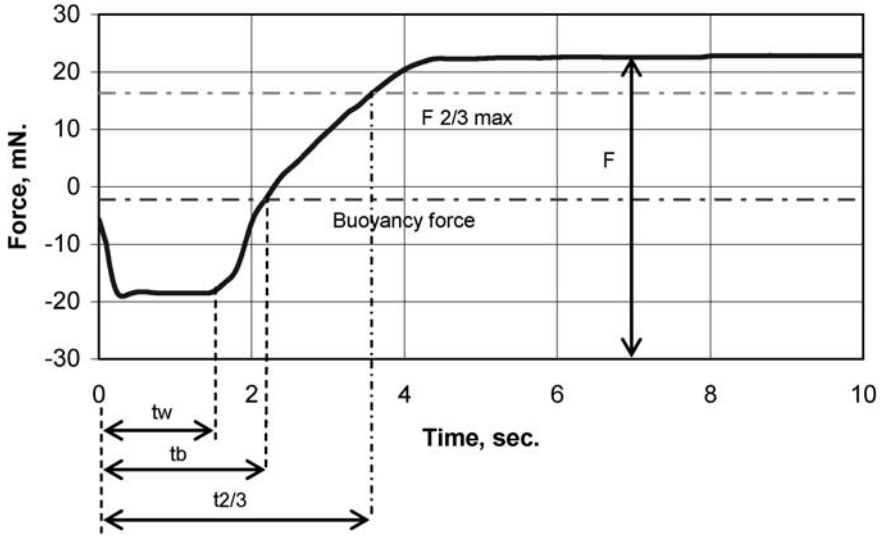


FIGURE 10.11. Wetting balance test interpretation.

Sn-based Pb-free alloys are not as good as those for the 63Sn-37Pb eutectic alloy on the copper substrate and other metallic finishes. However, most Pb-free alloys wet very well on Ni/Au metalization surface; and the wetting angles of these Pb-free alloys on Au/Ni are close or better than that for Sn-Pb eutectic on copper surface.

Another way for wettability evaluation under dynamic conditions is the wetting balance test, which involves investigation of both wetting times and wetting forces [18]. This method uses immersion of test samples into a solder bath, then recording the vertical force, which is the sum of buoyancy and surface tension forces over time. The typical wetting curve is shown in Figure 10.11. Commonly measured wetting balance characteristics are the initial wetting time t_w, the time for the wetting curve to re-cross the buoyancy force line t_b, the maximum wetting force F_{max} and the time required to rich the $2/3$ of the maximum force $t_{2/3}$. The wetting curve comparison for Sn-Pb and three Pb-free alloys, Sn-0.7Cu, Sn-3.5Ag, and Sn-3.8Ag-0.7Cu (SAC387), on copper substrate is shown in Figure 10.12. Measurements were taken at 235°C using $12.5 \times 25.0 \times 0.5$ mm copper test coupons. It can be clearly seen that the wetting times t_w, t_b and $t_{2/3}$ for the Sn-Pb alloy are significantly shorter than those for the Pb-free alloys. The best wetting performance for the tested Pb-free alloys is from the Sn-3.8Ag-0.7Cu alloy. Increasing the test temperature up to 260°C

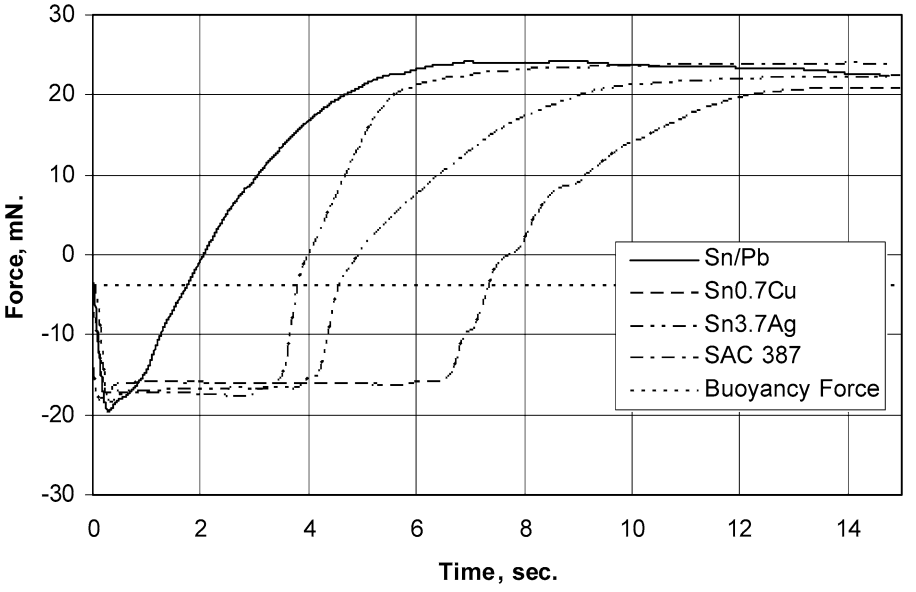


FIGURE 10.12. Wetting balance curves for Sn-Pb eutectic and Pb-free solders at 235°C.

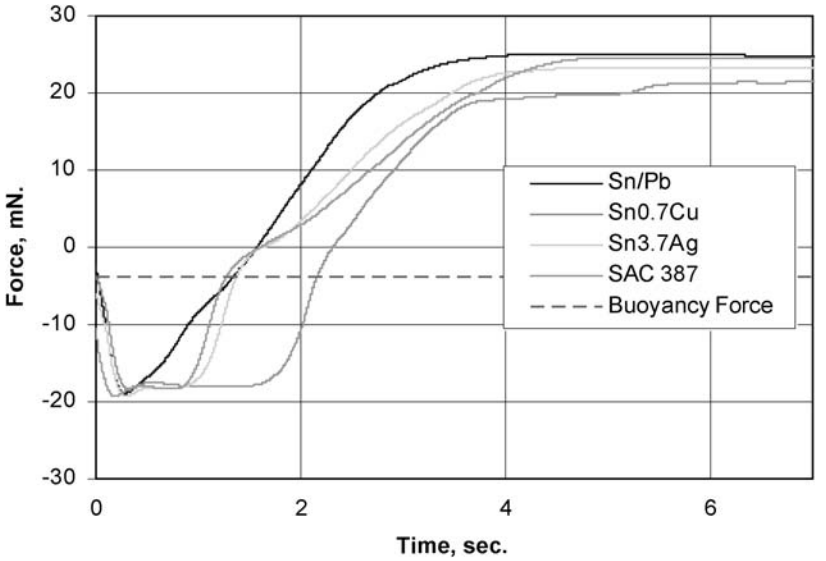


FIGURE 10.13. Comparison of wetting curves at 260°C for Sn-Pb eutectic and Pb-free solders.

decreases the wetting times significantly for these alloys (Figure 10.13). The buoyancy time t_b of Sn-3.5Ag and SAC387 alloys is equal to that of the Sn-Pb alloy at higher temperatures (>250°C). The wetting time t_b of the Sn-0.7Cu alloy is still significantly longer as compared with Sn-Pb eutectic, Sn-3.5Ag and SAC387 at temperatures above 250°C. Figure 10.14 shows the measurement results for initial wetting time at different testing temperatures for all these alloys. From these wetting balance test results, it is clear that the

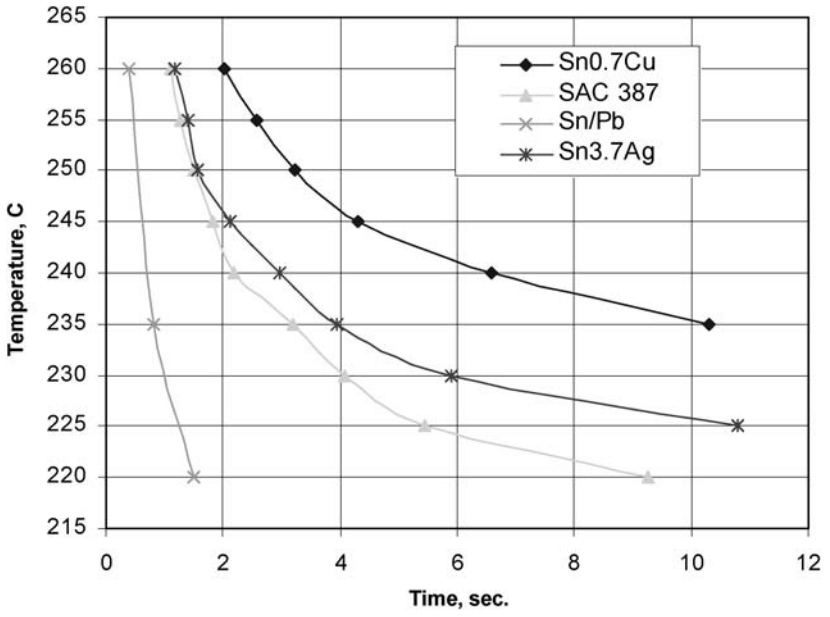


FIGURE 10.14. Comparison of initial wetting time for Sn-Pb eutectic and Pb-free solders.

Pb-free soldering process requires an increase in the process temperature up to 245–250°C for Sn-3.5Ag and Sn-3.8Ag-0.7Cu alloys, and up to 260°C for the Sn-0.7Cu alloy, which has been a candidate Pb-free alloy for wave soldering.

10.2.4. Interfacial Intermetallic Formation and Growth at Liquid–Solid Interfaces

As stated above, the intermetallic growth and substrate dissolution take place rather rapidly during a normal soldering operation. Since lead-free soldering requires substantially higher temperatures (around 250°C), the rates for intermetallic growth and substrate dissolution are expected to be significantly greater for Pb-free solders than those for the current Sn-Pb eutectic solder. A thorough understanding of lead-free solder/substrate interfacial reactions can lead to the optimum lead-free soldering processes and the optimum lead-free coating thicknesses for component and PCB termination finishes. Excessive intermetallic compounds (IMCs) can cause embrittlement of solder joints and decrease fatigue strength, leading to unfavorable reliability for Pb-free PCB assemblies.

10.2.4.1. Intermetallic Growth Kinetics on Cu Substrate Reactions between copper and molten tin at the copper/solder junction for three Pb-free solders, Sn-3.5Ag, Sn-0.9Cu, and Sn-3.8Ag-0.7Cu, produced a bilayer of Cu_6Sn_5 adjacent to the solder and Cu_3Sn adjacent to the copper [19]. Figures 10.15, 10.16 and 10.17 present SEM images of IMC layers formed at 235°C for 10 sec, 3 min, 30 min and 2 hr holding times for these three solders. After holding for 10 sec, thin Cu_6Sn_5 intermetallic layers of average thicknesses of 1.18, 1.33 and 1.24 μm, were observed respectively for the three alloys. The presence of the Cu_3Sn phase for these samples was not detected at this temperature for such a short time. Cu_3Sn phase, however, appears with increased holding time. The final average thicknesses of the IMCs after holding for 2 hrs were 6.97, 7.65 and 9.44 μm for the three solders on copper substrate. The maximum thicknesses for the three alloys were determined to be

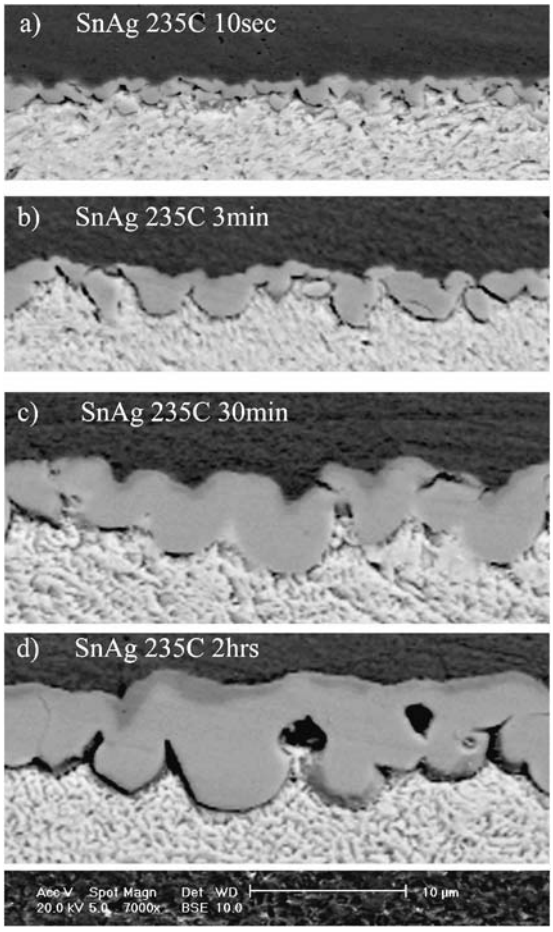


FIGURE 10.15. Comparison of SEM images of the intermetallic layers for the SnAg alloy cast at 235°C and held for 10 sec (a), 3 min (b), 30 min (c), 2 hrs (d).

8.42, 10.37 and 16.5 μm. The similarity of the scalloped intermetallic microstructures for all three alloys indicates that the mechanism of growth is the same for all three. Comparison of the scanning micrographs in Figure 10.17 with those in Figures 10.15 and 10.16 clearly shows that the intermetallic grows faster in the Sn-Ag-Cu (SAC387) alloy than in the other two alloys.

Figure 10.18 illustrates effects of temperature on IMC growth for Sn-3.5Ag alloy at 225 and 280°C. The sample held for 30 sec does not show the presence of Cu_3Sn. Increasing the temperature up to 280°C promotes Cu_3Sn layer formation with a thickness of 0.36 μm. The total average thickness of the IMC increases by 89% from 1.14 μm up to 2.16 μm with an increase in temperature from 225 to 280°C.

The IMC layer growth varies depending on the diffusion rates of Cu and Sn through the Cu_6Sn_5 and Cu_3Sn layers and the reactions at the layer interfaces. It is further complicated by the fact that the Cu_6Sn_5 layer is scalloped [20,21]. As such, the thickening or

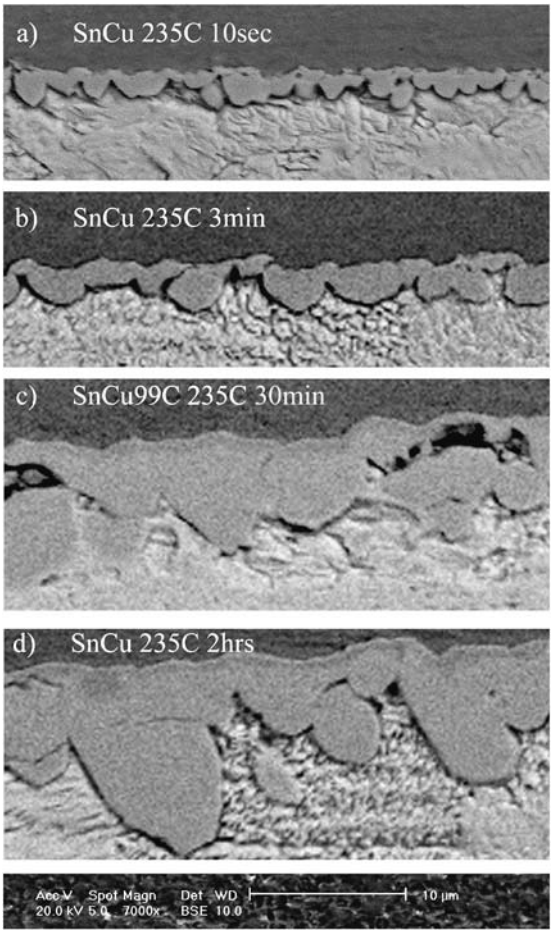


FIGURE 10.16. Comparison of SEM images of the intermetallic layers for the SnCu cast at 235°C and held for 10 sec (a), 3 min (b), 30 min (c) and 2 hrs (d).

growth kinetics of the total IMC layer has been evaluated by simply assuming parabolic growth. That is, it is assumed that the total layer thickness is given by

$$w = w_0 + kt^{\frac{1}{2}}, \tag{10.3}$$

where w is the intermetallic layer thickness, t is the holding time, k is the IMC growth rate constant, and w_0 is the initial thickness of the IMC layer formed on immersion of the copper sample in the solder bath. As such, the measured average layer thicknesses are plotted as a function of the square root of holding time for each temperature studied and fitted with a linear least squares curve to determine the values of k and w_0 for each temperature. The k values are subsequently analyzed to determine the apparent activation energy (Q) for IMC growth using the Arrhenius equation,

$$k = k_0 e^{-\frac{Q}{RT}} \tag{10.4}$$

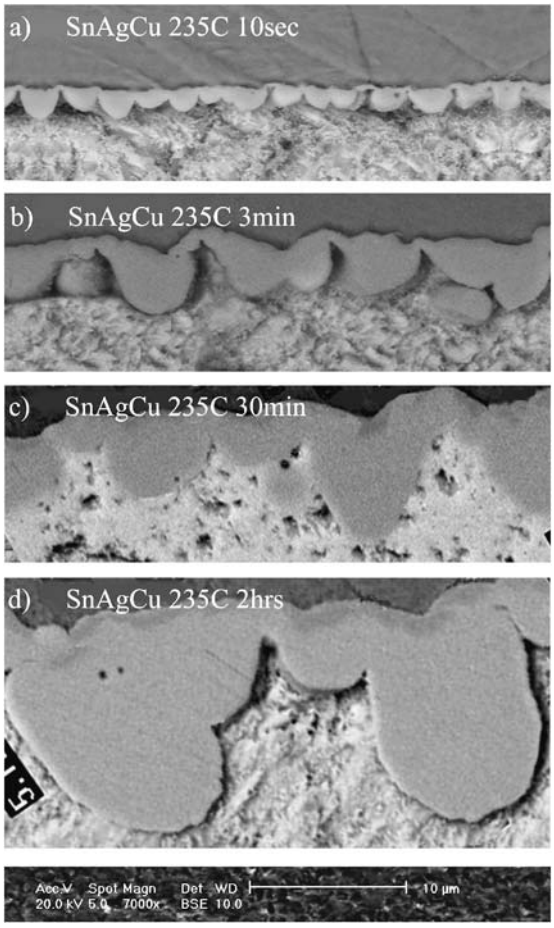


FIGURE 10.17. Comparison of SEM images of the intermetallic layers for the SnAgCu alloy cast at 235°C and held for 10 sec (a), 3 min (b), 30 min (c), 2 hrs (d).

where k_0 is the proportionality constant, Q is the apparent activation energy for IMC layer growth, R is the Boltzmann's constant, and T is the absolute temperature. The apparent activation energy Q was obtained from plots of $\ln(k)$ versus $1/T$.

Figure 10.19 shows the experimental data for IMC layer growth for the SnAgCu (SAC387) alloy as a function of the square root of the holding time for five different temperatures. The linear least squares lines show very good fit to the data, with the average deviation from the experimental data being in the range of $\pm 5\%$. The upper and lower limits of the error bars for the thickness measurements for each time represents the thickness of the biggest scallop and the deepest cusp between scallops, respectively. As can be seen, the minimum and maximum IMC thickness can deviate up to $\pm 75\%$ from the average IMC thickness values calculated from the integration of IMC areas. This difference in IMC thickness can clearly be seen in the SEM images (Figures 10.15–10.18).

Comparison of IMC growth for the SnAg (Sn-3.5Ag), SnCu (Sn-0.9Cu) and SnAgCu (SAC387) lead-free solder alloys at five different temperatures are presented in Figure 10.20. Growth rate constant (k) values obtained from the slopes of the lines fitted to

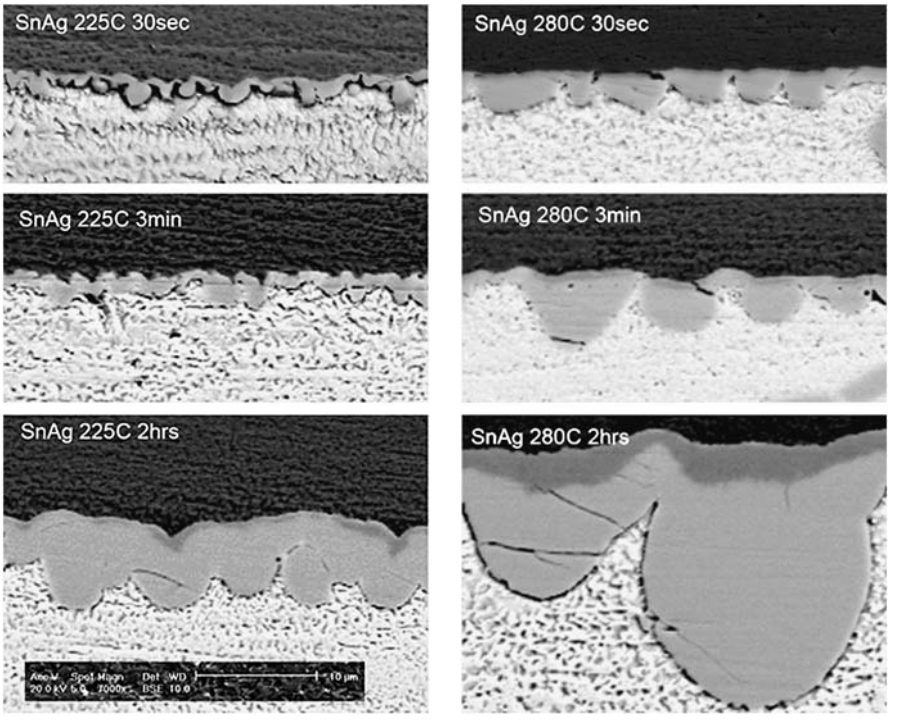


FIGURE 10.18. The IMC thickness comparison for the SnAg alloy cast at 225 and 280°C and held for 30 sec, 3 min and 2 hrs.

TABLE 10.4.
Calculated IMC parabolic growth rates and intercepts for SAC387, Sn3.5Ag and Sn0.7Cu alloys.

	IMC growth rate constant k (μm/sec)		
Temperature, °C	SAC387	Sn3.5Ag	Sn0.7Cu
225	0.096	0.068	–*
235	0.096	0.071	0.077
245	0.120	0.088	0.088
260	0.129	0.094	0.104
280	0.137	0.135	0.122

*225°C testing temperature is below the melting point for Sn0.7Cu alloy.

the data are presented in Table 10.4. It can be seen that IMC growth rate constants for the SnAg and SnCu solder alloys are very similar for all test temperatures, while those for the SnAgCu alloy are about 30% greater. This is consistent with the growth rates being greater for the SnAgCu alloy.

10.2.4.2. Activation Energy and Temperature Effects on Intermetallic Growth on Cu Substrate Comparisons of intermetallic growth for the SnAg, SnCu and SnAgCu alloys at different temperatures are presented in Figures 10.21 and 10.22. As can be seen in Figure 10.21(a), the IMC for the SnAg alloy is thinner than that for the SnAgCu alloy for

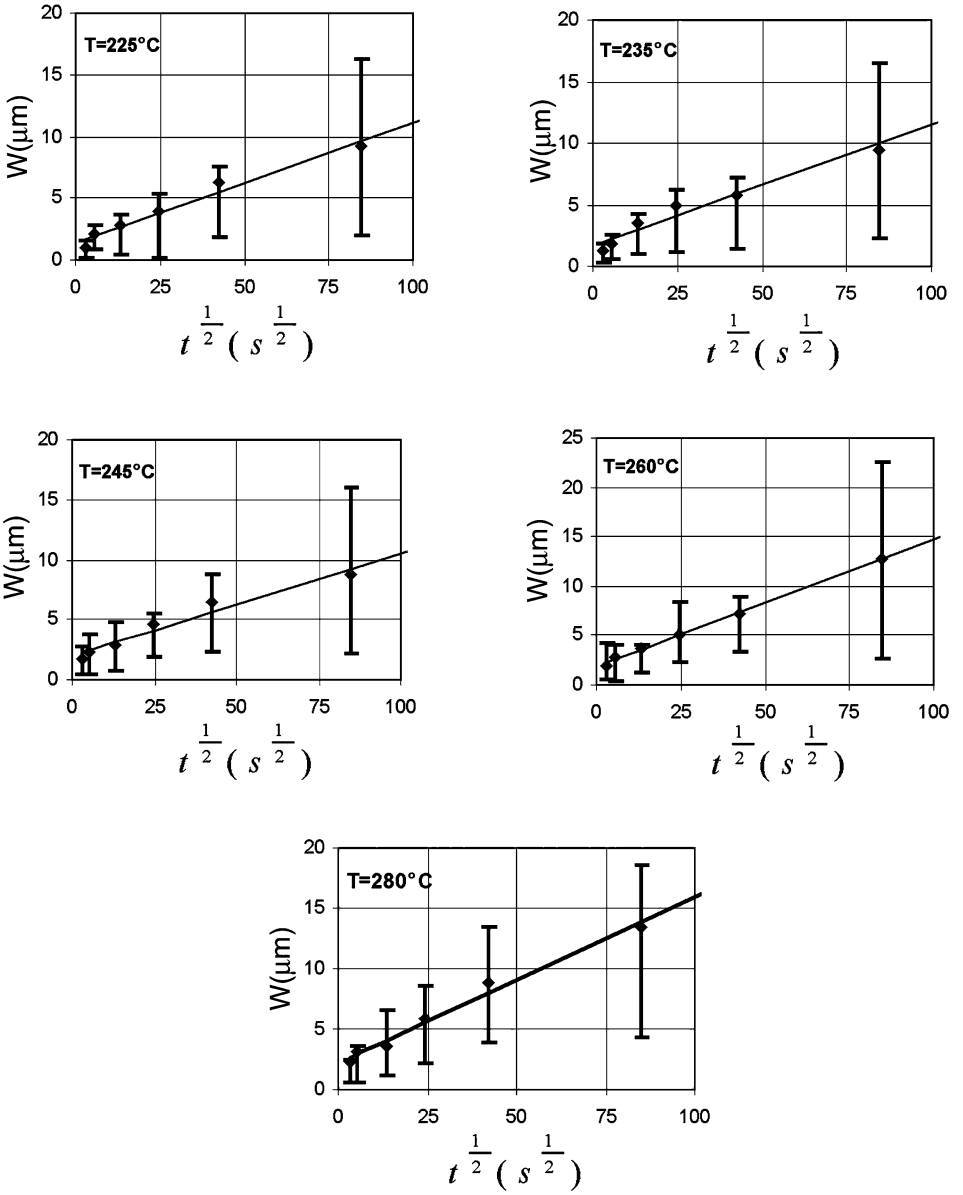
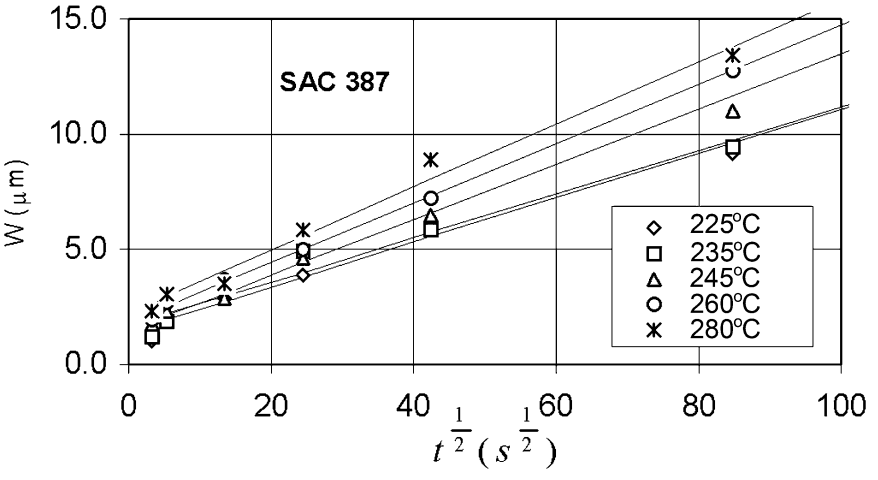
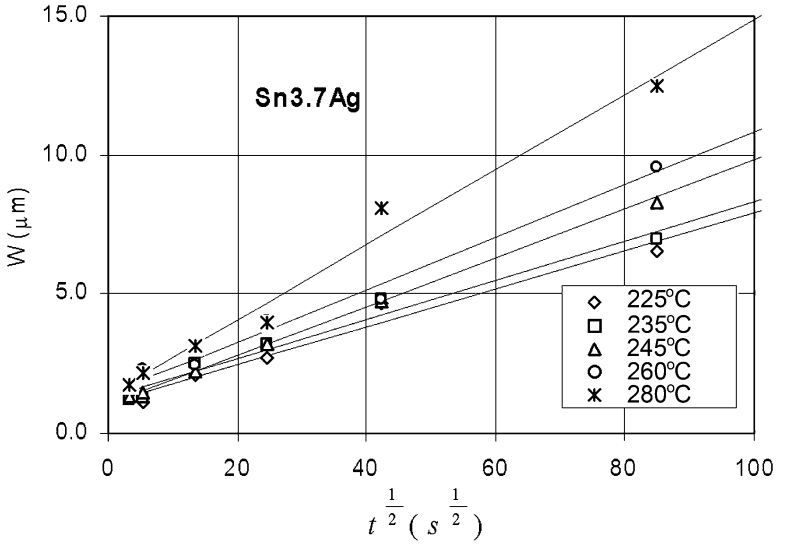


FIGURE 10.19. IMC thickness w as function of the square root of time t for the SnAgCu solder at 225, 235, 245, 260 and 280°C. The straight lines represent mean least square fits of the data.

all exposures. This may result from the SnAgCu alloy being closer to copper saturation at the start of intermetallic growth. It has previously been shown that Cu_6Sn_5/Cu_3Sn intermetallic layer growth for eutectic SnPb solders on copper substrates is slower for growth into solder initially containing no copper than for growth into solder initially saturated with copper. Based on previous studies of copper dissolution into molten eutectic SnAg solder, it is expected that saturation of the copper free solder in this study should require about 60 minutes saturating at 225°C [21].



(a)



(b)

FIGURE 10.20. IMC growth in copper samples for the SnAgCu (a), SnAg (b) and SnCu (c) lead free alloys as function of the square root of time. Markers represent experimental data and solid lines represent calculated regression data.

As can be seen in Figure 10.21(b) with soldering at 235°C the IMC layers for the SnAgCu alloy again grows faster than the IMC layer for the SnAg alloy. Further, the SnCu alloy, which is molten at this temperature, grows IMC at about the same rate as the SnAg alloy. Figure 10.22(a) and Figure 10.22(b) show a continuation of this trend at 260 and 280°C, with the difference between the IMC growth rates for the SnAgCu solders and the SnAg and SnCu solders being the smallest at 280°C. One reason for this latter effect may be that, as the temperature increases, the solubility of copper in the molten solders increases. Thus, the presence of copper in the SnAgCu alloy is further from the saturation level and IMC growth in all of the solders is characteristic of growth into a solder with relatively low

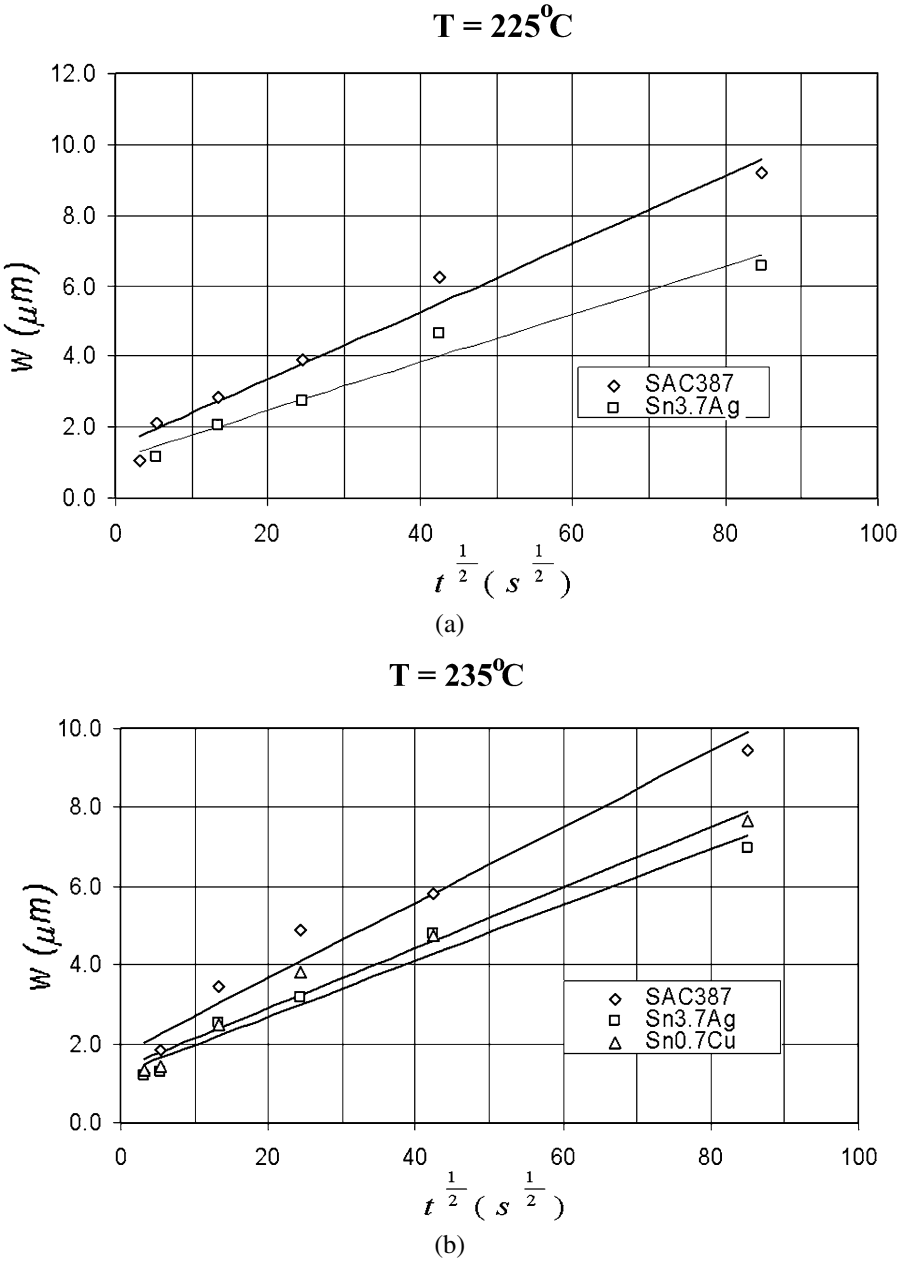
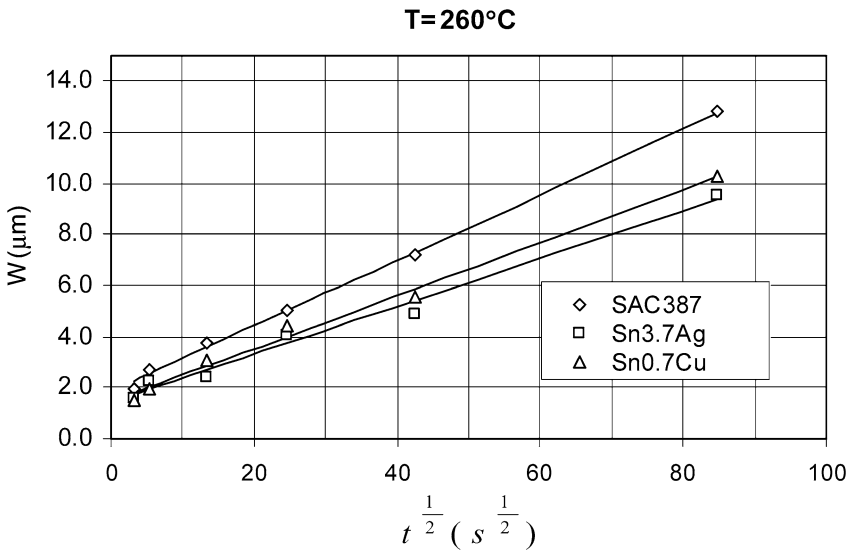


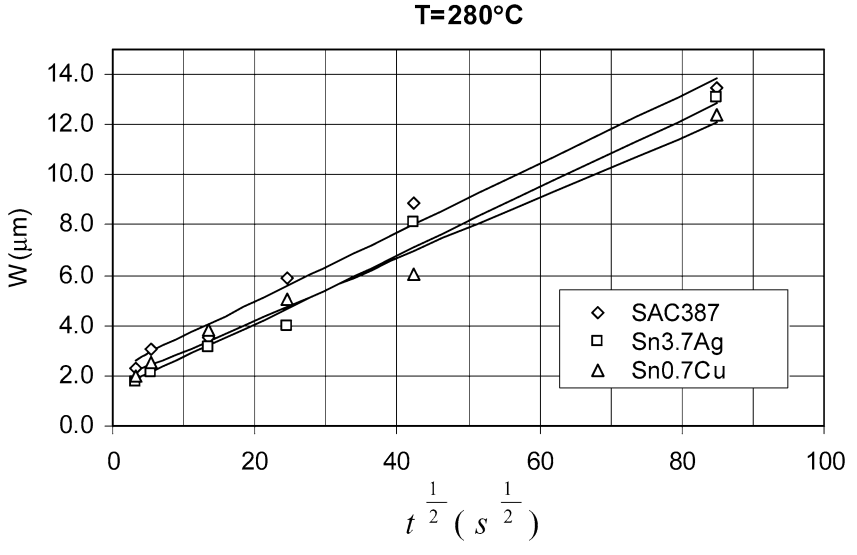
FIGURE 10.21. IMC thickness as function of the square root time for 225°C (a) and 235°C (b). Solid lines represent a calculated fit through the experimental data.

copper content. Following this reasoning, the IMC growth rates for all three solders studied should approach each other as the soldering temperature increases.

Arrhenius plots of the IMC growth constant values given in Table 10.4 are presented in Figure 10.23 and the apparent activation energies are given in Table 10.5. As can be seen, the apparent activation energy for intermetallic growth in the SnAgCu alloy is lower than



(a)



(b)

FIGURE 10.22. IMC thickness as function of the square root of time for 260°C (a) and 280°C (b). Solid lines represent a calculated fit through the experimental data.

TABLE 10.5.
Calculated activation energy for IMC growth of SAC387, Sn3.5Ag and Sn0.7Cu lead free alloys.

Alloy	Activation energy Q, kcal/mol
SAC387	16.44
Sn3.5Ag	25.74
Sn0.7Cu	23.71

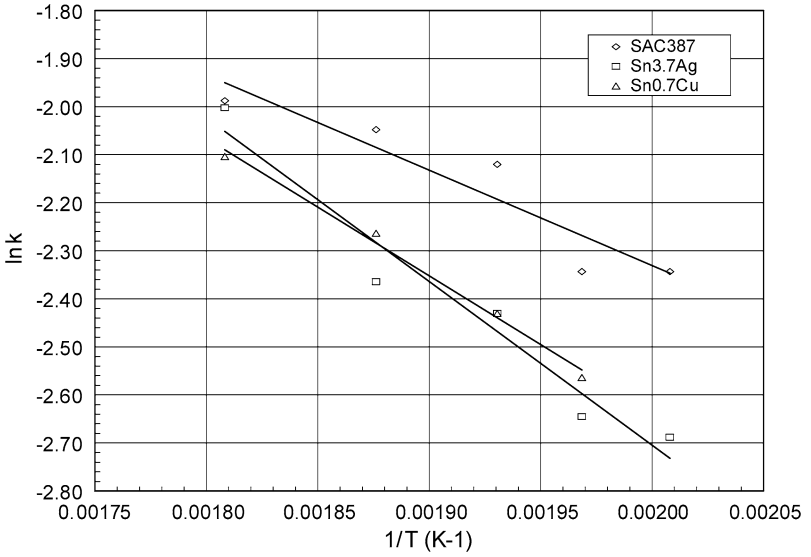


FIGURE 10.23. Arrhenius plot of IMC growth constants (k).

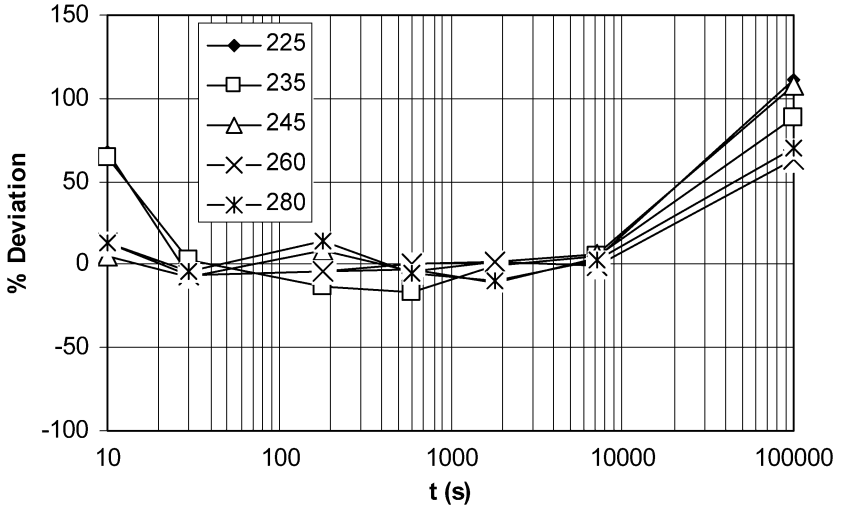


FIGURE 10.24. Deviation of the linear least squares fit values of IMC thickness from experimental data for the SnAgCu alloy at 225, 235, 245, 260 and 260°C.

that for the SnAg and SnCu alloys. This is consistent with the growth rates for this alloy being faster. The values for the SnAg and SnCu alloys are consistent with those reported previously for IMC layer growth for eutectic Sn-Ag solder [22,23].

10.2.4.3. Initial Stage of Intermetallic Growth on Cu Substrate As shown in Figures 10.21 and 10.22, some IMC thickness values, notably those for short and long holding times, deviate from the least square fitted average growth line. Figure 10.24 shows the deviations of experimental thickness values from the least squares fit values for the SnAgCu alloy. It is clear that the deviations are indeed the greatest on a percentage basis at short and

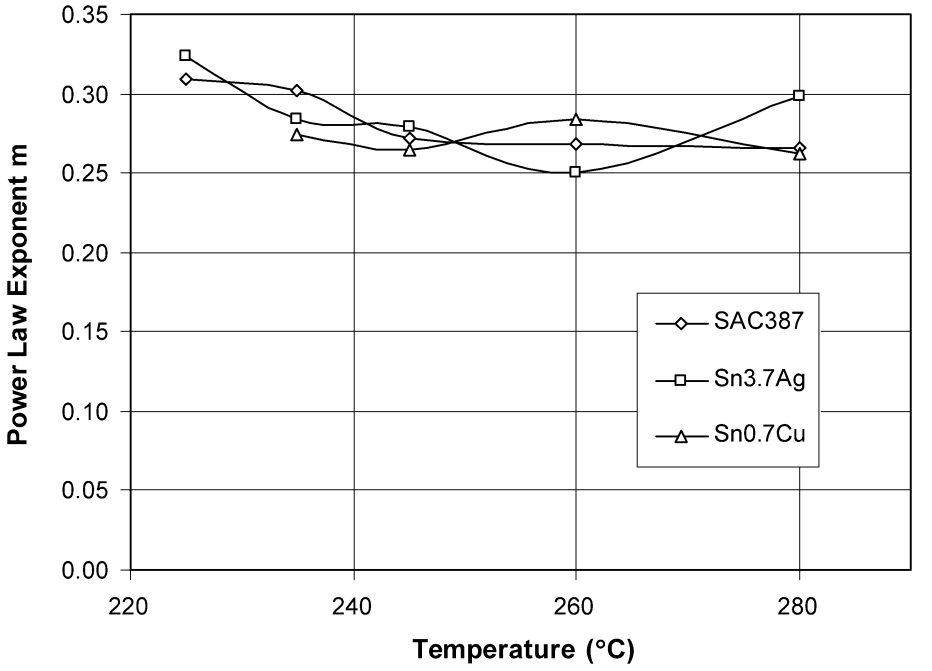


FIGURE 10.25. Calculated power (m) for IMC growth as a function of temperature for the SnAgCu, SnAg and SnCu alloys.

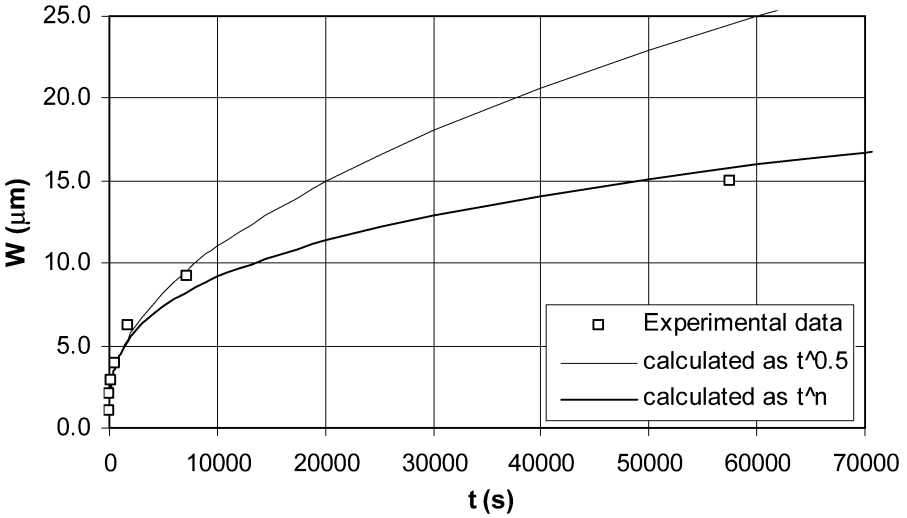
long times and that they are the greatest for the lower temperatures. The deviation at short times is probably the result of several factors. During the early stage of IMC formation, the IMC layer is in most cases growing into solder not saturated with copper. Also, the early stage of layer formation may be reaction controlled instead of diffusion controlled. Diffusion models [24] of steady state intermetallic growth of Cu_6Sn_5 between Cu and pure Sn and Sn containing solders also predict power law dependency rather than parabolic dependency.

The thickness (w) data shown in Figures 10.19–10.22 can also be fit by regression analysis with an equation of the form,

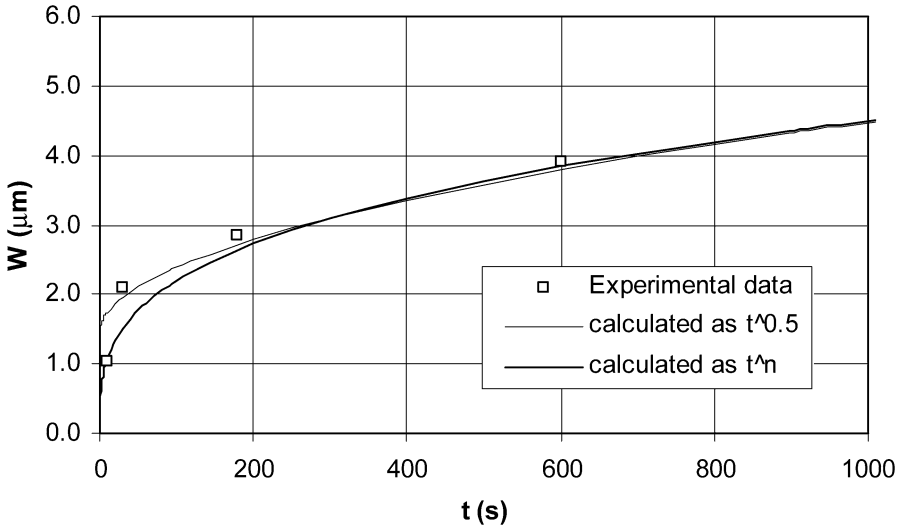
$$w = kt^{m(T)}, \tag{10.5}$$

where k and $m(T)$ are constants resulting from the fits. Figure 10.25 shows $m(T)$ values extracted from curve fits as a function of temperature for all three alloys studied. It can be seen that values of m vary from 0.25 to 0.32. The values are consistent with those predicted by diffusion models [25,26]. Figure 10.26 shows calculated best fit curves based on both Equations (10.3) and (10.5) for IMC growth for the SnAgCu alloy at 225°C. As can be seen, the parabolic ($m = 0.5$) curve fits the data best at time interval of 30 sec up to 2 hrs, while the power law equation works best for short times below 30 sec and again at longer times. Similar results were observed at other temperatures for the SnAg and SnCu alloys.

10.2.4.4. Intermetallic Growth on Nickel and Alloy 42 Substrates Unlike extensive study on intermetallic growth of lead-free solders on Cu substrate, there is relatively little study on other substrates. Experiments for intermetallic kinetics of SAC387 alloy were reported



(a)

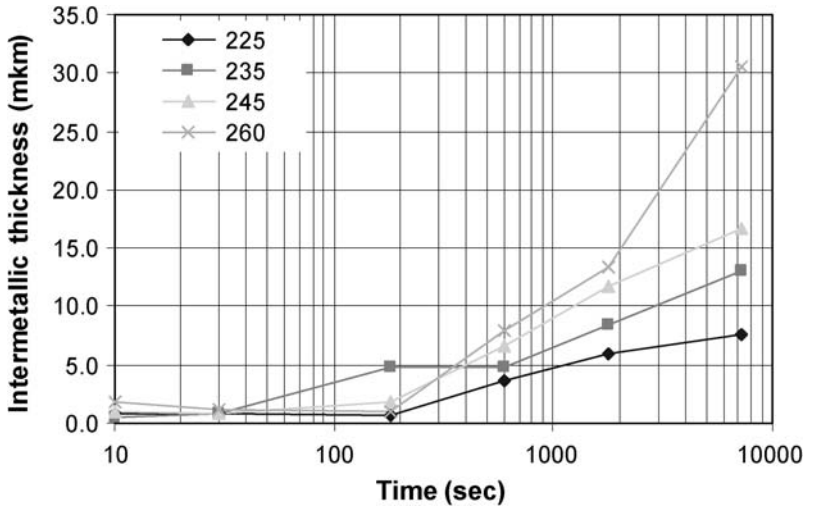


(b)

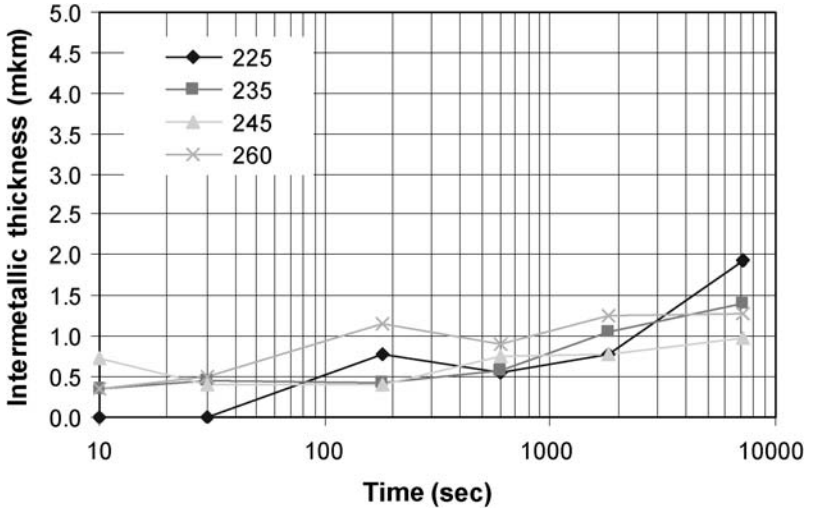
FIGURE 10.26. Comparison of two calculated lines ($w = w_o + kt^{0.5}$ and $w = kt^n$) with experimental data for IMC growth for the SnAgCu alloy at 225°C. The graph (b) presents the magnified portion of the graph (a).

for nickel and Alloy 42 (Fe-42Ni) substrates. The comparison of the intermetallic thickness as function of time at different temperatures is presented in Figure 10.27. SEM images of the intermetallic layers for the Nickel and Alloy 42 are presented in Figures 10.28 and 10.29. It can be seen that intermetallic thickness of alloy 42 samples is significantly thinner compared with those on both copper and nickel substrates.

Kinetic parameters of the intermetallic formation were calculated for Nickel substrate. The values of parameter $m(T)$ in Equation (10.3) for temperature interval of 225°C to 260°C is about 0.475. The calculated activation energy Q for IMC growth for SAC387 alloy on the Nickel substrate is equal to 12.6 kJ/mol, which is 24% lower compared with Q



(a)



(b)

FIGURE 10.27. Thickness of intermetallic as function of time for SAC387 on (a) Nickel, (b) Alloy 42.

for copper substrates. Due to the large scattering for thin intermetallic layer measurement on the Alloy 42 substrate, statistically meaningful growth kinetic equation is not available.

Sn-Ni binary systems have three stable intermetallic compounds, Ni_3Sn (at Ni-rich side), Ni_3Sn_4 (at Sn-rich side), and Ni_3Sn_2 in the between [27]. The additions of Ag and Cu further complicate the metallurgy at the interface for Sn-Ag-Cu solder alloys. Figure 10.30 shows the EDAX phase composition for the IMC for SAC387 on the Ni substrate, indicating a compound of possible $(NiCu)_3Sn_4$ composition. No Ni-rich ICM is detectable in the samples tested.

For Alloy 42 substrate (see Figure 10.31), the alloy itself is a mixture of α-Fe plus Fe_3Ni compound. Since atomic Ni is not available, the intermetallic phase at the interface of SAC387 and Alloy 42 would be mainly Sn-Fe intermetallic compounds. The Fe-Sn

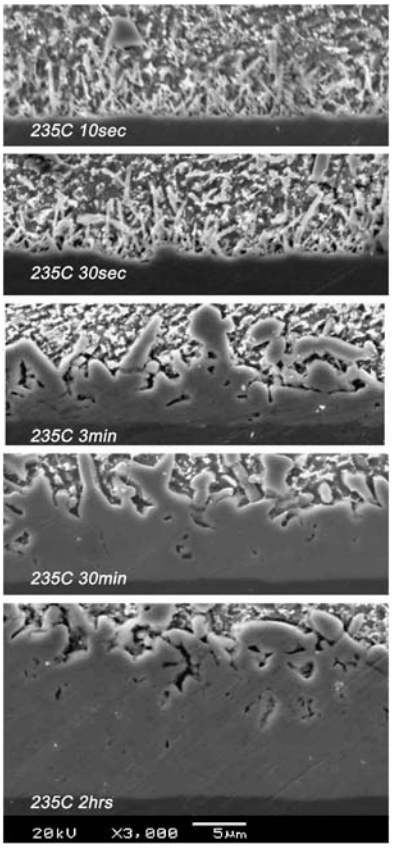


FIGURE 10.28. Comparison of SEM images of the intermetallic layers for the Ni samples cast at 235°C held for 10 sec, 3 min, 10 min, and 2 hrs.

binary system has two stable intermetallic phases: FeSn at the Fe-rich side and $FeSn_2$ at the Sn-rich side. The EDAX analysis clearly shows a $FeSn_2$ intermetallic compound is the predominant ICM for SAC387 on Alloy 42 substrate.

10.2.4.5. Summary on Intermetallic Growth Kinetics In summary, the IMC formation and growth at liquid Pb-free solders and solid interface shows the following characteristics:

(1) Morphologically the formation and growth of IMC layers between all three Pb-free alloys and copper substrate are identical, with formation of a thin layer of Cu_3Sn adjacent to the copper substrate and a scalloped layer of Cu_6Sn_5 between Cu_3Sn and the molten solders.
(2) The IMC layer formed with the SnAgCu alloy grows faster than those formed with the SnAg and SnCu alloys on Cu substrates.
(3) The faster growth of the IMC layers for the SnAgCu alloy may be the result of the layers growing into a solder initially containing more copper for the SnAgCu alloy.
(4) The thickening of the IMC layers in all alloys can largely be described with a parabolic ($m = 0.5$) growth equation on both Cu and Ni substrates, indicating a diffusion controlled growth process.

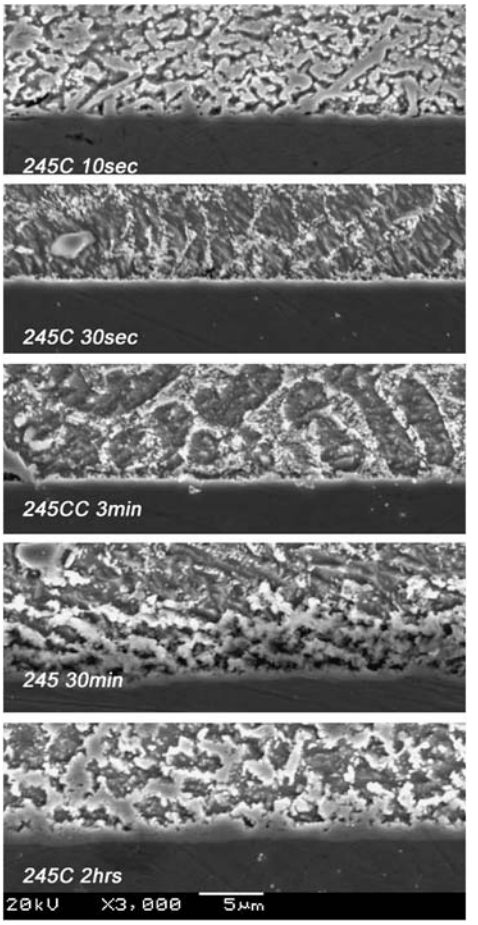
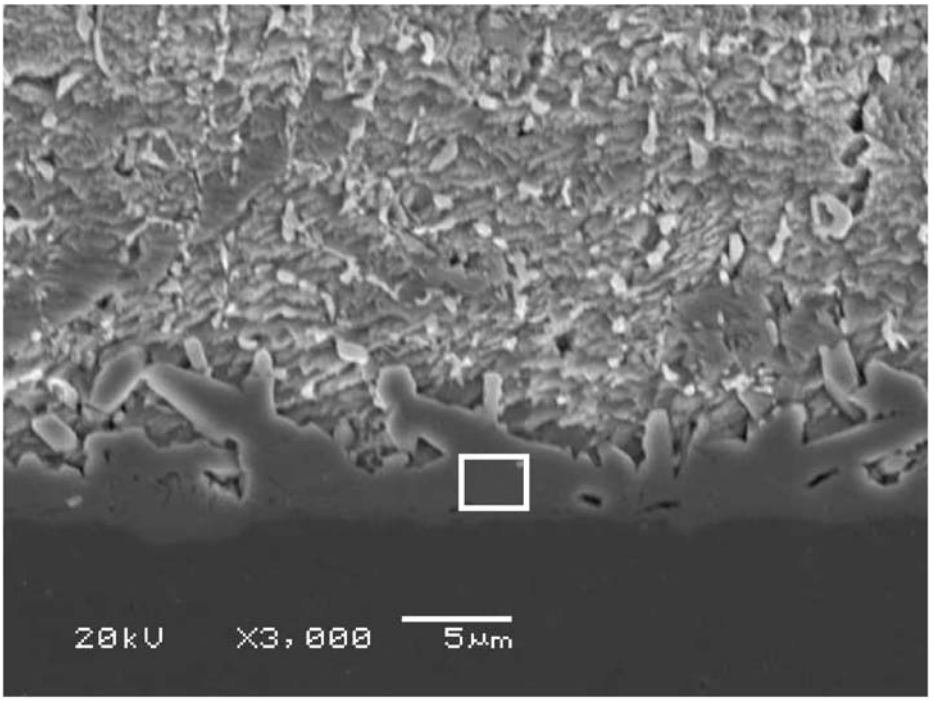


FIGURE 10.29. Comparison of SEM images of the intermetallic layers for the alloy 42 samples cast at 245°C held for 10 sec, 3 min, 10 min, and 2 hrs.

(5) Among Cu, Ni and Alloy 42, Cu substrates grow intermetallic layers fastest, followed by Ni, and Alloy 42.

10.3. LEAD-FREE SOLDERING PROCESSES AND COMPATIBILITY

As Pb-free soldering moves from the laboratory to the manufacturing floor and the worldwide electronics industry gradually implements Pb-free soldering in printed circuit board (PCB) assemblies, it becomes clear that lead-free processing has many unique requirements during to the metallurgical and chemical attributes of Pb-free solders, fluxes and pastes. Volume manufacturing with lead-free solders is a complex undertaking for the industry. As shown in Figure 10.32, there are many compatibility issues needed to be concerned in volume manufacturing of lead-free solder PCB assemblies, including processing, materials, components, equipment, design, quality control and reliability.

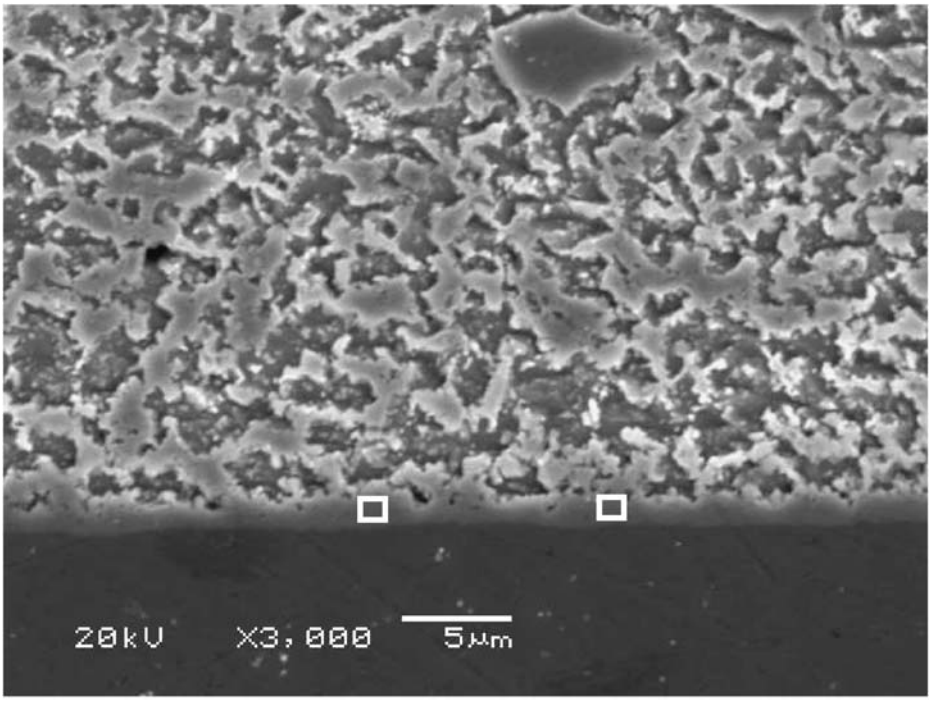


Element	Wt %	At %
AgL	0.86	0.79
SnL	64.27	48.78
NiK	8.41	12.95
CuK	26.46	37.48

FIGURE 10.30. Nickel sample (254°C, 30 min, ×3000).

10.3.1. Lead-Free Soldering Materials

10.3.1.1. Lead-Free Solder Alloy Compositions As discussed above, the industry is now converging on Sn-Ag-Cu (SAC387, SAC396 and SAC305) ternary eutectic alloys for reflow processes, and Sn-Ag-Cu or Sn-Cu eutectic alloys for wave soldering. It is generally believed that the different variations of the Sn-Ag-Cu alloys, with Ag content from 3.0% to 4.0%, are all acceptable compositions. The Sn-Cu alloy has been found to be inferior to Sn-Ag-Cu in terms of wettability, dross formation, and reliability; however, its much lower cost as compared with Sn-Ag-Cu alloys makes it an attractive alternative alloy for wave soldering, especially for cost sensitive products. The steep slope of the Sn-Cu binary phase diagram (approximately 30°C change in liquidus for every percent change in composition, roughly 20 times the slope for Sn-Pb) suggests that the chemical composition of the Sn-



Element	Wt %	At %
AgL	0	0
SnL	80.11	65.83
FeK	16.17	28.24
NiK	1.77	2.94
CuK	1.94	2.98
Total	100	100

FIGURE 10.31. Alloy 42 sample (245°C, 2 hrs, ×3000).

Cu solder pot needs to be closely monitored and controlled. Several variants of the Sn-Cu alloys have also been introduced, including Ag, Ni, and others as alloying elements.

During the transition, many products may be assembled with lead-free solders with Pb-containing component termination. For wave soldering, the level of Pb in the solder pot, due to dissolution of Sn-Pb plated component finishes, needs to be monitored. The impact of Pb in Pb-free solders on long-term reliability is not clearly known at the current time. Preliminary studies have indicated that the impact varies with the amount of Pb in

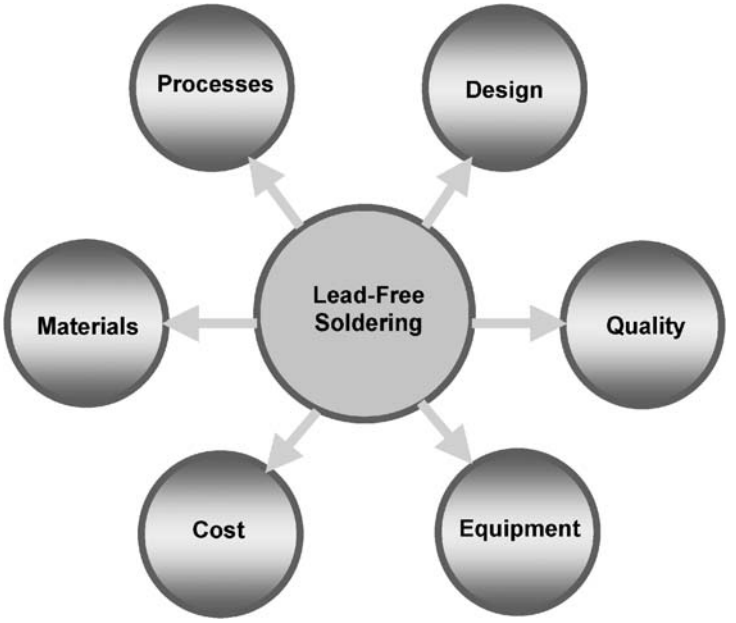


FIGURE 10.32. The compatibility issues with Pb-free soldering technology.

the solder joints, and the impact may be the greatest when the amount of Pb is within some intermediate range, because of the formation of segregated phases (e.g., coarse Pb grains) in the last-to-solidify inter-dendritic Sn grain boundaries, where cracks may initiate and propagate under cyclic loading. It has been shown [28] that 2%–5% Pb can be detrimental to the fatigue life of Pb-free solder joints.

10.3.1.2. Lead-Free Flux and Pastes Earlier attempts to simply mix the no-clean flux (developed for Sn-Pb alloys) with the Pb-free alloys yielded miserable results. The no-clean flux needs to be re-formulated for the Pb-free alloys in order to accommodate the characteristics of the Pb-free alloys. The chemical reactions between the flux and the solder alloys in the paste affect the rheology characteristics of the solder pastes (which is critical for printing performance). The difference in the density between the Pb-free solder alloys and the Sn-Pb alloys means that the metal loading of the solder paste needs to be different. The higher soldering temperature needed for the Pb-free solders also requires greater stability of the flux chemistry at higher temperatures. The performance of the flux residues after reflow, in terms of in-circuit test (ICT) probeability and electromigration, is also an important consideration. Similarly, no-clean and VOC-free fluxes need to be formulated specifically for lead-free wave soldering. Water-soluble fluxes for lead-free solder pastes and wave soldering applications are also needed for certain applications.

10.3.2. PCB Substrates and Metalization Finishes

As the soldering temperature increases, the CTE (coefficient of thermal expansion) mismatch between the laminate material, the glass fiber, and the Cu, will exert greater stresses on the Cu, potentially causing failures by cracking the Cu in the plated vias and holes. This is a rather complex issue because it depends on a number of variables, such as

the thickness of the PCB, the laminate material, the soldering profile, the Cu distribution, the via geometry (such as aspect ratio), etc. Other issues, such as delamination and blistering at higher soldering temperatures, are also areas to be studied. Much work is needed to determine under what conditions, alternative laminate materials (such as high T_g, low CTE, and high decomposition temperatures) to the traditional FR4 may be needed for Pb-free soldering. This is not to say that lower cost materials (such as CEM, FR2, FR4, etc.) can not be used with Pb-free soldering. In fact, such applications do exist in volume manufacturing with lead-free manufacturing. The situation needs to be examined on a case-by-case basis.

The search for PCB surface finishes as alternatives to HASL (hot air solder leveling) has been on-going for many years, primarily because of the inherent inconsistency in the quality of the HASL finish. For example, the thickness (and therefore solderability) of HASL is difficult to control. In areas with a very thin layer of HASL, consumption of Sn by the formation of Sn-Cu intermetallic compounds will render the areas non-wettable. The HASL finish is typically non-flat (with a dome shape), making it difficult to deposit a consistent amount of solder paste during solder paste printing and difficult to place fine pitch (<25 mil) devices. The HASL process itself is not as clean and easy to control as other plating processes. The move towards Pb-free soldering has provided the additional impetus towards Pb-free surface finish alternatives.

Pb-free HASL, using Pb-free alloys (such as Sn-Cu, Sn-Ag, or Sn-Ag-Cu alloys) in place of Sn-Pb, is commercially available. Electroless nickel and immersion gold (ENIG) provides good solderability and contact/switch interfaces for most applications; however, tight plating process control is necessary in order to prevent the occurrence of catastrophic "black pad" failures. For higher end applications, electrolytic nickel (Ni) and gold (Au) provides a more reliable surface finish. Care must be taken to limit the thickness of Au, as too much Au dissolved in the solder joints may cause "Au embrittlement," even though the Sn-Ag-Cu alloys are less sensitive to Au embrittlement than the Sn-Pb alloys. Immersion silver is a more recent and less costly alternative. Its solderability, ICT probe-ability, and contact/switch pad performance, are not as good as Ni/Au, but are adequate for most applications. For immersion Ag, the exact chemistry, thickness, surface topography, as well as the distribution of organic constituents within the Ag layer, should be carefully selected and specified [29–32].

For Pb-free soldering, immersion tin and Ni/Au surface finishes provide the best wetting results on fresh boards, followed by immersion Ag and organic solderability preservative (OSP). However, in a humid condition, the wetting of the immersion Sn finish degrades the fastest, whereas the wetting of the Ni/Au remains excellent as evidenced in Figure 10.33 to Figure 10.34, which show the relative maximum wetting force on 1 inch by 0.5 inch wetting coupons with Sn, OSP and Ni/Au finishes before and after a one-week aging at 85°C/85%RH. The Ni/Au finish remains excellent with variety of pre-conditioning treatments and heat exposures, while OSP loses wettability rapidly with exposure to a high temperature environment, as shown in Figure 10.35. Fresh immersion Ag boards can withstand up to four lead-free reflow cycles before the final reflow soldering process, and at least two reflow cycles before wave soldering process. On the other hand, neither fresh nor aged immersion Sn finished boards can withstand multiple lead-free reflow cycles or a reflow cycle prior to wave soldering process without significant degradation in wettability.

10.3.3. Lead-Free Soldering Processes

10.3.3.1. Reflow Soldering In terms of printability, tackiness, slump, and solder balling, there is no clear and consistent difference between Sn-Pb and Pb-free solder pastes, because

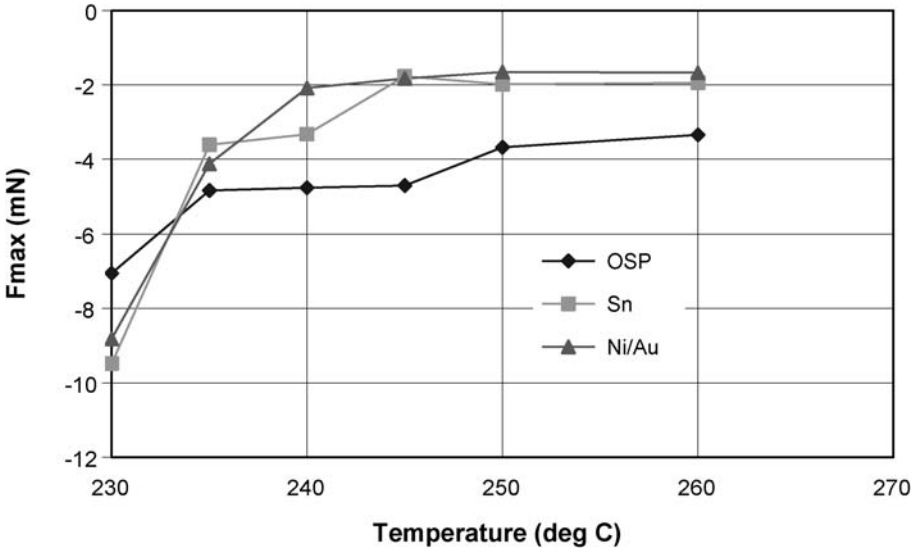


FIGURE 10.33. Relative maximum wetting balance force for virgin coupons with OSP, Sn and Ni/Au finishes (coupons size 1 inch by 0.5 inch).

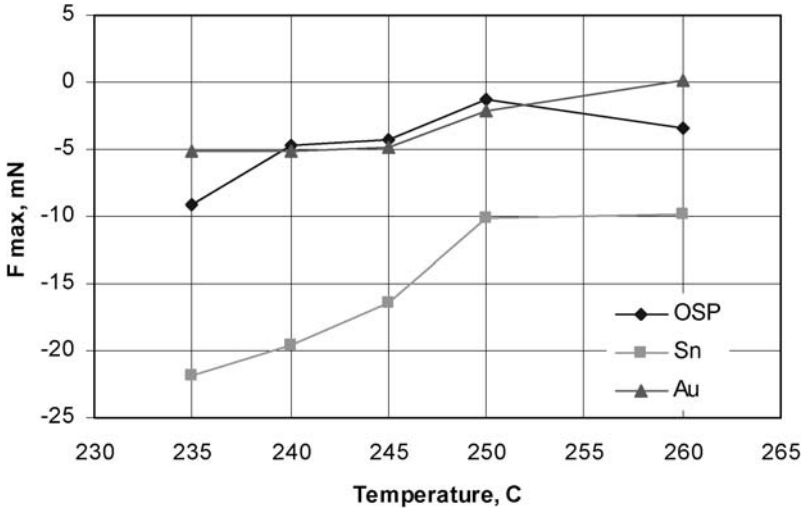


FIGURE 10.34. Relative maximum wetting balance force for coupons after aging at 85°C/85%RH.

the performances depend mainly on the solder paste formulation, not directly on the solder alloys. Very clear and consistent differences have been observed, however, in wettability between Sn-Pb and Pb-free solder pastes. In general, as expected and discussed in the previous section, the wettability of Pb-free solder pastes are not as good as Sn-Pb eutectic solder pastes on most metalization finishes [33,34]. The current commercially available Pb-free solder pastes exhibit very limited spreading on OSP during reflows up to 250°C. Exposed corners on OSP boards after reflow are quite common, unless overprint or round corner pads are used. The key parameters for the reflow profile are the peak temperature and time above the liquidus. Adequate super-heat reflow temperature above the liquidus is needed

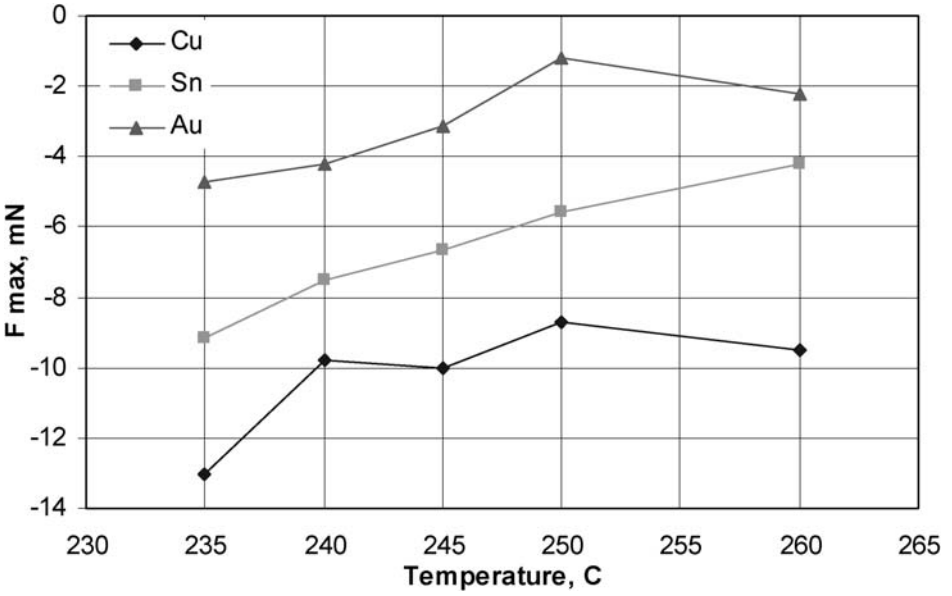


FIGURE 10.35. Relative maximum wetting balance force for coupons after aging at 130°C for 24 hrs.

for the Pb-free solder to melt, wet and spread, metallurgical reactions with the Cu or other metallic surface on the pads and the component termination to form strong intermetallic bond when cooled and solidified [35]. Typically, 30°C superheat (above the melting temperature) is desired. For Pb-free soldering, because of concerns about the thermal stability of the components, efforts have been made to minimize the soldering temperatures. For SAC alloys with the eutectic temperature at 217°C, many experimental studies show that the minimum reflow peak temperature should be 235°C for large volume manufacturing, taking into account process robustness, yield, variety of component finishes, oven thermal stability and tolerance, etc. The dwell time (or time above liquidus) is typically 40–90 seconds. The reflow profiles may be straight ramp, or with a pre-heat plateau for the purpose of homogenizing the temperature distribution across the board, with soldering environment either in air or N_2.

10.3.3.2. Wave Soldering For wave soldering with Pb-free solders, a higher solder pot temperature, typically 255–270°C, will be required. Flux application (spray or foaming), flux amount, preheat temperature, and dwell time needs to be optimized for each particular Pb-flow wave soldering application. A longer pre-heat may be needed in order to keep the thermal shock (difference between the preheat and peak temperatures) below 100°C to protect ceramic components. Dual wave soldering, already popular in Sn-Pb soldering, will still be used for Pb-free solders, and inert (N_2) atmosphere may be used to improve yield and reduce dross.

The amounts of dross formed with SAC and Sn-Pb solders have been found to be very similar, at the same temperature, for the same duration, and under the same atmosphere, while the Sn-Cu solder forms considerably more dross. Just as with the Sn-Pb solder, N_2 helps reduce dross formation for Pb-free solders. The overall influences on the yield and quality for wave soldering of other process parameters, such as conveyor speed, dwell time and contact length, component orientation and soldering direction (parallel or perpendic-

ular), are similar for Pb-free and Sn-Pb solders [36,37]. For Pb-free wave soldering, the chemical composition of the molten solder in the pot needs to be closely monitored. For the Sn-Cu solder, the liquidus temperature will change by as much as 6°C when the Cu composition changes by 0.2%; such a change may cause significant change in the wave dynamics and the soldering quality.

10.3.3.3. Rework and Repair Rework for Pb-free solders has been found to be generally more difficult, because the Pb-free solder alloys typically do not wet or wick as easily as the Sn-Pb solder. In spite of the difficulty, successful rework methods (both manual and semi-automatic) have been developed with Pb-free solders for all types of components (discrete components, area array packages, etc.). The soldering parameters should be adjusted to accommodate the higher melting temperature and less wettability of the Pb-free solder. Care should also be taken to minimize any potential negative impact of the rework process on the reliability of the components and the PCBs. Surface insulation resistance (SIR) tests need to be performed to ensure the compatibility between the reflow/wave solder flux and the rework flux.

The issue of "component mixing" or cross-contamination warrants special attention, especially during the transition period. If a Pb-bearing board is to be repaired (for example for warranty repair at some future time) with the Pb-free solder, the repair temperature impact on the Pb-bearing components (especially plastics package parts) and PCBs would be a concern, with further reliability implication of non-homogeneous mixing of Pb-free solders and the original Sn-Pb solder. Careful consideration should be given to the use of area array packages with Pb-free balls to repair a Pb-board. In this case, if the temperature is not high enough, reliability concerns may also arise.

10.3.4. Components for Lead-Free Soldering

10.3.4.1. Termination Metallurgy Traditionally, most of the components have Pb-bearing lead finishes or terminations (for chip capacitors/resistors), or Pb-bearing solder balls, to provide and to preserve solderability. A number of Pb-free component termination finishes have been evaluated and used over the years, including Ni/Pd, Ni/Au, and Ni/Pd/Au for lead frames (such as QFP). More recently, Sn-Bi, Sn-Cu, and Sn-Ag-Cu-Bi coatings were proposed and evaluated for Pb-free component lead finishes and terminations with some successes. Due to its processing compatibility to current Pb-bearing finishes/terminations and good solderability, matte Sn as component termination finishes have been the preferred termination metallurgy for many component designers and manufacturers. Although it has been used for passive components (such as chip capacitors and resistors) for many years, pure tin for Pb-free fine pitch component lead finishes have waved a red flag for many OEMs with higher reliability requirements due to ramifications related to the possible Sn whiskering.

10.3.4.2. Tin Whiskering The issue of Sn whisker formation and growth on pure Sn finish has generated a great deal of concern, especially for fine pitch QFPs and for high reliable and long service life products such as those used in avionics, aeronautics, satellites, missiles, mission-critical storage systems and servers, as well as medical systems. Tin whiskering currently is one of most difficult issues for the implementation of a complete Pb-free transition. Although, it is generally agreed that the root cause of Sn whisker growth is the compressive stress which builds up in the Sn coating after plating, the origins and magnitude of such stresses and kinetics of Tin whiskering are still not fully understood. It has

been reported in the literature that whiskering phenomenon is dependent on grain size, organic (carbon, sulfur, oxygen) and inorganic inclusions in the tin coatings, plating conditions (current density, pulse, temperature, time), residual stress, external stress, thermal stresses, substrate materials, surface oxides and roughness, environments, even electric and magnetic field. Field failures attributable to tin whiskers have resulted in many million, if not billions, of dollar loss in mission critical communication satellites. The tin whiskers have been described in the press [38] to be a potential problems similar to Y2K issue, should Sn-coated components go into every electronic and semiconductor products due to the requirements for elimination of lead in any form in electronic products.

In terms of tin whiskering mechanisms, it has been reported that whiskers be controlled by the Sn/Cu interfacial reactions, such as intermetallic growth, inter-diffusion, which lead to buildup of compressive stress in the tin coatings either through volume change of intermetallic formation and growth [39], or excessive Cu diffusion into the Tin at the interface. If the oxide films on the tin coating surface is porous and broken, whiskers will initiate and grow to release the compressive stresses. However, these theories fail to explain why a small amount of alloying elements, such as Pb and Bi, is able to eliminate or significantly reduce the tin-whiskering since there is no noticeable change in the Sn/Cu interfacial reactions.

In order to understand the fundamental physical and mechanical metallurgy for Tin whisker initiation and growth, it has been proposed recently to use nano-indentation to induce tin whiskers on pure tin coated component leads [40]. The indentations generate variable small scale compressive stress–strain field with sharp stress gradient. The nucleation and growth of whiskers were monitored *in-situ* with ESEM. Finite element analysis was carried out to theoretically calculate the stress/strain distribution around the indentations and to identify the stress level at which whiskers nucleate. Experimental and theoretical calculation results show that whiskers form at a certain stress level and locations. This suggests that there might exist a critical stress that governs the whisker initiation. It is believed that establishment of a quantitative relationship between stress level and whisker formation/growth could lead to a breakthrough in risk and reliability assessment with pure Tin application in the electronic industry and in safeguard for smooth Pb-free transition. In additions, a joint research project [41] was established recently to use vapor deposition techniques to make thin Cu and Sn coatings with a variety of thickness on silicon wafers, then use a curvature based optical method (Multi-beam Optical Stress Sensor—MOSS) to measure stress evolution in the Sn film following deposition and during aging. Any stress build up due to any reason in the thin film system will cause the whole thin silicon wafer to bending (change in curvatures). Coupled with study of the microstructure evolution, interfacial reactions, and monitoring whisker initiation and growth kinetics, we might be one step closer to fully understanding the intrigue tin whiskering phenomenon.

The current strategy to reduce or mitigate the risk of Sn whisker is generally based on empirical observations, including:

(1) Matte Sn instead of bright Sn;
(2) Sn alloys with a small percentage of alloying elements, such as Bi, Ag and Cu;
(3) Ni barrier layer underneath Sn coating;
(4) Reflowing Sn coating or annealing;
(5) Conform coating or strip of Tin coatings.

The effectiveness of these and other mitigation techniques are still under evaluation by the industry. Accelerated testing methods and acceptance criteria for pure tin coated components have been proposed by JEDEC and IPC recently [42]. However, since it is not

able to definitely assess the accelerating factors for most proposed acceleration and acceptance tests, the real risk associated with Tin whiskers in the vast different field application conditions will still be unknown. It is hoped that as the tin whiskering mechanisms are understood better in the future, with better measuring and monitoring methodologies and systems being developed, the real solutions may be eventually developed to eliminate or reduce the tin whiskering problems.

10.3.4.3. Component Thermal Stability Different components have different sensitivities to the soldering temperature. For example, ceramic resistors and especially capacitors are very sensitive to the ramp rate but are not so sensitive to the actual temperature. Aluminum electrolytic capacitors, on the other hand, are extremely sensitive to temperatures. Connectors and some of the plastic package parts may also exhibit increased failures (such as delamination, pop-corning, deformation, etc.) at higher soldering temperatures. Roughly, for every 10°C increase in temperature, the MSL (moisture sensitivity level) degrades by one level.

Because of the temperature impact on components, conscious efforts have been made to minimize the Pb-free soldering temperatures. However, as discussed above, the soldering temperature have to be high enough to offer a robust process window to enable good yields for large volume production for a variety of products. For reflow soldering, assuming the minimum peak reflow temperature to be 235°C for a specific product with a specific solder paste, the actual maximum temperature will be still dependent on the board size, thickness, layer count, layout, Cu distribution, component size and thermal mass, thermal capacity of the oven and certain unavoidable process variations. Large thick boards with large complex components (such as CBGA, CCGA, etc.) typically have temperature delta as high as 20–25°C [43,44]. Rework is another process which needs to be conducted at an even higher and non-uniform elevated temperature environment. For wave soldering, the solder pot temperature is typically higher than the reflow temperature.

When all of the application requirements are taken into consideration, 260°C peak temperature has been proposed as the temperature required for components for Pb-free soldering. It is recognized, however, that for certain products (such as small/medium boards), which typically have a smaller temperature delta, a lower maximum temperature may be adequate. The actual component body temperature may be different from the temperature measured on the board, and different components may have different temperatures depending on component thermal characteristics and location on the board. Overall, most components may have to be qualified for temperatures as high as 260°C.

10.3.4.4. Component Forward/Backward Compatibility Sn, Ni/Pd and Ni/Pd/Au finished components are forward compatible with Pb-free soldering, as well as backward compatible with the Sn-Pb solder in terms of soldering processes. This makes it much easier to manage production lines with Sn-Pb solders and Pb-free solders within the same factory during the transition. The backward compatibility of SAC solder balls with Sn-Pb soldering, however, is very much questionable. This is primarily due to the fact that the SAC alloys may not always completely melt during reflow with the Sn-Pb solder, typically at reflow peak temperatures between 205–225°C (or even as low as 200°C in extreme cases). As such, there will be little or no self-alignment, which is critical especially for finer pitch area array packages, with coplanarity issues further aggravating the situation due to the lack of collapse. Further, very little mixing takes place leading to grossly segregated and non-homogeneous solder microstructures. Poor interfacial bonding and increased voids are some of the issues, which may render the SAC balled area array packages "incompatible" with the current Sn-Pb soldering [45,46]. When the reflow temperature is high enough

(>230°C) and self-alignment does occur, alloy mixing and composition homogenization takes place between the SAC and Sn-Pb, and the reliability of the interconnects using Sn-Pb solder paste and SAC balls can be as good as that using Sn-Pb solder paste and Sn-Pb balls for area array packages.

10.3.5. Design, Equipment and Cost Considerations

It is anticipated that no major changes in design rules will be needed when switching to Pb-free soldering. For wave soldering of through-hole components, some changes in the design may be necessary to accommodate the difference in the physical properties between Pb-free and Sn-Pb solder alloys. The general guidelines, such as board orientation relative to the soldering direction, still apply to Pb-free wave soldering [36,37].

As with the Sn-Pb solder, it is beneficial to optimize board layout and Cu distribution in the PCBs in order to minimize the temperature delta across the board. This is especially important for Pb-free soldering in order to minimize the temperature impact on components.

For reflow ovens, the ability to minimize the temperature delta for large complex boards is a key differentiator for Pb-free soldering. It has been suggested that in order to minimize the temperature delta, the convey speed should be lowered for the Pb-free reflow process. This, however, is limited, not only by the throughput requirement, but also by the durability of the flux and by the dwell time.

Wave soldering machines need adequate pre-heating capacity in order to keep the thermal shock below 100°C. Solder pot erosion is another consideration for Pb-free wave soldering, and special materials are needed for the solder pot and the other equipment components which are in contact with the molten Pb-free solder. Most of the rework equipment for Sn-Pb can still be used for the Pb-free soldering.

Even though the types of defects for Pb-free solders, for both reflow and wave soldering, are the same as for Sn-Pb, it takes considerable efforts in order to achieve the same yield, especially for wave soldering. This is especially true during the transition when every party of the supply chain has to go through the learning curve. AOI, AXI, and ICT parameters will need to be adjusted for Pb-free solder boards to pick up defects due to poor printing and poor wetting in Pb-free soldering processes.

It is generally accepted that the IPC 610 standards are still valid for Pb-free solder PCBA for workmanship acceptance. Discussions are currently on-going to further refine the standards for Pb-free soldering. Operator (and AOI) training is needed for Pb-free soldering because the appearance of the Pb-free solder joints are generally more dull and grainy, and less shiny, than the Sn-Pb solder joints. As discussed previously, this difference in appearance is determined by the metallurgy of the solder alloys and is not a reflection of the workmanship. The Sn-Pb solder solidifies as a typical eutectic microstructure. The SAC alloy, even though it is a eutectic alloy, solidifies as an off-eutectic microstructure, under typical soldering conditions, due to non-equilibrium solidification. Sn dendrites with shrinkage, which are formed as a result of the non-equilibrium solidification, create the grainy and dull appearance of Pb-free solders joints.

There are no simple ways to assess the precise cost impact for converting to Pb-free solders. The field is still dynamic and fluid, and volume dependency is an important factor. However, it is certain that Pb-free solder pastes cost more than the Sn-Pb solder paste. The price differential is expected to decrease as the volume increases. SAC bar solder and wire-core solder will cost several times more than the Sn-Pb solder because of the higher

metal cost for Sn and Ag. Components and PCBs for higher soldering temperatures may also add to the cost. Any special handling or baking for components will also increase the cost. In terms of oven energy consumption for reflow, study has shown that once the oven has reached steady state, the difference in power consumption between Sn-Pb and Pb-free soldering is about 10%–20% for boards of typical sizes and complexities [43]. Yield is another factor which can significantly impact cost. Equipment upgrading, if needed, is also a contributor to initial cost increase. Other cost factors are due to the complexity in managing the transition, including training, materials handling and tracking.

10.4. RELIABILITY OF PB-FREE SOLDER INTERCONNECTS

There are many published reliability results for Sn-Pb solder joints based on the field data, power/temperature cycling tests, mechanical shearing, bending, and twisting tests, shock and vibration, and electrochemical tests [47–50]. However, it is not the case for Pb-free solder joints because of limited volume production and limited variety of products manufactured with Pb-free solders.

Many environmental stress factors, such as temperature, voltage, humidity, corrosion, current density (electromigration), and mechanical loads, can lead to Pb-free solder joint failure. The most common failure modes in practice are overload and thermal fatigue. Overload failure occurs whenever the stress in the solder joints is greater than the short time mechanical strength of the solder alloys, such as in impact tests or with improper handling. On the other hand, thermal fatigue failure takes place via the initiation and propagation of cracks. The stress that typically causes fatigue failure is usually below the overload failure levels. Reliability of the Pb-free solder joints of the high-density package assemblies is dependent on temperature cycling test conditions, PCB and package materials, as well as the metalization surfaces of the components and of PCBs, such as HASL, Ni/Au, and OSP finishes.

10.4.1. Reliability and Failure Distribution of Pb-Free Solder Joints

Reliability of the solder joint of a particular package is defined as the probability that the solder joint will perform its intended function for a specified period of time, under a given operating condition, without failure. Numerically, reliability is the percentage of survivors, i.e., [51]:

$$R(t) = 1 - F(t), \tag{10.6}$$

where $R(t)$ is the reliability (survival) function and $F(t)$ is the cumulative distribution function (CDF) for failure.

Life distribution is a theoretical population model used to describe the lifetime of a solder joint and is defined as the CDF for the population. Thus, the objective of a reliability test is to obtain failure data and to best fit the failure data to the CDF of a chosen probability distribution, in the most cases, the Weibull distribution. The number of items (sample size) to be tested should be such that the final data are statistically significant.

Most reliability tests are acceleration tests in nature (with increased stress intensity, and realistic sample sizes and test times). Thus, acceleration models (to determine the acceleration factors) are needed to transfer the failure probability, reliability function, failure

rate, and mean time to failure from a test condition to a normal operating condition. In establishing the acceleration models of the solder joints, their surrounding materials (e.g., solder, molding plastic, ceramic, copper, fiber reinforced glass epoxy, and silicon), loading conditions (e.g., stress, strain, temperature, voltage, humidity, current density, and voltage), and failure mechanisms and modes (e.g., overload, fatigue, corrosion, and electromigration) must be considered. The acceleration factors are determined by the predominant material failure physics within the environmental stress range under the test and/or the operation conditions. It is possible that the predominant failure mechanism under the acceleration test condition may not be the one that is operational under the real application conditions. In these cases, the reliability models established in the acceleration tests may not be applicable in practice.

Some common tests for solder joint reliability are temperature cycling, power cycling, functional cycling, shock and vibration, mechanical shear, pull, push, bend and twist, humility, corrosion, voltage, current density, etc. In the microelectronics industry, the solder joint failure caused by temperature cycling under normal application conditions and during testing is the predominant one. Reliability tests based on actual electronic assemblies yield valuable data on the failure distribution only for the specific test condition. The reliability results also depend on the component and PCB substrate materials, and the size of the components. The results are functions of the reliability test environments. The factors that affect reliability of Pb-free solder joint interconnections in electronic assemblies are discussed in the next section.

10.4.2. Effects of Loading and Thermal Conditions on Reliability of Solder Interconnection

Thermal cycling in normal operations or stress-screening tests of electronic products generates very complicated creep/fatigue deformation and damage processes in solder joints. The current Pb-free solder candidates, such as Sn3.0Ag0.5Cu (SAC305), Sn3.9Ag0.6Cu (SAC396), and Sn3.8Ag0.7Cu (SAC387), are based on near eutectic composition of the Sn-Ag-Cu ternary system; they all have higher melting points, different mechanical properties and physical properties (surface tension and viscosity) from the current Sn-Pb eutectic based solders [52]. These Pb-free alloys tend to have a higher yield strength and a higher resistance to creep deformation [53]. From the point of view of traditional engineering design, these attributes are very much desirable since normal engineering components and structures operate well within the material elastic limits, and not much time-dependent deformation (creep or stress relaxation) is anticipated in service. However, solders and solder joints also function as a strain relief buffer for thermal mismatch between PCB substrates and components, in addition to mechanically supporting components and conducting electricity (and heat). A stronger or less ductile solder may not fit all application conditions, particularly when a large thermal strain is expected. For Pb-free and other solders, the homologous temperatures (T/T_{m}) are more than 0.6, leading to the ultimate solder joint lives to be dependent on a variety of parameters, such as initial microstructures, microstructural change during service, temperatures, frequency, holding time, solder joint geometry, as well as restraining effects from adjacent materials and structures [54].

In order to understand the actual reliability of Pb-free solder joints, time and path-dependent creep models are needed to determine the severity of the thermal mechanical loads in solder joints under different service conditions [55]. Methods to reduce such thermal loading in terms of both material properties and solder joint design for any specific ap-

plication thus need to incorporate solder alloy creep and fatigue life models, package/PCB design parameters, and anticipated service conditions.

It is important to distinguish the intrinsic and extrinsic factors that influence the reliability of a solder joint in service and in testing. For a given solder alloy, the reliability of solder joints made with a normal soldering manufacturing process is largely dependent on loading and thermal conditions that are determined by the joint geometric shape and adjacent materials' mechanical, thermal and geometric properties, as well as service or testing conditions. These factors can be considered to be extrinsic in nature. On the other hand, creep, monotonic and cyclic deformation, and resistance to fatigue and creep crack initiation and propagation, are intrinsic to a solder alloy, which are determined basically by the composition-processing-property relationship of the solder alloy.

Since solder joints in general have a complicated geometric shape and geometric discontinuity, and are generally under multi-axial stress–strain conditions, measurement of its mechanical behavior under these conditions are not possible in most cases. Fundamental material mechanical tests of solder alloys are usually conducted on laboratory samples under much simpler stress–strain conditions (uniaxial tension or torsion). These results may or may not be used directly for actual solder joint stress/strain and failure analyses due to a lack of representativeness of the actual microstructure and the actual multi-axial stress condition. Likewise, most industry reliability tests on actual solder joints with certain package design and configuration under specific test conditions may not be extended to other conditions since the solder joints are under uncontrollable and non-measurable non-uniform stress–strain condition, where both extrinsic and intrinsic factors are playing important roles in determining the final reliability lifetime of the solder joints.

Although much has been known to the commonly used Sn-Pb eutectic alloy, the reliability and life assessment methods are still far from perfection due to complicated microstructural evolution and intermetallic reactions, as well as time and path dependent creep and fatigue damage processes. For Pb-free solders, mechanical data from well controlled tests are still scarce. Comparisons between the solder joint reliability performance of different solder alloys were mostly conducted on individual components in specific geometric and thermal conditions [56–58]. Due to many other factors, such as different solder joint geometry, different thermal and mechanical properties of components and board materials, such a comparison may not be generalized, and may sometimes be misleading.

10.4.2.1. Strength of Pb-Free Alloys under Monotonic Shearing and Creep Conditions Many factors can affect the service reliability of a solder joint [59]. Among them, the deformation and fracture behavior of solder alloys under a variety of complex load conditions is foundation to understand and model the reliability performance of real solder joint interconnections. Most of mechanical study on solder alloys was conducted with real solder joints or simplified shearing tests such as sandwiched shearing samples. Due to variations in solder joint shape and small sample size, direct measurement and control of stress–strain are often not possible. Recently, mechanical shearing and creep strength of Pb-free alloy (SAC387) has been reported [60,61] based on thin-walled samples which provide a nearly uniform stress–strain condition for measurement and control, with a computer-controlled bi-axial tension-torsional servo-hydraulic system.

10.4.2.1.1. Shearing Strength of Pb-Free SAC387. Figure 10.36 shows stress–strain curves for SAC387 alloy with shearing strain rates of 6.7×10^{-7} to 1.3×10^{-1} (1/sec). It is clear that strain rates have very significant effects on general deformation behavior of SAC387. With strain rate change from 6.7×10^{-7} to 1.3×10^{-1} range, the change in both yielding stress and flow stress are running as high as high as 400%. What is clear here is

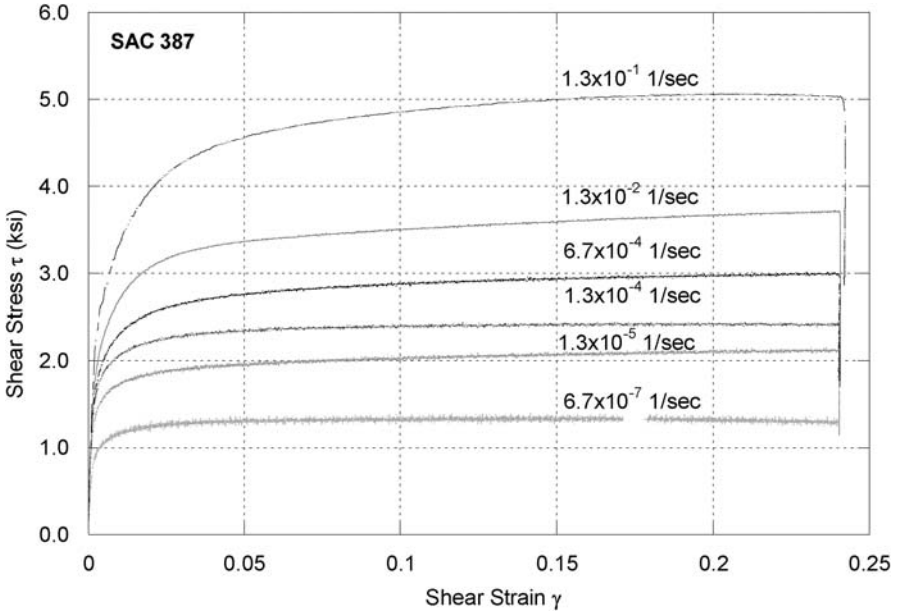


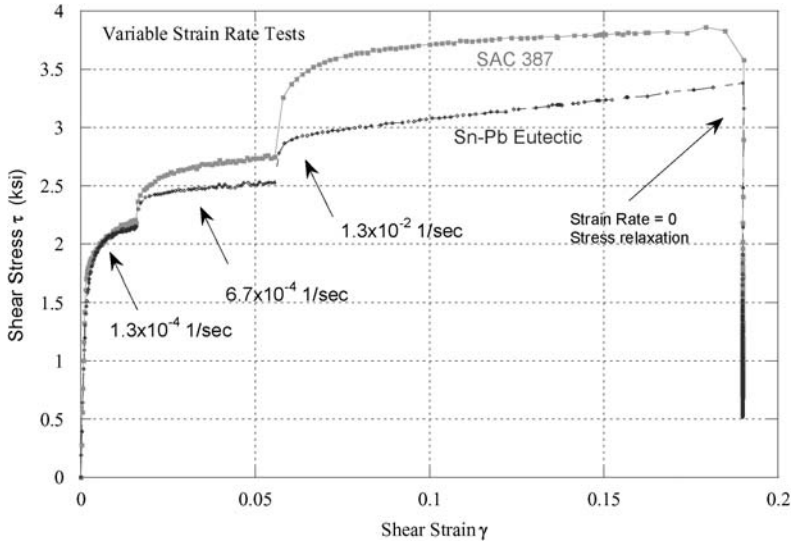
FIGURE 10.36. Strain rate effects on Pb-free SAC387 solder.

that all mechanical properties of elasticity, yielding, flow stress, or ultimate strength of Pb-free SAC387 alloy are highly strain-rate-sensitive. It is ill conceived to compare strengths of Pb-free alloys with others without indicating an exact test condition. The difference between yielding strengths and ultimate strengths, as usual, indicates how strong a material goes through strain hardening. It is clear that a much strong strain hardening occurs at higher strain rates for Pb-free SAC387 alloy.

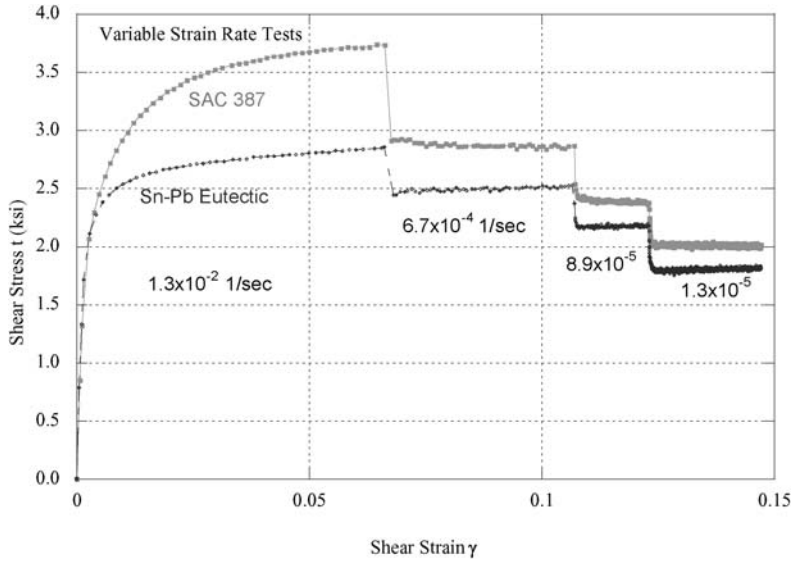
10.4.2.1.2. Variable Strain Rate Shear Test. Variable strain rate and load drop tests on one single test sample have been a very useful mechanical test technique to develop general state variable material constitutive equations [62]. This kind of tests usually requires advanced programming capability and accurate measurement for the mechanical testing system to catch the instantaneous response of materials to sudden change in stress or strain. Figure 10.37 shows such variable strain rate test results for Pb-free SAC387 and Sn-Pb eutectic alloys. In the case of increasing strain rates [Figure 10.37(a)] from 1.3×10^{-4} to 6.7×10^{-4} to 1.3×10^{-2}, it can be seen that there is a rapid increase in inelastic flow stress after each strain rate increase, followed by gradual positive strain hardening with further increase in shear strain. The magnitude of increase in the stress at the beginning of the strain rate change (e.g., from 6.7×10^{-4} to 1.3×10^{-2}) for SAC387 is much more significant than the Sn-Pb eutectic alloy. It is clear that strengthening effects or flow stress increase is not coming only from inelastic deformation; the strain rates play a major role also.

The shear stress–strain curves for the decreasing strain rate tests (from 1.3×10^{-2} to 6.7×10^{-4} to 8.9×10^{-5} to 1.3×10^{-5}) are shown in Figure 10.37(b). Again, it can be seen that the sudden strain rate decreases cause a sharp stress drop, particularly for SAC387 alloy from strain rate of 1.3×10^{-2} to 6.7×10^{-4}. The drop in stress for both alloys takes place rapidly after a sudden decrease in strain rates.

It needs to point out that even the strain rate is reduced, there is no reverse in shear strain direction, and the inelastic strain is still increasing. It is interesting to note that after



(a)



(b)

FIGURE 10.37. Variable strain rate shearing tests. (a) Increasing rates from 1.3×10^{-4} to 1.3×10^{-2}/sec. (b) Decreasing rates from 1.3×10^{-2} to 1.3×10^{-5}/sec.

the initial stress drop, SAC387 flow stress keeps dropping gradually for more than 5% more inelastic deformation at 6.7×10^{-4} after drop from 1.3×10^{-2}, while Sn-Pb eutectic shows a slight strain hardening after the initial stress drop. The next two steps of strain rate drop induce further flow stress decrease for both alloys. It is worth to know that the flow stress levels for both alloys at each strain rate are not exactly same for the two different variable tests, but generally they are within the same order of magnitude. What is interesting is that there is no normal strain hardening after the initial strain rate drop at 1.3×10^{-2} for both

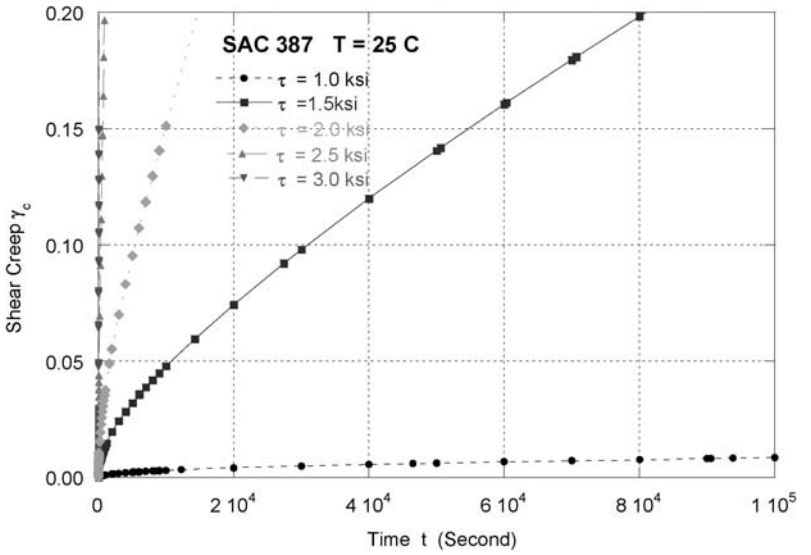


FIGURE 10.38. Shear creep curves of SAC387 solder at RT.

alloys even an additional more than 10% inelastic deformation is accumulated at the low strain rates.

It is clear that the complex stress–strain behavior for both Sn-Pb eutectic and Pb-free SAC387 alloys indicates that the classic elasto-plasticity and strain hardening theories (either isotropic or kinetic hardening) fail to model the mechanical deformation processes in solder alloys. Even if the elastic strain, inelastic strain, and strain rate are known, there is no analytic mathematic equitation to define a unique stress–strain condition. Apparently, the stress–strain is highly path-dependent. Such a behavior will need more sophisticated internal state variable approach to model, where back stress and recovery kinetic can be formulated and integrated with a given deformation history.

10.4.2.1.3. Creep. Creep deformation develops with time under a constant stress. After the initial loading, which may be purely within linear elasticity or elasto-plasticity, all further deformation is accumulative and is time-dependent. At high temperature (usually $T/T_m > 0.5$), metals undergo creep deformation. The primary and secondary stages of a creep curve are determined by a combined action of strain hardening and thermally activated recovery of dislocation obstacles [63,64]. In the primary stage, the strain hardening by formation of dislocation tangles or a dislocation substructure predominates, while the secondary stage is characterized by a balance between strain hardening and recovery softening. The acceleration of creep in the tertiary stage is often caused by formation and joining of micro-voids or cavities on grain boundaries, that is, onset of internal or external damage processes which result in a decrease in the resistance to load or a significant decrease in the effective section area to carry load [65]. The accumulative strain at each stage of creep is very much dependent on stress level, temperature, and stress condition (tensile, torsional or multi-axial stress conditions).

Figure 10.38 and Figure 10.39 show the creep curves for Pb-free SAC387 alloy and Sn-Pb eutectic alloy at different stress levels at room temperature under torsional test condition. In additions to the obvious effects of stress level on creep and creep rates, there is subtle difference in general creep behavior between Sn-Pb eutectic alloy and Pb-free

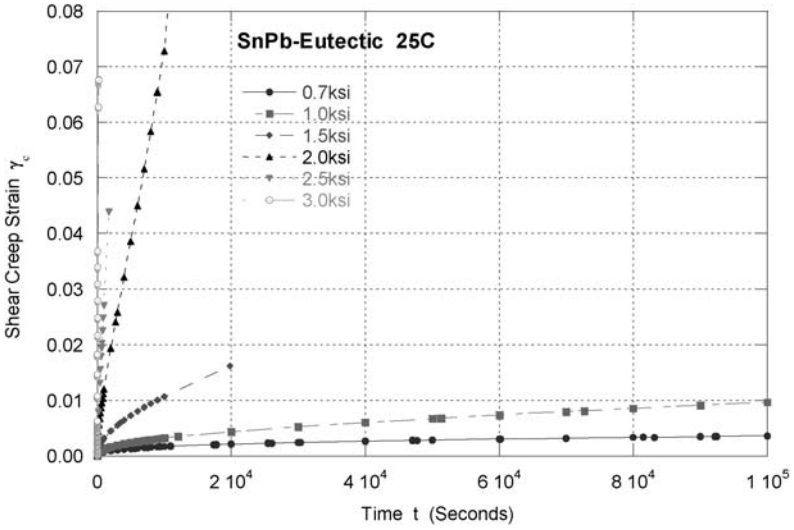


FIGURE 10.39. Shear creep curves of Sn-Pb solder at room temperature.

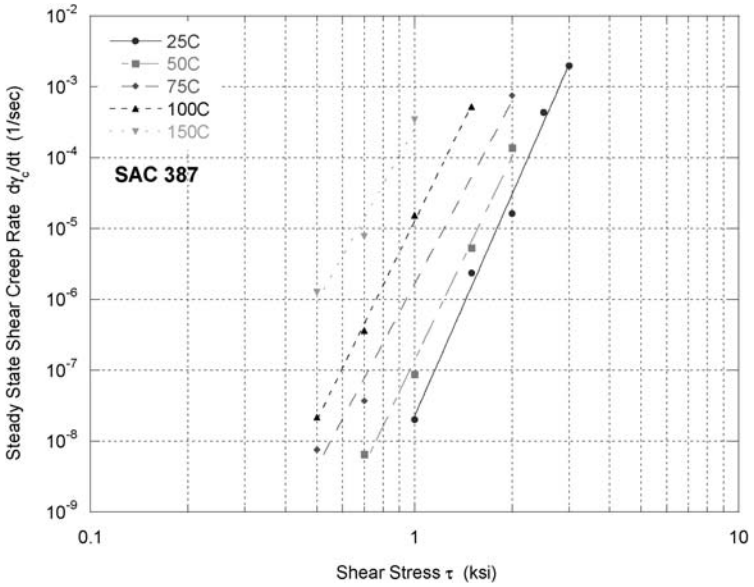


FIGURE 10.40. Shear creep rates vs stress for SAC387.

SAC387 alloy. For Sn-Pb eutectic alloy, the accumulative primary creep is relatively small (less than 1.0% for all the stress levels tested: 0.7 ksi to 3.0 ksi), and also relatively short. For Pb-free SAC387 alloy, the primary creep is a strong function of stress level. At low stress level (1.0 ksi and below), the accumulative primary creep is very low (less than 1.0%). But at high stress level, the primary creep strain can easily reach more than 5.0%, see Figure 10.38. Such primary creep behavior was also observed with other high tem-

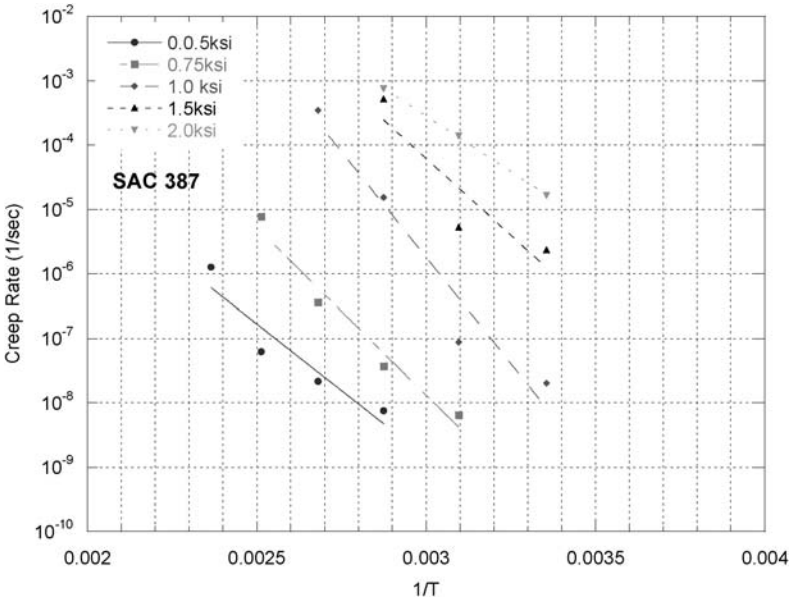


FIGURE 10.41. Temperature effects on creep rates for SAC387.

perature Pb-free and Pb-bearing solder alloys by the authors [66], which could further complicate the general mathematic modeling of creep deformation and fatigue life.

The stress dependency of the secondary (steady state or minimum) creep rates usually indicate the underlining creep mechanisms. For the SAC alloy tested, the secondary creep rates vs shear stress at temperatures ranging from 25°C to 150°C are presented in Figure 10.40. It is clear that SAC alloy shows a high stress dependency of creep rates, and follow the power creep well with the stress exponents approach 10, indicating that there could be a precipitation strengthening mechanism existing.

The temperature effects on stead-state creep rates under different stress levels for SAC387 alloy are presented in Figure 10.41. For Pb-free SAC387 alloy, the activation energy data are is scattering, but indicating that there is single creep activation energy.

10.4.3. Reliability of Pb-Free Solder Joints in Comparison to Sn-Pb Eutectic Solder Joints

Reliability issues can manifest themselves in different forms of failure modes under different operating or testing conditions. The failure could be mechanical, such as fatigue failures: crack initiation and propagation; overstress failures, or it can be electrochemical, such as corrosion, electro-migration and dendrite growth [67]. Only mechanical reliability due to thermal CTE mismatch in service or in testing conditions is discussed in the following sections.

The thermo-mechanical reliability of Pb-free solder interconnects due to CTE mismatch in an electronic system (components, solder joints, and substrates) is determined by creep and fatigue interaction of the solder alloy [68]. Under typical conditions, Pb-free alloys are more creep resistant than the Sn-Pb alloy due to differences in microstructure (such as fine Ag_3Sn phases in the matrix and on the grain boundaries) [69]. In most applications, the Pb-free solder interconnects have been found to show better reliability than the

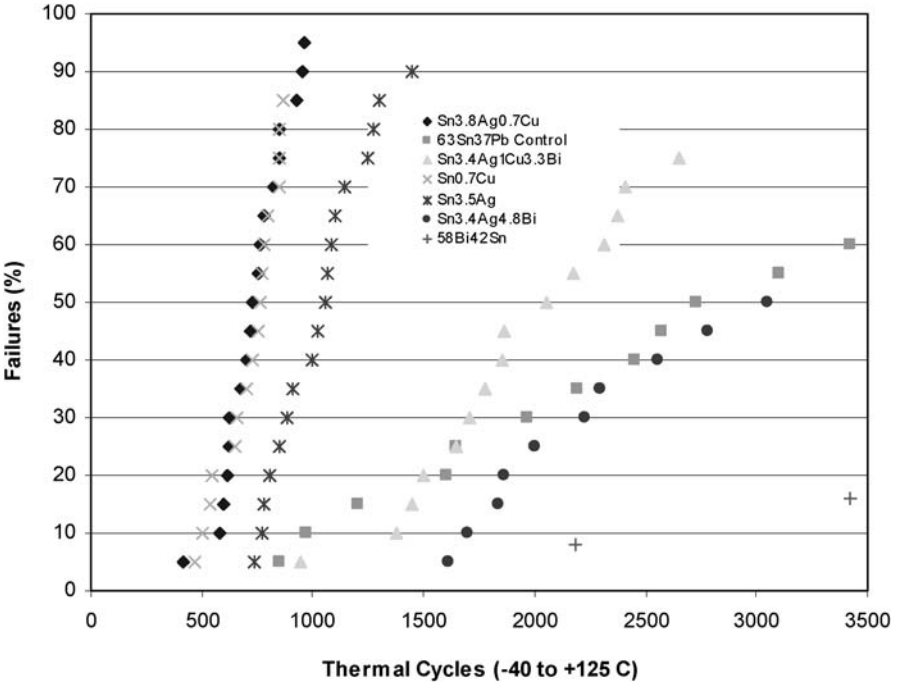


FIGURE 10.42. Thermal cycle reliability distribution curves for LCCC20 (−40°C to 120°C) [83].

Sn-Pb solder joints. However, this may not always be the case. The NCMS study has found that the ranking of Pb-free alloys in terms of reliability relative to Sn-Pb varies with the thermal cycle conditions and component types [70]. In most typical application conditions, the SAC solders are more reliable than the Sn-Pb solder. However, for very large components and components with a very large CTE mismatch with substrates, such as ceramic body components on FR4, and/or under very severe thermal cycling conditions, the Pb-free solder joints are indeed less reliable compared to the Sn-Pb solder joints, as evidenced in Figure 10.42, Figure 10.43 and Figure 10.44 [71–74]. Tests also show that for some leadless chip carriers and large size leadless chip resistors on FR4, and for Alloy 42 leaded components (such as TSOP 48), the fatigue life of Pb-free solder joints are significantly worse than those of the current Sn-Pb eutectic counterparts [69]. In the same NCMS study, it is found that for solder fatigue failure at 0°C to 100°C (10°C/min, 5-min dwell), the Pb-free Sn3.5Ag is similar to Sn-Pb eutectic, but at −55°C to 125°C (11°C/min, 20-min dwell), Sn-Pb eutectic shows much better fatigue life than Sn3.5Ag [52].

It is clear that a simple statement that Pb-free solders will generate more reliable joints could be misleading in certain design and application conditions, since these data above clearly indicate that the fatigue performance ranking of Pb-free alloys relative to Sn-Pb solder varies with thermal cycling conditions and component package types.

To understand the seemly confusing and contradictory reliability performance of Pb-free solder joints, one will need to examine carefully all extrinsic and intrinsic factors that most likely will determine the final fatigue life of a solder joint. Since Pb-free solders have quite different physical and mechanical properties from the current Sn-Pb eutectic solder, it is believed that the solder joint physical shape (geometry), stress and strain conditions, and materials resistance to deformation and fatigue fracture are also quite different for Pb-free

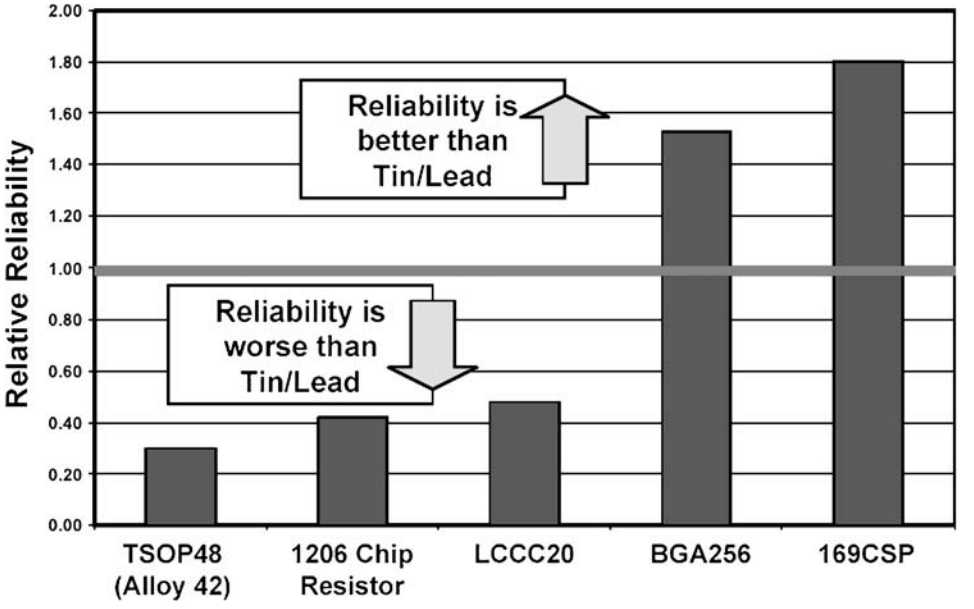


FIGURE 10.43. Comparison of reliability of Pb-free and Sn-Pb eutectic solder joints for different types of component packages [83].

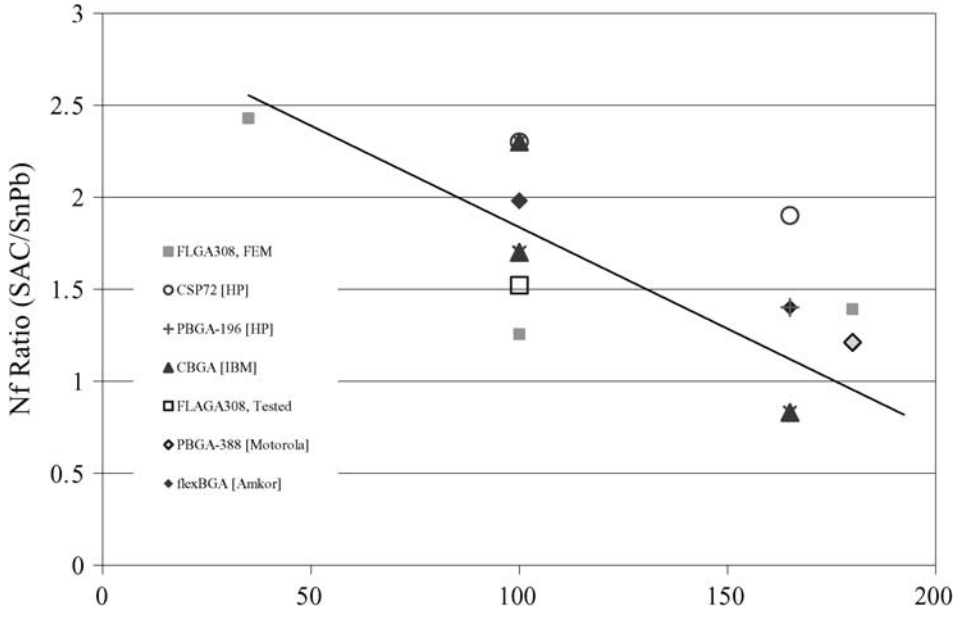


FIGURE 10.44. Fatigue life vs thermal cycling temperatures for a few different BGA packages [84,58,74].

and Sn-Pb eutectic solder joints, even with the same package type under a similar testing condition.

10.4.3.1. Factors That Affect Solder Joint Fatigue Life For any packaging technology, there are a few major extrinsic factors that affect thermal mechanical fatigue reliability of solder joints [75–77]. They include: board and package materials' mechanical and thermal properties, size and geometric dimensions; cyclic temperature profiles of the board and component, including the ramp rate, extreme temperatures, dwell times, etc; solder joint geometry (size, shape, height, etc.); environmental conditions: oxygen, nitrogen, or other reactive species. The intrinsic factors include solder alloy's chemical composition, microstructures, resistance to creep, monotonic and cyclic deformation, crack initiation and propagation, and material degradation sensitivity to complicated multi-axial stress–strain and environmental interactions. Failure of solder joints is usually a result of combined damages from fatigue, creep and environmental interaction. The above factors act together to determine what kind of stress and strain a solder joint will experience at a given time and temperature, and ultimately determine the life of the solder joints.

10.4.3.2. Loading Conditions of BGA Solder Balls The increased use of area-array technology in electronic packaging in recent years has given greater importance to the tasks of predicting and improving the thermal fatigue life of area array solder joints. These tasks typically involve the development of a "macro-model" of the entire assembly (component, substrate, and area array) to identify the critical joint and its end displacements; and a "micro-model" of the critical joint with the prescribed end displacements. The objective of the microanalysis is to determine the distribution of the stress, strain, and/or strain energy (or strain work) density in the critical joint. The output of the microanalysis may then be used as input to an appropriate fatigue life model.

In most cases, the maximum thermal mismatch at the critical joint can be expressed as:

$$\Delta u_{\max} = \beta_{sh} \Delta u_0, \tag{10.7}$$

where β_{sh} is the "shear correction factor" which will account for the combined effects of restraining factors that prevent the free thermal expansion, Δu_0, which is estimated to be:

$$\Delta u_0 = (\alpha_c \Delta T_c(t) - \alpha_s \Delta T_s(t)) \frac{\sqrt{L^2 + W^2}}{2}. \tag{10.8}$$

There also exists an out-of-plane displacement which will cause the board warpage and a tensile (or compression) stress on solder joints. In general, the critical joint (usually at the corner of a BGA array) will experience both shear Δu and z-axial Δw deformation [75,76]. The maximum z-directional (axial) displacement $\Delta w_{\max}$ can be related to the theoretical free thermal expansion Δu_0 as:

$$\Delta w_{\max} = \beta_z \Delta u_0, \tag{10.9}$$

where β_z is the z-axial correction factor, which is also dependent on all of the other restraining factors from PCB substrates and components, as well as solder joint shape and elastic or plastic characteristics, such as Young's modulus and Poisson ration, yielding strength and stress–strain hardening.

Exact analytic solutions for β_{sh} and β_z have been worked out and published for a few special cases of practical interest by the authors [75,76]. These correction factors are functions of the number of solder balls, solder joint geometry, pad size, Young's modulus and Poisson's ratio of solders. Figures 10.45 and 10.46 show the correction beta factors

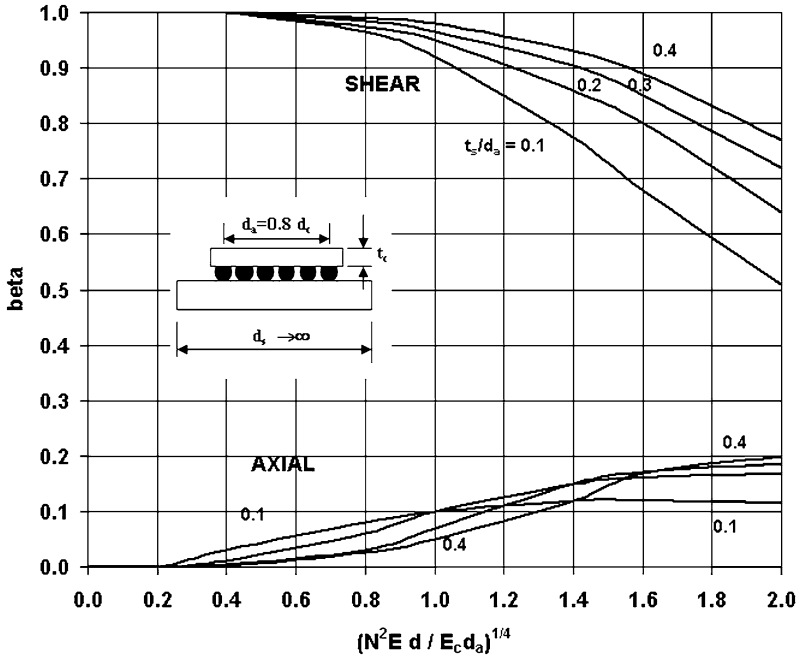


FIGURE 10.45. Shear and axial correction factors for a rigid-component assembly. Notice a positive (tensile) axial displacement [76].

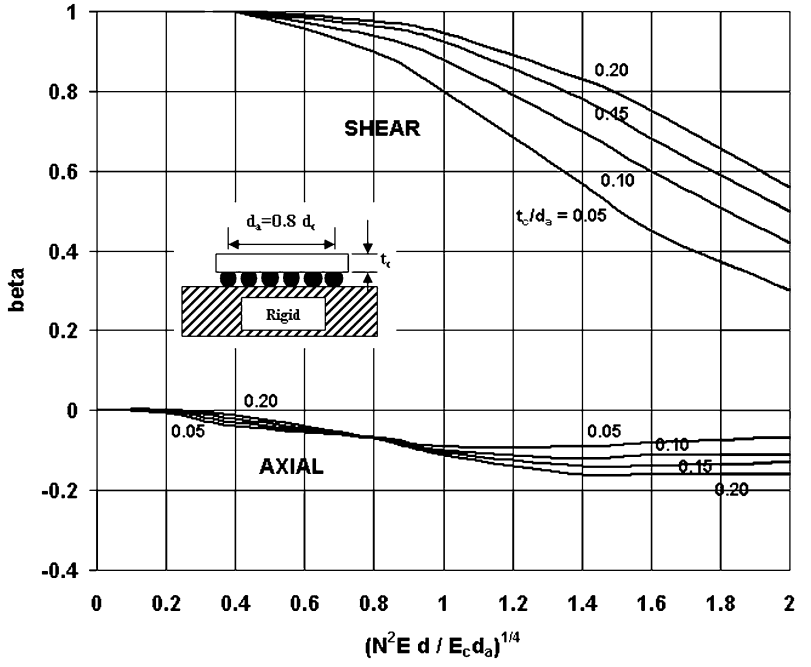


FIGURE 10.46. Shear and axial correction factors for a rigid substrate assembly. Notice a negative (compressive) axial displacement [76].

for two special cases where either component is rather rigid (Figure 10.45), such as in the case of a typical ceramic BGA or flip-chip package, or the PCB board substrate is rigid (Figure 10.46). It is clear that the analytical results establish a clear relationship between the correction factors and the physical parameters of the whole package and PCB system: (a) the dimensions, elastic and thermal properties of the component and substrate; (b) the mechanical properties of solder alloys (i.e., Young's modulus and Poisson's ratio); (c) the array size and population; and (d) the geometric parameters of the individual joints. For example, as shown in Figures 10.45 and 10.46, an increase in the number of solder balls, pad diameter and elastic modulus will increase the overall stiffness, thus causing a reduction of the shear beta factor, which effectively reduces the shear thermal deformation, and possibly, introduces a z-axial deformation, i.e., warping at the same time [76]. Changes in other geometric and mechanical parameters in substrates and components also have direct effects on the final shearing and tensile deformation on solder balls.

It is clear that the loading conditions on a BGA solder joint are mostly shearing and axial displacements, and are determined by geometric parameters and material mechanical properties of the entire assembly. The thermal cycling temperatures and profiles in service or testing will determine the loading amplitude and spectrum that a solder joint will experience with time. These shearing and axial displacements can be estimated fairly accurately. The result may be used as loading conditions (or boundary conditions) for solder joint life time assessment either based on first principles or based on time-dependent inelastic finite element analysis, should the creep and fatigue life models be available [54].

Shear strain on solder joints in service thermal and/or power cycling is generally the most damaging mechanical load to determine solder fatigue life and reliability. The shear strain $\gamma_{th}(t)$ is a function of the thermal mismatch, solder joint geometry and material elasto-plastic and creep constitutive relationships [54]. Depending on temperature and the magnitude of the total thermal mismatch, the strain at any given time is a combination of elastic, plastic and creep deformation. Furthermore, the solder strain and stress are not uniformly distributed in the solder joints; they are functions of location (x, y, z), temperature, time and pre-stress/strain history. A more accurate estimation of stress and strain on solder joints requires a complete understanding and mathematical modeling of solder alloy's plasticity and creep; and can be carried out only with a non-linear and time-dependent finite element modeling.

For a thermal cycle where ΔT_c and ΔT_s are given, the thermal fatigue life of the critical solder joints can be, based on the widely used modified Coffin-Mason equation, estimated as [77]:

$$N_f = \frac{1}{2}\left(\frac{\Delta\gamma_{in}}{\varepsilon_f}\right)^{\frac{1}{c}}, \tag{10.10}$$

where $\Delta\gamma_{in}$ is the inelastic strain range (for simplicity, it is roughly approximated as the average strain along a BGA solder ball); and c and ε_f are materials constants that are functions of thermal cycling temperatures and frequency [77]. As discussed above, all geometric and material parameters have their own contribution to the shear displacement, these effects are reflected in terms of the resultant shear strain, which then determines the final fatigue life.

Equations (10.7–10.10) and beta correction factors incorporate some of the most important factors of package design parameters, relevant material properties and service (or testing) conditions, into analytical expression for solder joint reliability assessment based

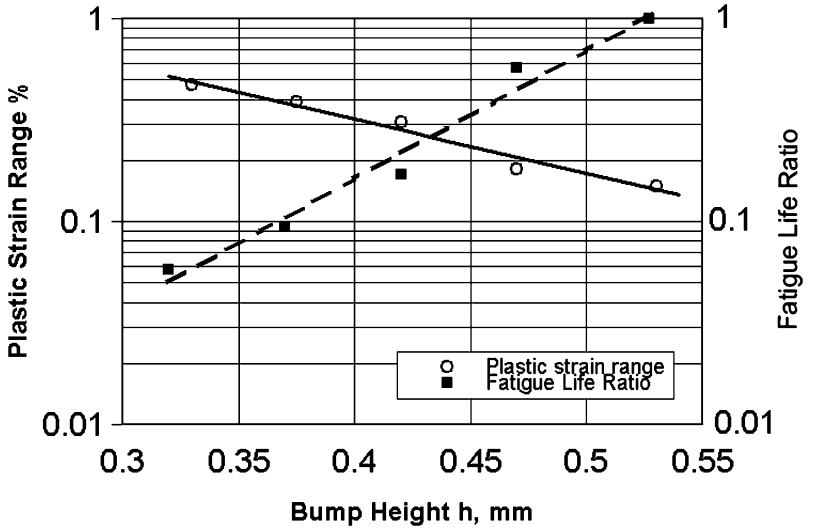


FIGURE 10.47. Solder ball height vs plastics stain and fatigue lives [78].

on simplified first principles. It is clear that the BGA balls' fatigue life is very sensitive to any slight change in solder joint height, solder volume, and CTEs for a given thermal cycle condition.

Figure 10.47 schematically shows solder joint bump height effect on estimated plastic strain range and thermal fatigue life of solder joints based on Sn-Pb eutectic solder [78]. It can be seen that when a solder joint's height is reduced from about 0.54 mm to 0.32 mm, the fatigue life can drop by more than 90%. It is apparent that the solder ball height is one of the most important and sensitive parameter to BGA reliability. In general, the solder ball cross-section along the height vary; the ball will assume different geometrical shape determined by the total solder volume, molten solder surface tension, component size and weight, as well as the soldering surface finishes. Stress and strain will concentrate around the smallest cross section area, leading to early fatigue crack initiation, and to a shorter fatigue life. Figure 10.48 shows the effects of solder ball geometry shape, height, and chip size on the fatigue life of Sn-Pb eutectic solder balls [79].

Since most published data on Pb-free solder reliability did not clearly state in what conditions the comparison to Sn-Pb solder is made (except for the thermal cycling profile), the difference in fatigue reliability performance is also likely due to the solder ball geometry and package materials because of different soldering temperatures and solder alloys' surface tension energy. Because Pb-free solders are generally stronger, the stiffening restraining effects on the maximum thermal shear displacement at the critical solder joint will be larger than that for the softer Sn-Pb eutectic solder. Likewise, the higher surface tension energy of Pb-free solder alloys could produce taller solder balls. Both factors would reduce the maximum thermal strain on Pb-free solder joints. Therefore, in theory, Pb-free solder joints can be more reliable than Sn-Pb eutectic joints for a similar package design due to these extrinsic factors, assuming both solders have a similar fatigue resistance. Elimination of the difference in these factors with well controlled test setup and samples could shed light on the real fatigue resistance of the alloys. Unfortunately, making identical solder joints out of two completely different metallurgical alloy systems are difficult. Therefore, an exact comparison and conclusive extrapolation about alloy performance based only on

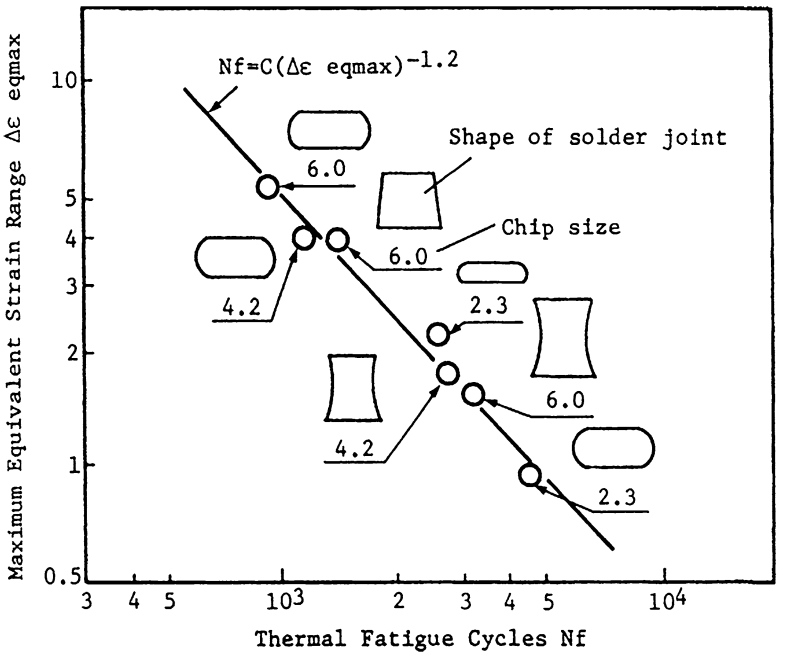


FIGURE 10.48. Effect of BGA ball geometry and chip size on fatigue life of Sn-Pb eutectic solder balls [79].

a few specific test configurations with Pb-free and Sn-Pb eutectic solder joints would not be viable.

10.4.3.3. Reliability of Pb-Free Alloys and Joints: Isothermal or Thermal Mechanical Fatigue? Since most factors indicate a longer fatigue lives of Pb-free solder balls for BGA applications, the discrepancy that in certain conditions Pb-free solder joints show a poor reliability compared to Sn-Pb eutectic points to some fundamental difference in thermal fatigue processes between Pb-free and Sn-Pb eutectic alloys.

It has been reported early by Soloman that temperature has no significant effect on low cycle fatigue (LCF) life of the Sn-Pb eutectic solders with strain control LCF tests up to 150°C [80]. Our previous study [66,81] for the modified eutectic alloy 62.5Sn-36.1Pb-1.4Ag, as shown in Figure 10.49, also proved this point, as well as with published data from M. Mukai [82], see Figure 10.50. However, temperature can have a significant effect on Pb-free alloys, such as 95Sn-5Ag, see Figure 10.51, which shows that an increase in temperature generally decreases the fatigue lives of the Sn5Ag alloy.

It is likely that temperature, strain rate, and stress levels may have significant effects on the thermal fatigue life of Pb-free alloys. Therefore, the simplified strain-range based fatigue life model, such as Equation (10.11), may not work well for Pb-free alloys. Most thermal fatigue life models are based on isothermal fatigue tests, where an isothermal equivalent temperature for a thermal cycle profile can be defined. For Pb-free alloys, however, not much experimental work has been conducted to establish a generally acceptable thermal fatigue life model. From the limited published and unpublished literature, fatigue life models based on stress–strain work density of hysteresis loops during thermal cycle tests, such as shown in Figure 10.52, clearly indicate that based on normalized creep strain energy density for thermal cycling tests, Sn-Pb eutectic solder has a cross-over with

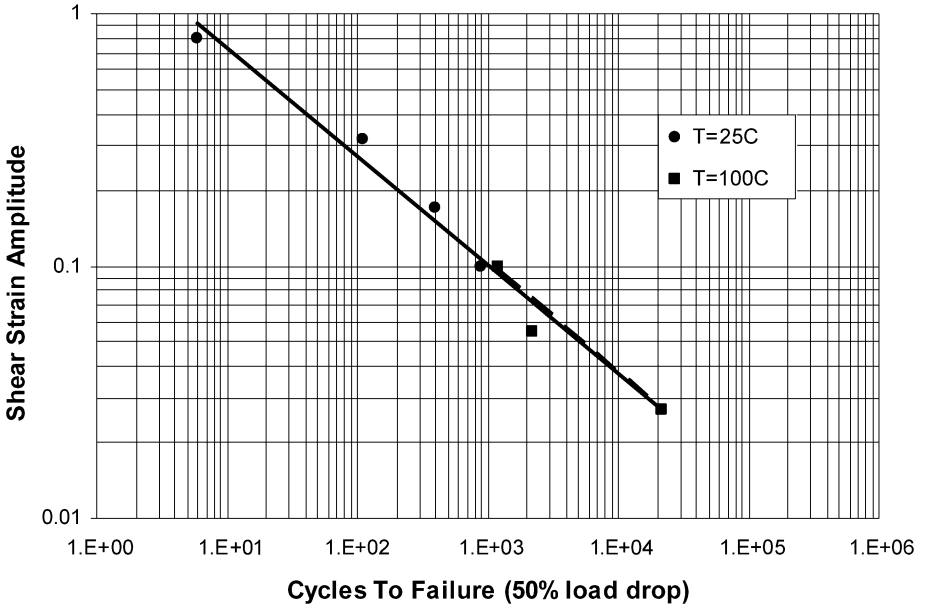


FIGURE 10.49. Temperature effects on isothermal low cycle fatigue life of Sn-Pb eutectic solder alloy [66,81].

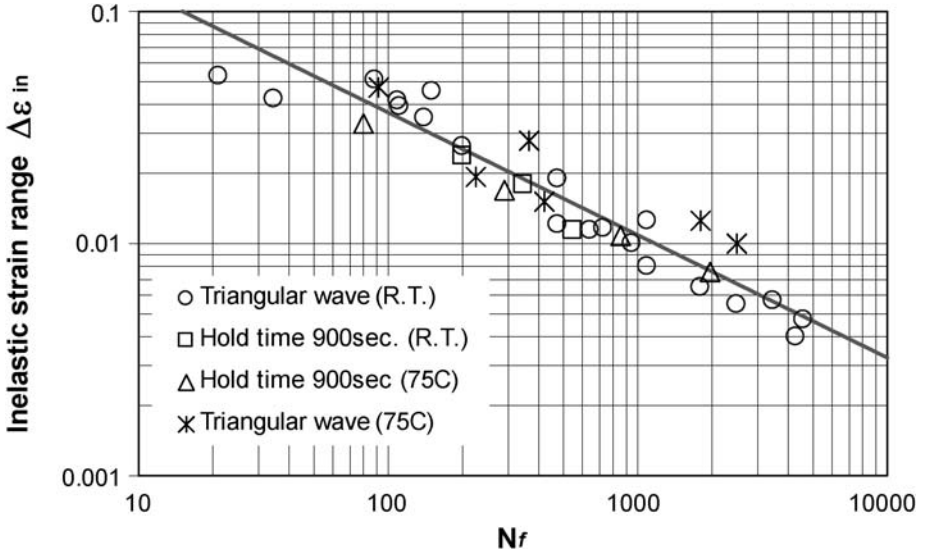


FIGURE 10.50. Fatigue life of eutectic solder vs cyclic inelastic (plastic) strain range at room temperature and 75°C with dwell times [83].

Pb-free alloys [83]. At higher stress–strain levels, the Sn-Pb eutectic solder has a superior fatigue resistance to Pb-free alloys; at lower or modest stress–strain levels, Pb-free alloys would outperform the Sn-Pb eutectic alloy. This is in a good agreement with above mentioned comparisons of BGA solder ball fatigue reliability, as shown in Figure 10.43 and Figure 10.44. Since for a relatively large component with a large thermal mismatch (such

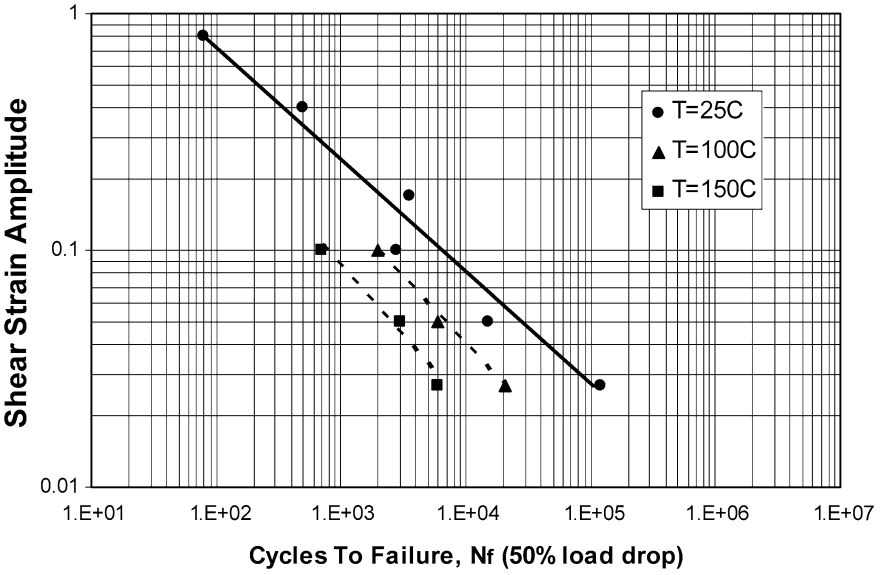


FIGURE 10.51. Temperature effects on isothermal low cycle fatigue life of a Pb-free solder alloy, 95Sn-5Ag [81,82].

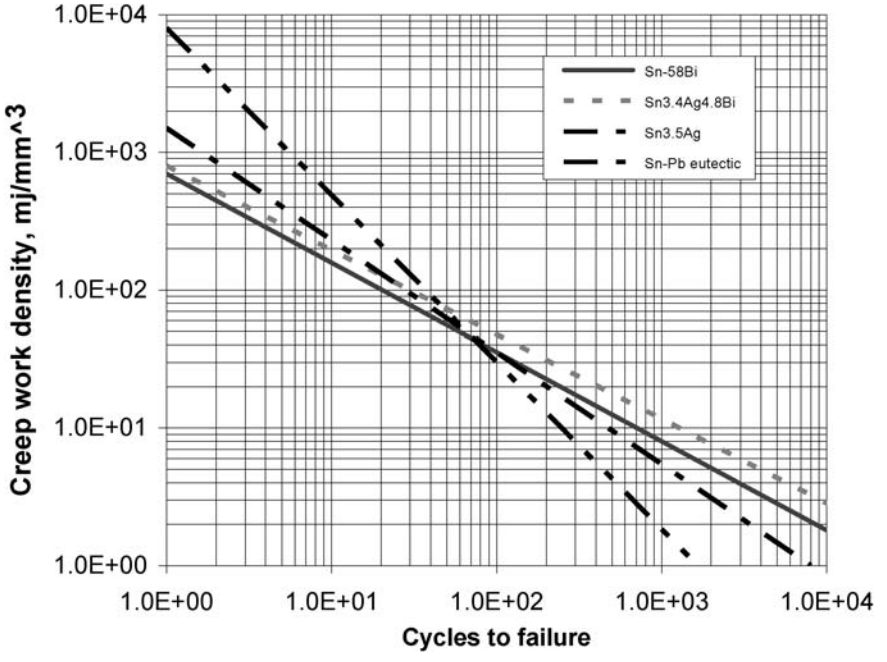


FIGURE 10.52. Thermal fatigue reliability comparisons of Pb-free and Sn-Pb alloys based on stress–strain work density in a thermal cycle [45,84].

as silicon and ceramic vs RF-4 substrate), a large thermal cycling temperature range will increase the stress–strain energy level, thus could result in lower thermal fatigue lives. For

ceramic BGA, the intrinsic large CTE mismatch and a near 1.0 shear beta factor could generate a condition where the thermal strain range for Pb-free alloys can be very high, thus result in a less favorable fatigue performance. Figure 10.44 clearly shows that with large thermal cycling, CBGA solder balls have lower fatigue lives than the Sn-Pb eutectic alloy joints. When the thermal cycle range is reduced, Pb-free solder balls then outperform the Sn-Pb solder joints.

From the above analyses and observation of most "discrepancy" for Pb-free alloys, it can be deduced that Pb-free alloys are much more sensitive to temperature in terms of reduction of creep resistance and isothermal low cycle fatigue performance. Therefore, the thermal fatigue life of a Pb-free solder joint will depend not only on the strain range, but also on the thermal cycle profile and strain rates, as well as the stress level. Thus, the energy-based fatigue life models, such as shown in Figure 10.52, would do a better job to characterize the fatigue reliability of Pb-free alloys and their joints.

This analysis shows that for more realistic service conditions encountered with most of the SMT leaded joints (such as QFP), when the lead stiffness is very low, the solder joint is most likely stressed well below its yielding point, and quite possibly, below the stress level where no significant time-dependent creep deformation may take place. Thus, solder alloys with a high creep resistance, such as SAC Pb-free alloys, will significantly improve SMT leaded joint reliability, as long as the lead stiffness and service temperature are not extremely high. On the contrary, for solders used in leadless interconnection or BGAs with large thermal mismatch and temperature swings, a low creep resistance and high ductility (such as the current soft Sn-Pb eutectic alloys) may prolong the solder joint fatigue life [54].

Since Pb-free solders are generally stronger and more creep resistant at the normal operational temperature range, a complaint lead design will lead to a better reliability of these leaded solder joints. However, if the leads are very stiff in geometry and/or of high modulus material, such as Alloy 42, the thermal mismatch will be transferred to the solder joints either with an instantaneous plastic deformation or through rather high rate creep. If the level of thermal strain is high due to CTE mismatch or a large thermal cycling temperature range, these leaded Pb-free solder joints could also show less favorable thermal fatigue life, as evidenced by TSOP 48-alloys 42 in Figure 10.43. It is clear that to fully utilize the good attributes of high yielding strength and high creep resistance of Pb-free alloys, stiff leads should be avoided.

10.4.3.4. Thermal Mechanical Fatigue Life Assessment In general, cyclic strain-range based fatigue life models, such as Equation (10.10), work well for isothermal low cycle fatigue of most engineering materials where material microstructure (except sub-cell dislocation structure) is relatively stable during the fatigue damage process. However, in most thermo-mechanical fatigue conditions, a simple strain range based fatigue life model may not work since it fails to take into consideration of the effects of peak stress and strain rate. Other fatigue life models, based on strain energy density, or models based on fatigue crack growth rate with help of non-linear fracture mechanics, may be needed for more accurate fatigue life assessment under thermo-mechanical conditions [84,85].

From early thermo-mechanical study of superalloys for aerospace applications, it has been learned that a stress–strain energy density fatigue life model could be a better fatigue failure criterion since it incorporates some effects of strain rates and temperatures. In general, a stress–strain energy-based fatigue life model can be expressed as:

$$N_f = c(W_{ss})^m, \tag{10.11}$$

where c and m are temperature-dependent material constants, which can be derived from isothermal low cycle fatigue tests at different temperatures; W_{ss} is the stress–strain hysteresis energy density. However, in practice, an accurate W_{ss} can not be estimated without establishing a complete temperature–stress–strain material constitutive relationship. The constitutive equation is necessary to model the effects of loading conditions and strain rates (or heating-up/cooling-down rates, dwell, etc.), which usually affect the peak stresses for a thermal cycling profile. In addition, inelastic strain (both plasticity and creep) can also be partitioned if necessary to use more advanced fatigue life models, such as strain-range partition or damage mechanics.

For Pb-free solders, there are not enough data to establish widely acceptable fatigue life models for the alloys themselves at this time. The models should be applicable for any testing or service condition, particularly in thermo-mechanical loading conditions encountered by solder joints. Well-controlled fatigue tests on Pb-free alloys with precise control of either stress and or strain are mostly conducted under isothermal conditions. To extend these fatigue life models to thermal cycle conditions where temperature and strain/stress changes take place simultaneously, will need much more investigation on Pb-free alloys. For solder joints, such as solder balls for BGA and leaded solder joints, which are mostly under multi-axial loading conditions with much more complicated non-uniform deformation and damage processes, thermal fatigue life assessment would require nonlinear and time dependent finite element analysis of stress/strain distribution. Doing so will also need much more study on Pb-free solder alloy's fundamental creep mechanisms and mathematical modeling.

10.5. GUIDELINES FOR PB-FREE SOLDERING AND IMPROVEMENT IN RELIABILITY

In summary, the mechanical loading, metallurgical structures and design parameters, as well as service (or test) conditions, all affect the reliability of Pb-free and Sn-Pb eutectic solder joints. Solder joint thermal deformation is determined not only by external environments, but also by the solder alloy itself and joint geometry. Comparison of solder joint reliability between Pb-free and Sn-Pb eutectic alloys does not always favor the high strength Pb-free alloys in all conditions. For isothermal low cycle fatigue with modest strain ranges, Pb-free alloys outperform Sn-Pb eutectic alloys. However under thermal mechanical conditions, Pb-free alloys show significant deterioration of reliability when large thermal strain and/or stress are encountered.

Pb-free solder alloys with a high creep resistance will significantly improve the reliability of BGA solder balls and compliant leaded solder joint as long as thermal mismatch and thermal cycling do not generate large inelastic deformation. On the contrary, for leadless joints, non-compliant leaded joints, and BGA with high CTE mismatch, a low creep resistance and high ductility Pb-free solder similar to the current Sn-Pb eutectic alloy would be more favorable.

REFERENCES

1. G. Humpston and D.M. Jacobson, Principles of Soldering, ASM International, Materials Park, OH, 2004.
2. J.F. Roeder, M.R. Notis, and H.J. Frost, in D.R. Frear, W.B. Jones, and K.R. Kinsma, Eds., Solder Mechanics, A State of the Art Assessment, TMS, 1991, p. 1.

3. E.E. de Kluizenaar, Soldering and Surf. Mount. Tech., 4, pp. 27–38 (1990).
4. Directive 2002/96/EC of the European Parliament and of the Council of 27 January 2003 on Waste Electrical and Electronic Equipment, Official Journal of the European Union, 13.2.2003, pp. L37/24–L37/38.
5. Directive 2002/95/EC of The European Parliament and of the Council of 27 January 2003 on the Restriction of the use of Certain Hazardous Substances in Electrical and Electronic Equipment, Official Journal of the European Union, 13.2.2003, pp. L37/19–L37/23.
6. Z. Mei and J.W. Morris, Jr., Journal of Electronic Materials (A.I.M.E. Metallurgical Society), 21, pp. 599–607 (1992).
7. N.-C. Lee, Lead-free soldering of chip-scale packages, Chip Scale Review, March/April, p. 42 (2000).
8. C.M.L. Wu, D.Q. Yu, C.M.T. Law, and L. Wang, The properties of Sn-9Zn lead-free solder alloys doped with trace rare earth elements, J. Electronic Mater., 31, pp. 921–927 (2002).
9. K.W. Moon, W.J. Boettinger, U.R. Kattner, F.S. Biancaniello, and C.A. Handwerker, J. Electronic Mater., 29, pp. 1122–1236 (2000).
10. C.-S. Oh, J.-H. Shim, B.-J. Lee, and D.N. Lee, J. Alloys and Compounds, 238, pp. 155–166 (1996).
11. J.H. Shim, C.-S. Oh, B.-J. Lee, and D.N. Lee, Z. Metallkde., 87, pp. 205–212 (1996).
12. K.S. Kim, S.H. Huh, and K. Suganuma, Mater. Sci. Eng. A, 333, p. 106 (2002).
13. J. Park, R. Kabade, C. Kim, T. Carper, S. Dunford, and V. Puligangla, J. Electronic Mater., 32, p. 1774 (2003).
14. B.P. Richards, C.L. Levoguer, C.P. Hunt, K. Nimmo, S. Peters, and P. Cusack, Lead-free soldering, an analysis of the current status of lead-free soldering, Department of Trade and Industry, UK, 2001.
15. J. Liang, N. Dariavach, P. Callahan, and D. Shangguan, Metallurgy and kinetics of liquid-solid interfacial reaction during lead-free soldering, Trans. Mater. (2005) (to be published).
16. F.G. Yost and A.D. Romig, Materials Research Society Symposium Proceedings, 108, pp. 385–390 (1988).
17. H. Wang and H. Conrad, Metall. Mater. Trans., 26A, pp. 495–469 (1995).
18. S.V. Sattiraju, R.W. Johnson, D.Z. Genc, and M.J. Bozack, Wetting performance vs. board finish and flux for several Pb-free solder alloys, PackCon 2000, October 2–3, Santa Clara, CA, 2000.
19. N. Dariavach, P. Callahan, J. Liang, and R. Fournelle, A study of intermetallic growth kinetics for Sn-Ag, Sn-Cu and Sn-Ag-Cu lead-free solders on Cu, Ni and Fe-42Ni substrates, TMS 05, 2005.
20. K. Suganuma and G. Nakamura, J. Japan Inst. Metals., C59, p. 1299 (1995).
21. M. Schaefer, W. Laub, R.A. Fournelle, and J. Liang, in R.K. Mahidhara, et al., Eds., Design and Reliability of Solders and Solder Interconnects, TMS, Warrendale, PA, 1997, p. 247.
22. S. Chada, R.A. Fournelle, W. Laub, and D. Shangguan, J. Electronic Mater., 29, p. 1214 (2000).
23. S. Chada, W. Laub, R.A. Fournelle, and D. Shangguan, J. Electronic Mater., 28, p. 1194 (1999).
24. H.K. Kim and K.N. Tu, Phys. Rev. B, 53(23), p. 16027 (1996).
25. M. Schaefer, R.A. Fournelle, and J. Liang, J. Electronic Mater., 27, p. 1167 (1998).
26. R.A. Gagliano and M.E. Fine, J. Electronic Mater., 32, p. 1441 (2003).
27. ASM Handbook, Vol. 8, Metallography, Structures and Phase Diagrams, 8th Edition, ASM, 1973.
28. J. Oliver, M. Nylen, O. Rod, and C. Markou, Fatigue properties of Sn3.5Ag0.7Cu solder joints and effects of Pb-contamination, Proceedings of SMTA International Conference, Chicago, IL, Sept. 2002.
29. R. Gordon, S. Marr, and D. Shangguan, Evaluation of immersion silver PWB finish for automotive electronics, Proceedings of SMTA International Conference, Chicago, September 2000, pp. 583–591.
30. M. Arra, D. Shangguan, and D. Xie, Wetting of fresh and aged lead-free PCB surface finishes by Sn/Ag/Cu solder, Proceedings of APEX 2003, Anaheim, CA, March 2003, pp. S209–2-1/5.
31. M. Arra, D. Shangguan, J. Sundelin, T. Lepistö, and E. Ristolainen, Aging mechanisms of immersion tin and silver PCB surface finishes, Proceedings of the 3rd IPC/JEDEC Annual Conference on Lead-Free Electronic Assemblies and Components, 2003, pp. 170–179.
32. M. Arra, D. Shangguan, D. Xie, J. Sundelin, T. Lepistö, and E. Ristolainen, Study of immersion silver and tin PCB surface finishes in lead-free solder applications, Proceedings of Soldertec/IPC International Lead-free Conference, Brussels, Belgium, 11–12 June 2003, pp. 423–446.
33. M. Arra, D. Geiger, D. Shangguan, S. Yi, F. Grebenstein, H. Fockenberger, K.H. Kerk, and H. Wong, Performance evaluation of lead-free solder paste, Proceedings of SMTA Conference, Chicago, IL, Oct. 2001, pp. 850–857.
34. M. Arra, D. Shangguan, E. Ristolainen, and T. Lepisto, Solder balling of lead-free solder paste, J. Electronic Mater., 31(11), pp. 1130–1138 (2002).
35. M. Arra, D. Shangguan, E. Ristolainen, and T. Lepistö, Effect of reflow profile on wetting and intermetallic formation between Sn/Ag/Cu solder, components and printed circuit boards, Soldering and Surf. Mount. Tech., 14(2), pp. 18–25 (2002).

36. M. Arra, D. Shangguan, S. Yi, H. Fockenberger, and R. Thalhammer, Development of lead-free wave soldering processes, IEEE CPMT Transactions on Electronics Packaging Manufacturing, 25(4), pp. 289–299 (2002). Also, Proceedings of APEX 2002, San Diego, CA, Jan. 2002, pp. 4(1)–4(10).
37. J. Lau, D. Shangguan, W. Dauksher, D. Khoo, G. Fan, W. Loong-Fee, and M. Sanciaume, Lead-free wave-soldering and reliability of a light-emitting diode (LED) display, Proceedings of Soldertec/IPC International Lead-free Conference, Brussels, Belgium, 11–12 June 2003, pp. 116–124.
38. I. Amato, Tin whiskers: the next Y2K problems? January, 10, FORTUNE, 2005, p. 27.
39. P.J.T.L. Oberndorff, M. Dittes, and L. Petit, Intermetallic formation in relation to tin whiskers, Proceedings of IPC/SolderTec, International Conference on Lead Free Electronics, Brussels, Belgium, June 11–12, p. 171.
40. J. Liang, X. Li, Z. Xu, and D. Shangguan, Nano-indentation study on whisker formation on tin plated component leads, IPC/JEDEC 8th International Conference on Lead-free Electronic Components and Assemblies, San Jose, CA, April 18–20, 2005.
41. EMC/Brown, Stress Evolution and Whisker Formation in Cn-Sn Bimetallic Layers, June 1, 2004.
42. JEDEC, Qualification requirements for tin whisker mitigation practices of component lead finishes, Draft, Nov., 2004.
43. D. Geiger, D. Shangguan, and S. Yi, Thermal study of lead-free reflow soldering processes, Proceedings of the 3rd IPC/JEDEC Annual Conference on Lead-Free Electronic Assemblies and Components, 2003, pp. 95–98.
44. D. Geiger, J. Yu, and D. Shangguan, Development of assembly and rework processes for large and complex PCBs using lead-free solder, Proceedings of APEX 2004.
45. D. Shangguan, Supply chain impact of lead-free soldering, Proceedings of the 10th Annual Pan Pacific Microelectronics Symposium, Kauai, Hawaii, January 2005.
46. D. Shangguan, A holistic approach to lead-free transition and environmental compliance, Proceedings of SMTA 2004, Chicago, IL, 2004.
47. J.H. Lau and D. Rice, Solder joint fatigue in surface mount technology: state of the art, Solid State Technology, 28(October), pp. 91–104 (1985).
48. J.H. Lau, Thermal stress analysis of SMT PQFP packages and interconnections, ASME Transactions, Journal of Electronic Packaging, 111(March), pp. 2–8 (1989).
49. J.H. Lau, G. Harkins, D. Rice, J. Kral, and B. Wells, Experimental and statistical analyses of surface mount technology PLCC solder joint reliability, IEEE Transactions on Reliability, 37(5), pp. 524–530 (1988).
50. R.R. Tummala, Fundamentals of Microsystems Packaging, McGraw-Hill, New York, NY, 2001.
51. J. Lau, N. Hoo, R. Horsley, J. Smetana, D. Shangguan, W. Dauksher, D. Love, I. Menis, and Sullivan, Reliability testing and data analysis of high-density packages' lead-free solder joints, Soldering and Surf. Mount. Tech., 16(2), pp. 46–68 (2004). Also, Proceedings of APEX 2003, Anaheim, CA, March 2003, pp. S42-3-1/24.
52. NCMS Report 0401RE96, Lead-Free Solder Project, Final Report, National Center for Manufacturing Sciences, August 1997.
53. J. Hwang, SMT March 2001, 2001, p. 70.
54. J. Liang, et al., Advances in Electronic Packaging, 1997, AMSE InterPack'97, ASME, 1997, pp. 1583–1592.
55. D.S. Stone and M.M. Rashid, in D.R. Frear, S.N. Burchett, H.S. Morgan, and J.H. Lau, Eds., The Mechanics of Solder Alloy Interconnects, Van Nostrand Reinhold, New York, 1994, pp. 87–157.
56. D. Shangguan, Study of compatibility for lead free solder PCB assembly, Proceedings of International Conference on Lead-Free Electronics, SolderTec, Brussels, June, 2003, pp. 297–308.
57. J. Clech, Proceedings of IPC/SMEMA APEX 2004 Conference, Anaheim, CA, Feb., 21–26, 2004.
58. J. Lau, N. Hoo, R. Horsley, J. Smetana, D. Shangguan, W. Dauksher, D. Love, I. Menis, and B. Sullivan, Reliability testing and data analysis of high-density packages' lead-free solder joints, Proceedings of APEX 2003, Anaheim, CA, March 2003.
59. D.R. Frear, S.N. Burchett, and H.S. Morgan, Introduction: the mechanics of solder alloy interconnection, in D.R. Frear, S.N. Burchett, H.S. Morgan, and J.H. Lau, Eds., The Mechanics of Solder Alloy Interconnects, Van Nostrand Reinhold, New York, 1994, pp. 1–6.
60. J. Liang, N. Dariavach, and D. Shangguan, Deformation behavior of solder alloys under variable strain rate shearing and creep conditions, IEEE International Symposium and Exhibition on Advanced Packaging Materials, Irvine, CA, USA, March 16–18, 2005.
61. J. Liang, N. Dariavach, P. Callahan, G. Barr, D. Shangguan, and C. Li, Deformation and fatigue fracture of solder alloys under complicated load conditions, InterPACK '05, The ASME/Pacific Rim Technical Conference and Exhibition on Integration and Packaging of MEMS, NEMS, and Electronic Systems, San Francisco, CA, Westin St. Francis Hotel, July 17–22, 2005.

62. E. Krempl, The role of servocontrolled testing in the development of theory of viscoplasticity based on total strain and overstress, in R.W. Rohde and J.C. Swearenger, Eds., Mechanical Testing for Deformation Model Development, ASTM STP 765, ASTM, 1982, pp. 5–28.
63. J.H. Gittus, Creep, Viscoelasticity and Creep Fracture in Solids, John Wiley & Son Inc., 1975.
64. E. Orowan, J. West Scotl. Iron Steel Inst., 54, p. 45 (1946).
65. H. Riedel, Fracture at High Temperatures, Springer-Verlag, 1986.
66. J. Liang, N. Gollhardt, P.S. Lee, S.A Schroeder, and W.L. Morris, A study of fatigue and creep behavior of four high temperature solders, Fatigue Fract. Engng Mater. Struct., 19, pp. 1401–1409 (1996).
67. J. Liang, N. Dariavach, and S. Downs, Load conditions and reliability of Pb-free solder alloys and solder joints, IPC/Soldertec Global 2nd International Conference on Lead Free Electronics, Towards Implementation of the RoHS Directive, Amsterdam, June 21–23, 2004.
68. W. Ren, M. Lu, S. Liu, and D. Shangguan, Thermal mechanical property testing of lead-free solder joints, Soldering and Surf. Mount. Tech., 9(3), pp. 37–40 (1997).
69. T.A. Woodrow, Reliability and leachate testing of lead-free solder joints, Proceedings of the International Conference on Lead-Free Components and Assemblies, San Jose, CA, May 1–2, 2002, pp. 116–125.
70. NCMS Report 0401RE96, Lead-Free Solder Project, Final Report, National Center for Manufacturing Sciences, August 1997.
71. D. Shangguan, Study of compatibility for lead free solder PCB assembly, Proceedings of International Conference on Lead-Free Electronics, SolderTec, Brussels, June, 2003, pp. 297–308.
72. J. Lau, N. Hoo, R. Horsley, J. Smetana, D. Shangguan, W. Dauksher, D. Love, I. Menis, and B. Sullivan, Reliability testing and data analysis of high-density packages' lead-free solder joints, Proceedings of APEX 2003, Anaheim, CA, March 2003.
73. T.A. Woodrow, Reliability and leachate testing of lead-free solder joints, Proceedings of the International Conference on Lead-Free Components and Assemblies, San Jose, CA, May 1–2, 2002, pp. 116–125.
74. D. Xie, M. Arra, H. Phan, D. Shangguan, D. Geiger, and S. Yi, Life prediction of leadfree solder joints for handheld products, Proceedings of the Telecomm Hardware Solutions Conference & Exhibition, SMTA/IMAPS, Legacy Park, TX, May 2002, pp. 83–88.
75. S. Heinrich, J. Liang, and P.S. Lee, Advances in Electronic Packaging 1999, ASME InterPack'99, ASME, 1999, pp. 43–53.
76. S.M. Heinrich, P.S. Lee, and J. Liang, Analytical expressions for shear and axial joint deformations in area-array assemblies due to CTE mismatch, Proceedings, 2002 ASME International Mechanical Engineering Congress and Exposition, Vol. 2, Paper No. IMECE2002-39636, 13 pp., New Orleans, LA, November 17–22, 2002.
77. W. Engelmaier, in J.H. Lau, Ed., Solder Joint Reliability—Theory and Applications, Van Nostrand Reinhold, New York, 1991, pp. 545–587.
78. M. Kitano and M. Honda, in Advances in Electronic Packaging, 1997, AMSE InterPack'97, ASME, 1997, pp. 1407–1412.
79. R. Satoh, K. Arakawa, M. Harada, and K. Matsui, IEEE Trans. on CHMT, 14, pp. 224–232 (1991).
80. H.D. Solomon, in H.D. Solomon, G.R. Halford, L.R. Kaisand, and B.N. Leis, Eds., Low Cycle Fatigue, ASTM STP 942, ASTM, 1988, pp. 342–370.
81. J. Liang, N. Gollhardt, P.S. Lee, S.A. Schroeder, and W.L. Morris, in J.C. Suhling, Ed., Applications of Experimental Mechanics to Electronic Packaging, ASME, 1995, pp. 1–9.
82. M. Mukai, et al., Fatigue life estimation of solder joints in SMT-PGA packages, Journal of Electronic Packaging, ASME, 120(June), p. 207 (1998).
83. Q. Zhang, P. Friesen, and A. Dasgupta, Risk assessment & accelerated qualification, of Pb-free electronics, CALCE Project C03-05, 2003.
84. J. Liang, and R.M. Pelloux, Chin J. Met. Sci. & Technol., 5, pp. 1–12 (1989).
85. J.D. Morrow, Cyclic plastic strain energy and fatigue of metals, ASTM STP 378, ASTM, pp. 45–87 (1964).

11

Fatigue Life Assessment for Lead-Free Solder Joints

Masaki Shiratori and Qiang Yu

Yokohama National University, Japan

Abstract The authors will indicate the basic reliability problems associated with use of lead-free solder joints. They investigated the thermal fatigue reliability of lead-free solder joints, and focused their attention on the formation of the intermetallic compound and its effect on the initiation and propagation of the fatigue cracks. An isothermal fatigue test method was used to improve the efficiency of fatigue study. Several lead-free solder alloys, Sn-Ag-Cu, Sn-Ag-Cu-Bi, Sn-Cu and Sn-Zn-Bi, were investigated.
It was found that there were two kinds of major failure mode in lead-free solder joints: the solder bulk fatigue mode, and the interface fatigue mode. It was found also that the mode shift of the fatigue crack was affected by not only the properties of the intermetallic layer, but also by the tensile strength of the solder material. In order to investigate the influence of plating treatment on the fatigue strength of Sn-Zn-Bi solder joint, specimens with Ni/Au or Cu plating treatment on Cu-pad were used. Through a series of isothermal mechanical shear fatigue tests and FEM (Finite Element Method) analysis, it has been found that the fatigue life of Sn-Zn-Bi solder joint was greatly affected by the environmental temperature and plating conditions.

11.1. INTRODUCTION

Surface mount technologies, such as BGA (Ball Grid Array), CSP (Chip Size Package), and flip-chip, have been adopted in many electronic devices: computers, household electric appliances, and so forth. The tendency in the developments of surface mount technologies is characterized by miniaturization, performance enhancement, and the higher pin count. In addition, it is crucial that lead is removed from the solder joints. There is a need to establish regulations for the removal of Pb, especially in European countries and Japan. Many studies have been carried out recently to develop technologies for replacing Sn-Pb solder with the lead-free ones.

The authors have studied, and established an evaluation method for thermal fatigue strength for the BGA structure of the Sn-Pb eutectic solder [1–10]. It became clearer that conventional Sn-Pb eutectic solder has many merits, such as a low melting point, good wettability and high electric and mechanical reliability. It is difficult to find a lead-free material that has superior or comparative characteristics compared with the eutectic solder. At

the moment, the most favorable candidates for the lead-free solder materials are Sn-Cu for the flow soldering, and Sn-Ag-Cu, Sn-Ag-Bi and Sn-Zn-Bi for the reflow soldering. Using these solders one can achieve almost as close a reliability level as using Sn-Pb eutectic solder. There is a concern, however that the intermetallic compound formed by the thermal fatigue cycle might seriously affect the thermal fatigue strength under some special conditions [11,12].

The authors have paid attention to the role of the intermetallic compound formation between the solder bulk and the Cu-pad due to thermal cycling, in Sn-Ag-Cu (Bi), Sn-Cu and Sn-Zn-Bi lead-free solder materials [11–15]. We addressed the relation between the intermetallic compound and the thermal fatigue strength, and the relation between the mechanical properties of solder materials and the fatigue failure modes in the solder joints.

11.2. THE INTERMETALLIC COMPOUND FORMED AT THE INTERFACE OF THE SOLDER JOINTS AND THE Cu-PAD

Figure 11.1 shows an enlarged view of the microstructure around the interface of a solder bulk material and the Cu-pad. The intermetallic compound is formed at the interface if the Cu-pad is not plated by Ni. The components of the compound are Cu-Sn (Cu6-Sn5 or Cu3-Sn), in the case of eutectic solder, and Sn-Ag-Cu (Bi). The fatigue fracture mode of the solder joints, in which the intermetallic compound is remarkably formed, depends upon the solder materials, processing conditions, testing conditions, and so forth.

The fracture modes can be roughly classified into the following three types:

(1) Fatigue fracture in the solder bulk layer. It is solder fatigue failure mode. It is caused by the low cycle fatigue due to the repeated nonlinear strain. In this case, the fatigue life assessment can be carried out by using Manson-Coffin's law as a basis [3].
(2) Fatigue failure in the intermetallic compound layer or at the interface of the intermetallic compound and the solder material. It is interface fatigue mode. This failure mode is caused by the decrease in ductility due to the growth of the intermetallic compound layer.
(3) Peeling failure of the Cu-pad from the printed circuit board. This is caused by the stress concentration.

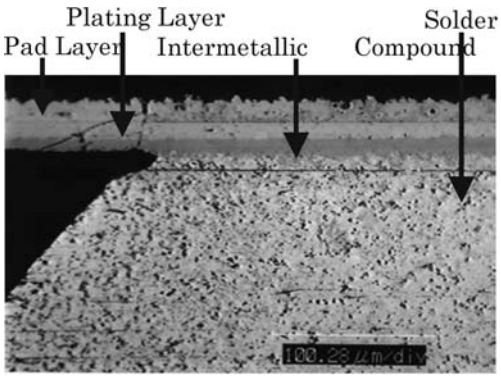


FIGURE 11.1. Structure of solder joint.

For the mode (2), the factors affecting this mode can be due to the following causes:

(1) For the intermetallic compound layer:

- Growth of the intermetallic compound under the reflow process;
- Growth of the intermetallic compound due to the dwell time at high temperature during the thermal fatigue test;
- Repeated stress concentration at the intermetallic compound layer at low temperature during the thermal fatigue test.

(2) For failure at the interface of intermetallic compound and the solder layer:

- The growth of the intermetallic compound and the repeated stress concentration;
- The coherency of the grain boundaries between the intermetallic compound and the solder bulk.

It was found that it scarcely happens that the different fracture modes arise simultaneously. In almost all the cases, only one mode of fracture is dominant. Therefore, it is important to investigate the conditions under which each fracture mode dominantly occurs. It has been found that the failure modes can be controlled and then a good guideline can be acquired to develop a new solder material and soldering process. In this paper the authors have paid attention to the difference in fracture mechanism of the solder joints. We investigated the initiation criteria of each fracture mode and the effect of the mechanical properties of solder material on these failure modes.

11.3. MECHANICAL FATIGUE TESTING EQUIPMENT AND LOAD CONDITION IN THE LEAD FREE SOLDER

The isothermal mechanical fatigue test was proposed to evaluate the fatigue reliability of the Sn-Pb solder joints instead of the thermal cyclic test [7,8,10,11,16]. It has been confirmed that the fatigue life of solder joints under power cycling can be predicted by the same Manson-Coffin's curve given by the isothermal mechanical fatigue tests. The shift in the failure modes can be checked by comparing the experimental results of the fatigue life with the value predicted from the fatigue life curve for the solder failure mode. If the fatigue life of a solder joint is much shorter than the predicted value, that means that the failure mode has changed from the "good," cohesive, failure mode to the interface failure mode or the other bad mode. The consistency of these two kinds of test methods has sufficiently been taken for the Sn-Pb solder joints case, where the failure mode was almost always the solder fatigue mode. However, the growth of the intermetallic compound in high-temperature dwell time must be considered, because the interface failure mode may become one of the major modes, when a lead free solder material is used. In this study, in order to investigate the interface fatigue behavior using the isothermal fatigue test method, a specimen with lead-free solder was heat-treated before the test to accelerate the growth of intermetallic compound. And then the cyclic mechanical deformation was directly given to the specimen to measure its fatigue life.

Figure 11.2 shows the schematic illustrations of isothermal mechanical fatigue test equipment used in this study. The specimen was fixed to the jigs by bonding both upper and lower surfaces. Then cyclic shear displacement was applied repeatedly to the upper jig. A displacement controlled fatigue test was carried out by applying triangular waves with a

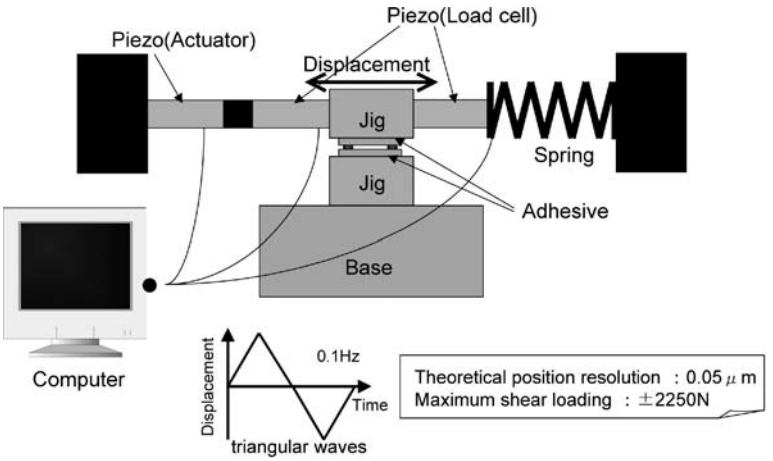


FIGURE 11.2. Isothermal fatigue-testing equipment.

constant displacement rate. The cross-sections of the solder joints were observed through a microscope with high magnifying power (maximum to 3000 magnifications on the display) during the whole fatigue cycles. The relative displacement on the upper and lower surfaces of the solder ball, $\Delta\delta_{re}$, was measured directly and automatically by the digital microscope system. Although there could be some different definitions for the number to failure N_f, in this study, it was defined as the number of cycles when the load measured by a piezo type load cell drops about 10% from the initial level of the reaction load. In order to investigate the stress and strain behavior in the solder joint under the mechanical cyclic loading, the elasto-plastic analyses were carried out by using ANSYS.

11.4. RESULTS OF MECHANICAL FATIGUE TEST

In this study, shear type mechanical fatigue tests were used to examine the failure modes and fatigue strength of lead free solder joints. Figure 11.3 shows the structure of a shearing type solder joint. Because the width of the solder layer is much longer than its height, this kind of solder joint could reduce the stress and strain concentrations at the edge of the solder layer. The lead-free solder materials used in this study are Sn-Cu-Ni, Sn-2.9Ag-0.5Cu-3Bi, and Sn-8Zn-3Bi. The test specimens were heat-treated before the testing by holding at 150°C for 200 hours to accelerate the growth of intermetallic compound as in the thermal cycle test. The Sn-8Zn-3Bi specimen was heat-treated at 110°C for 300 hours. The fatigue life of each solder joint was evaluated from the results of elasto-plastic analyses and the mechanical fatigue test. The results of the measured fatigue life for each lead free solder joint are shown in Figures 11.4–11.9.

Figure 11.4 shows that Sn-Cu-Ni solder joints have almost identical fatigue strength whether they were heat-treated or not. The fatigue failure mode was observed in details by the microscope to verify the above results. It was observed that the crack initiated within the solder bulk (solder fatigue mode) for both cases, as shown in Figure 11.5.

Figure 11.6 shows that the results of the fatigue life of the heat-treated specimen of Sn-2.9Ag-0.5Cu-3Bi decreased a little in comparison with the specimen that was not heat-treated. It was shown that the fatigue cracks were initiated within the matrix of sol-

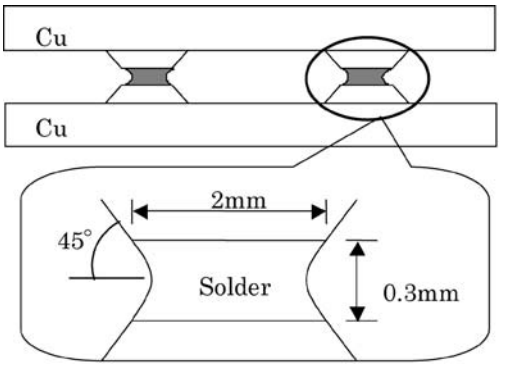


FIGURE 11.3. Structure of shearing type specimen.

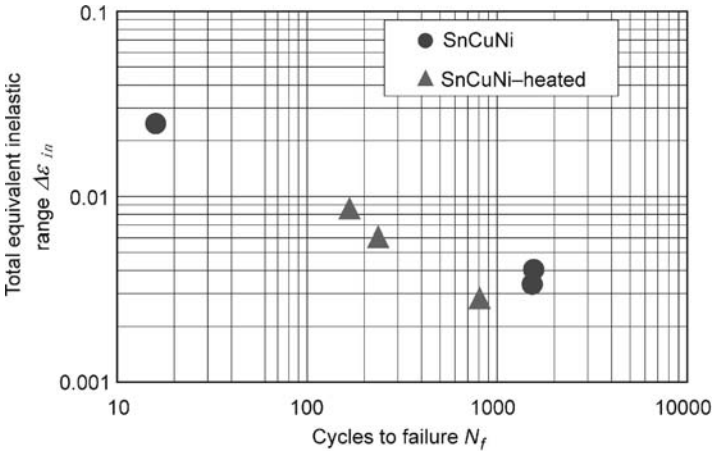
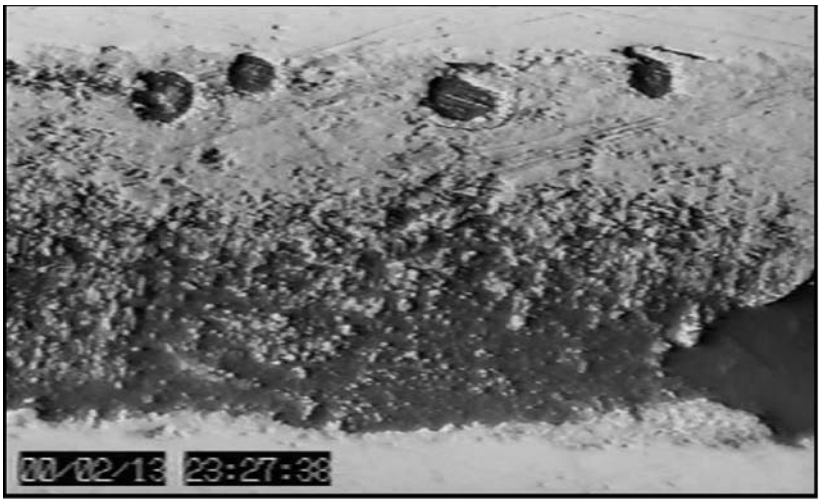


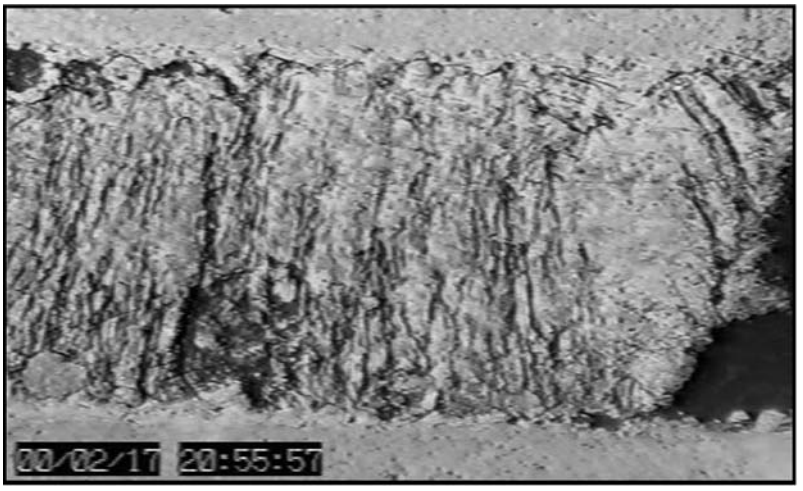
FIGURE 11.4. S–N curves of Sn-Cu-Ni solder joints.

der material (solder fatigue mode) in the case of not heat-treated specimen. On the other hand, for the heat-treated specimen, the fatigue cracks initiated in the solder layer where very close to the interface between the intermetallic compound layer and solder bulk. The cracks propagated with a path parallel to the interface (similar to an interface fatigue mode). This case was called solder/interface fatigue mode. Optical micrographs are shown in Figure 11.7.

Figure 11.8 shows that the fatigue life of the not heat-treated Sn-8Zn-3Bi specimen was much longer than that of the specimen heat-treated at 150°C for 200 hours. It was observed that the cracks initiated in the solder bulk (solder fatigue mode) in the case of not heat-treated specimen, and the cracks in the specimen heat-treated for 300 hours at 110°C initiated and propagated with the similar mode of Sn-2.9-0.5Cu-3Bi solder joint. However, the fatigue cracks initiated at the interface between the intermetallic compound layer and solder (a true interface fatigue mode) for the case of specimen heat-treated for 200 hours at 150°C. It was assumed that in this case there existed a reaction layer (CuSn) newly grown between the solder and intermetallic layer (ZnCu), and the cracks just started in this new layer. Figures 11.9 show the optical micrographs of the fatigue fracture modes of Sn-Zn-Bi solder joints. The above results showed that the shift in the fatigue fracture mode in the



(a)



(b)

FIGURE 11.5. Optical micrograph of fatigue crack in Sn-Cu-Ni solder joints. (a) Solder joint without high temperature exposure. (b) Solder joint after 200 hours exposure at 150°C.

solder joints from the cohesive failure in the solder to the interface mode might result in a significant decrease in the fatigue life.

The other type of fatigue test specimen used in this study is the BGA solder joints, where a dummy component was connected to a Cu-plated substrate by two lead-free solder balls. The lead-free solder materials used in these specimens are Sn-3.5Ag-0.75Cu and Sn-3.5Ag-5Bi. These specimens were heat treated by holding at 120°C for 360 hours to grow the intermetallic layer, and the specimens are classified into four types (Table 11.1).

Figure 11.10 shows the results obtained by the mechanical fatigue tests and elasto-plastic analyses. The figure shows the following:

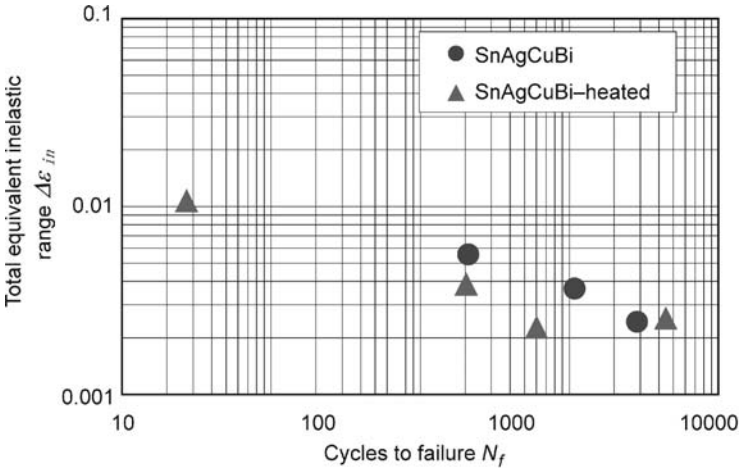


FIGURE 11.6. S–N curves of Sn-2.9Ag-0.5Cu-3Bi solder joints.

TABLE 11.1.
Types of BGA specimens.

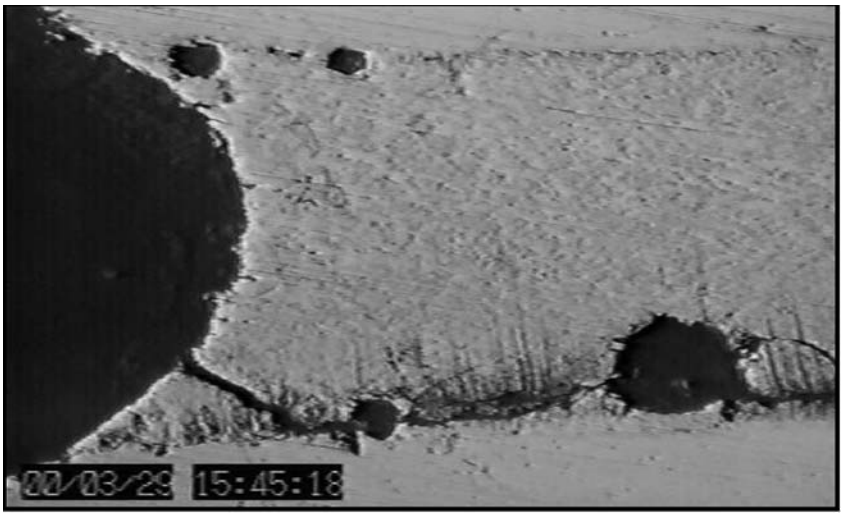
	No heat treatment	Heat treated
Sn-3.5Ag-0.75Cu	B1	B2
Sn-3.5Ag-5Bi	C1	C2

(1) the effect of heat treatment may be negligible in the case of Sn-3.5Ag-0.75Cu solder joints;
(2) the fatigue lives of the heat treated specimen of Sn-3.5Ag-5Bi (C2) decreased remarkably.

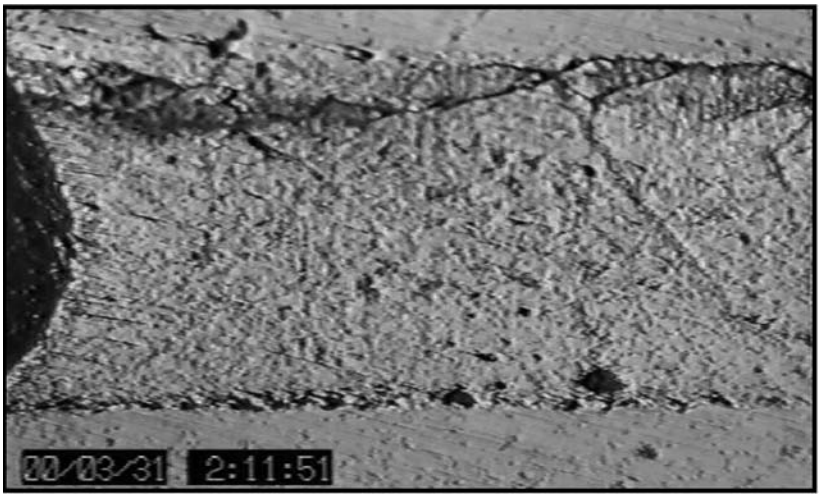
Optical micrographs of the BGA specimens after the fatigue test are shown in Figures 11.11 and 11.12. Similar to Sn-Cu-Ni solder joints, the fatigue cracks in both kinds of BGA Sn-3.5Ag-0.75Cu joints (B1 and B2) started and propagated within the solder bulk layer. It means that since Sn-Ag-Cu solder joints broke with a solder fatigue mode, the fatigue life was not affected by the conditions of reaction interface between the solder and the Cu pad. The situation was the same for the Sn-3.5Ag-5Bi specimen without heat treatment (C1), where the remarkable growth of the intermetallic compound layer could not be observed. On the other hand, the interface in heat-treated Sn-3.5Ag-5Bi solder joints (C2), where the remarkable growth of intermetallic compound layer was observed, broke and the interface fracture mode resulted in great decease in the fatigue life.

11.5. CRITICAL FATIGUE STRESS LIMIT FOR THE INTERMETALLIC COMPOUND LAYER

As a result of the mechanical fatigue test, it was found that the fatigue failure mode of the solder joints depended upon the solder material and the condition of the reaction layer. To explain this phenomenon, the authors have observed the nonlinear stress–strain characteristics of solder materials (Figure 11.13).



(a)



(b)

FIGURE 11.7. Optical micrograph of fatigue crack in Sn-2.9Ag-0.5Cu-3Bi solder joints. (a) Solder joint without high temperature exposure. (b) Solder joint after 200 hours exposure at 150°C.

The figure shows that the flow stress of Sn-3.5Ag-5Bi is remarkably higher than those of Sn-Pb and Sn-3.5Ag-0.75Cu. It means that Sn-3.5Ag-0.75Cu like Sn-37Pb has a much lower yield stress to let the solder layer to deform easily as a cushion for hard and fragile intermetallic compound layer. The stress applied on the interface layer can not increase high enough to break the reaction layer (intermetallic layer) during the fatigue test. However, lead-free materials with rich Bi are much harder than the Sn-Pb eutectic solder and more difficult to deform, and the higher yield stress may cause a very critical stress level on the interface layer. This may be the reason why a very fast fracture, which was closer to a brittle fracture than to a slow ductile fracture, occurred in and along the intermetallic compound. However, this kind of brittle fracture just propagated with a very

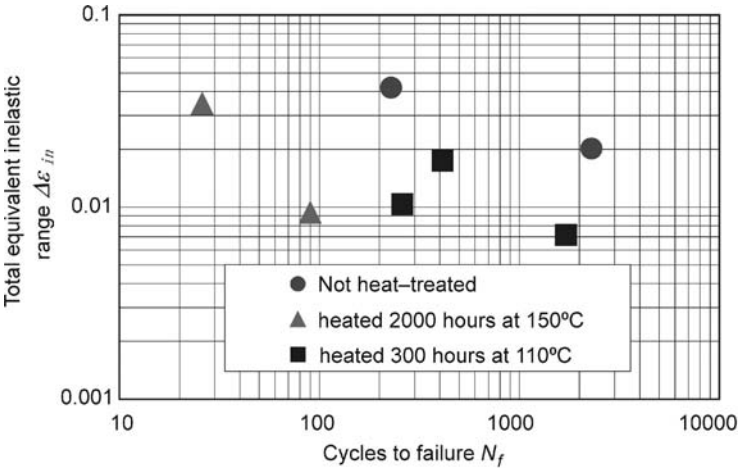


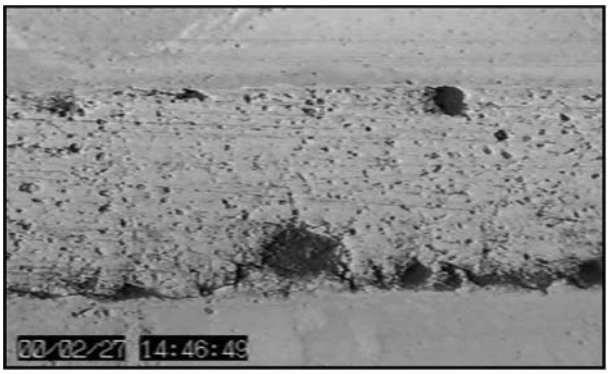
FIGURE 11.8. S–N curves of Sn-8Zn-3Bi solder joints.

TABLE 11.2.
Relation of fracture mode to tensile strength.

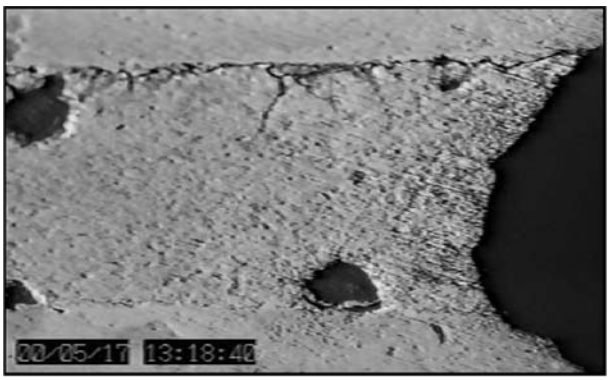
Solder material	Tensile strength (MPa)	Fracture mode
Sn-37Pb	38	Solder
Sn-Cu-Ni	35	Solder
Sn-3.5Ag-0.75Cu	49	Solder
Sn-2.9Ag-0.5Cu-3.0Bi	58	Solder/Interface
Sn-8Zn-3Bi	66	Solder/Interface (110°C) Interface (150°C)
Sn-3.5Ag-5Bi	68	Interface

small scale in every cycle because the displacement controlled load could cause of stress relaxation quickly after the crack propagation. In the case of the specimen C2, there existed the well-grown intermetallic compound layer due to the exposure at high temperature. The high stress level due to the high yield stress of Sn-Ag-5Bi caused the brittle fracture at the intermetallic compound (interface fatigue mode). This resulted in the remarkable decrease in the fatigue life. It can be assumed that the same brittle fracture occurred also in Sn-8Zn-3Bi joints after 200 hours exposure at 150°C, since the similar intermetallic compound layer to Sn-3.5Ag-5Bi joint existed at its interface. However, Sn-8Zn-3Bi could break only the interface of the ZnCu intermetallic compound layer in the solder joint after 300 hours exposure at 110°C. Because the toughness of the ZnCu intermetallic compound is said to be higher than that of SnCu, this failure mode may not be a big problem.

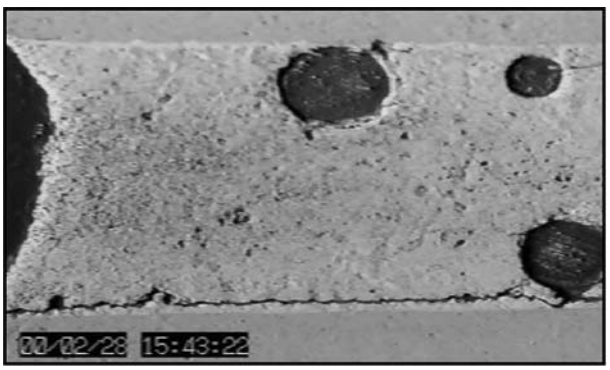
The relation between the tensile strength (flow strength at 2% strain) and the fatigue fracture mode of each solder joint is shown in Table 11.2. The fatigue fracture mode shifts as the solder hardens, from the solder mode to the interface mode, it is necessary to pay attention to the stress concentration, as well as to the strain behavior when one wants to assess the thermal fatigue strength of the solder joints with intermetallic compound layers [6]. Figure 11.14 shows the distributions of Mises equivalent stresses at the interface layer in the solder joint of Sn-37Pb, Sn-3.5Ag-0.75Cu, and Sn-3.5Ag-5Bi, respectively.



(a)



(b)



(c)

FIGURE 11.9. Optical micrograph of fatigue crack in Sn-8Zn-3Bi solder joints. (a) Solder joint without high temperature exposure. (b) Solder joint after 300 hours exposure at 110°C. (c) Solder joint after 200 hours exposure at 150°C.

Stress concentration at the end of the interface in the Sn-37Pb solder joint is the lowest, and that of the Sn-3.5Ag-5Bi is the highest. The maximum stresses are limited by the yield stresses of the solder materials. It is impossible for the maximum stresses to increase over the yield stresses. The maximum stress of each kind of solder joint corresponds to the sol-

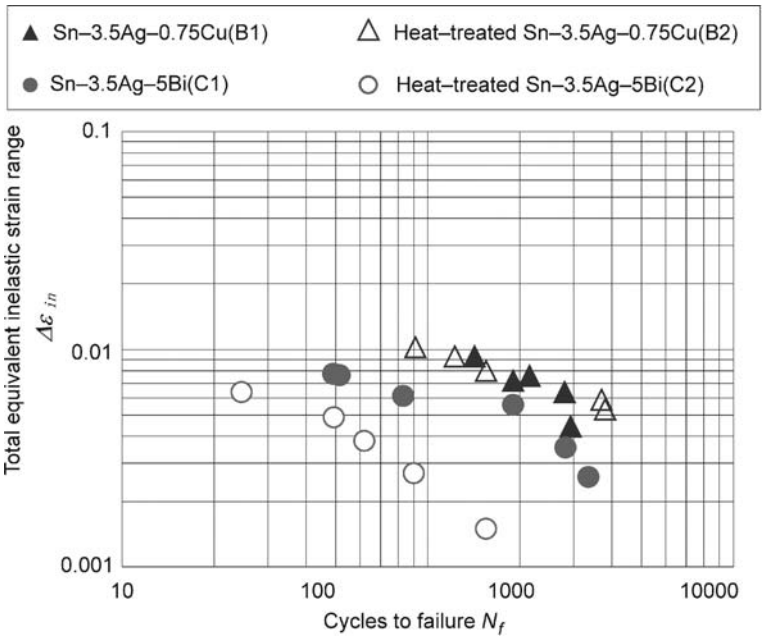


FIGURE 11.10. S–N curve of BGA solder joints.

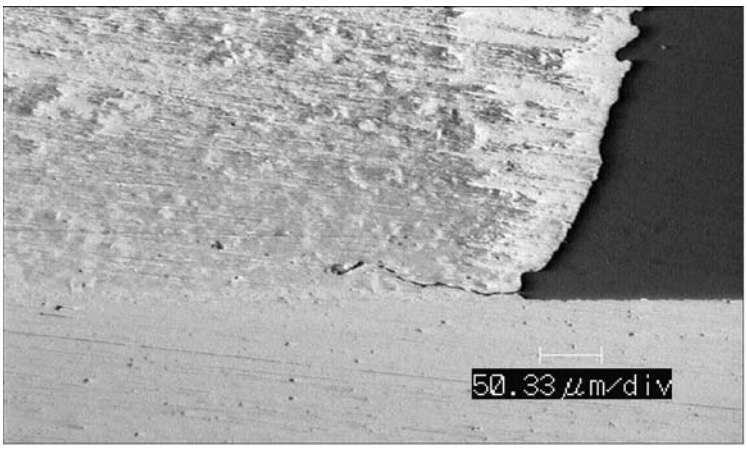


FIGURE 11.11. Crack in Sn-3.5Ag-0.75Cu BGA solder joint (B2).

der strength stress as shown in Table 11.2. This is because almost all the stress components in the solder domain are continuous to those in the intermetallic layer at the interface, and the stresses arising in the solder domain are constrained by its nonlinear stress–strain relationship. Therefore, there is no singularity problem for the stress distribution for the plastic deformation case.

Based upon the above considerations and also from the results of the mechanical fatigue test, it can be assumed that there might exist a critical stress (or tensile strength of solder material) limit against interface fatigue fracture mode for each intermetallic com-

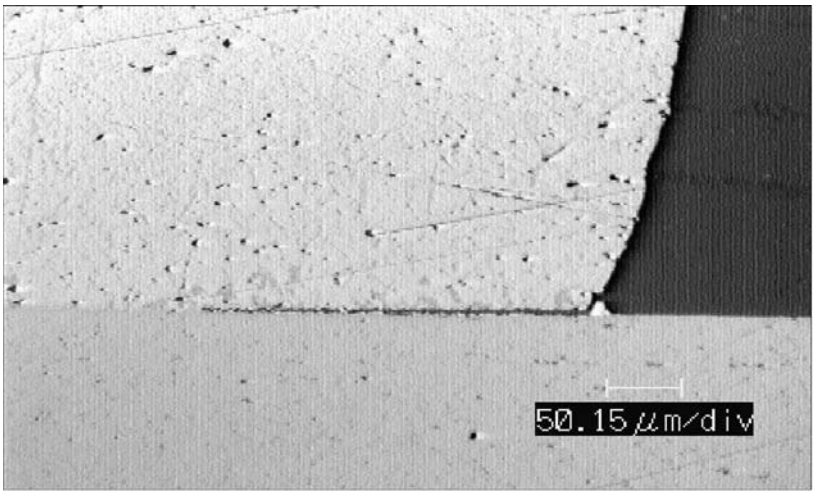


FIGURE 11.12. Crack in Sn-3.5Ag-5Bi BGA solder joint (C2).

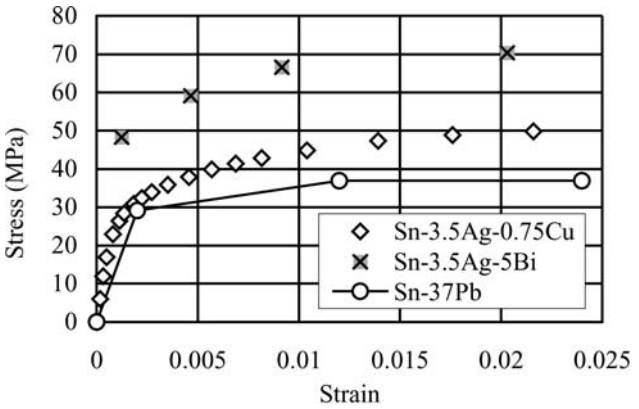


FIGURE 11.13. Stress–strain curve of solder materials.

pound SnCu and ZnCu. If the tensile strength of solder is lower than the limit, the fatigue cracks will appear within the solder bulk layer. This is the solder fatigue failure mode. The fatigue life of these solder joints is dominated by Manson-Coffin's law in the bulk solder material. However, if the tensile strength is higher than this limit, the brittle interface fatigue occurs much earlier than the solder fatigue life. This is the interface fatigue mode.

Summarizing the above results, we conclude that the critical stress limit for intermetallic SnCu should be drawn between the maximum stresses of Sn-3.5Ag-0.75Cu and Sn-3.5Ag-5Bi as shown in Figure 11.15, or, to be exact, the critical stress should be set close to the tensile strength of Sn-2.9Ag-0.5Cu-3Bi, 58 MPa, because it has been checked that the major failure mode of this solder joints was something like a mixed mode of solder and interface modes (solder/interface mode). This means that the failure mode of Sn-2.9Ag-0.5Cu-3Bi joints located just at the borderline of the two major failure modes. In the case of Sn-3.5Ag-0.75Cu, most of the cracks initiated within the solder layer, since its tensile

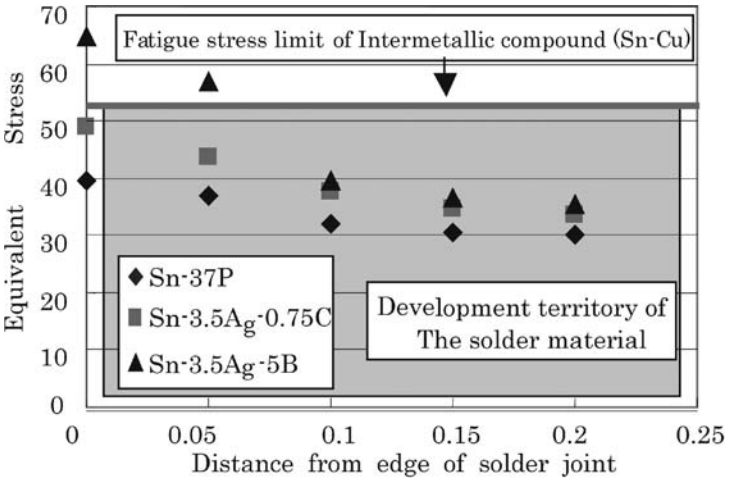


FIGURE 11.14. Distributions of Mises equivalent stresses.

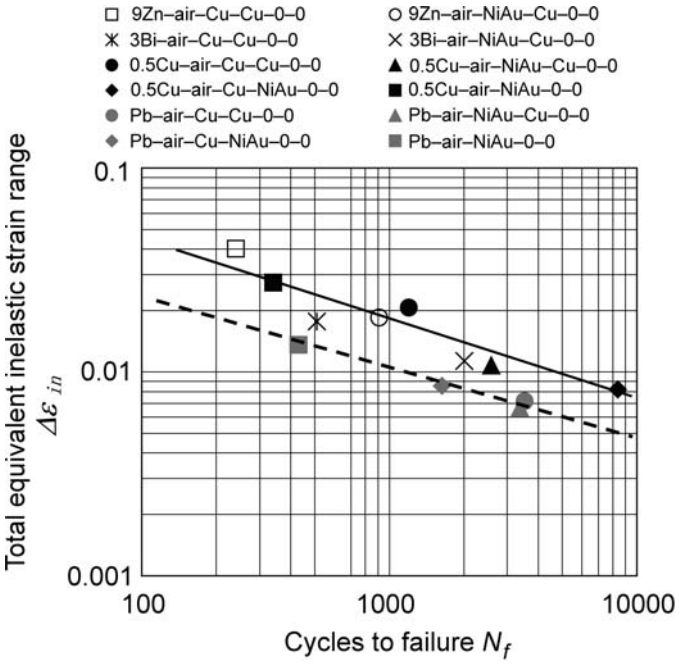


FIGURE 11.15. Fatigue life of Sn-Zn solder joints with initial state.

strength limits the maximum equivalent stress concentration at the end of the solder joint not to cross over the critical stress. For the same reason, it was very rare to find an interface fatigue failure mode in the Sn-Pb solder joints.

On the other hand, in the case of Sn-3.5Ag-5Bi, the tensile strength is higher than the critical stress, and, as a result, the brittle fracture at the intermetallic compound caused a

great decrease in the fatigue life. Based upon the results shown in Figures 11.8 and 11.9, it can be assumed that the limit stress for ZnCu is roughly between 60 to 68 MPa.

The above consideration shows that the reliability engineers have to think of the balance of the strengths of both the intermetallic compound and the solder matrix, if they want to achieve high enough fatigue reliability of a lead-free solder joint. In order to avoid the brittle fracture at the intermetallic compound layer, one has to consider the flow characteristics of the lead-free material, or to manage to prevent the growth of the intermetallic compound.

11.6. INFLUENCE OF THE PLATING MATERIAL ON THE FATIGUE LIFE OF Sn-Zn (Sn-9Zn AND Sn-8Zn-3Bi) SOLDER JOINTS

In order to investigate the influence of plating treatment on the fatigue strength of Sn-Zn solder joints, the specimens with Ni/Au or Cu plating treatment on the Cu-pad at the side of substrate were prepared. These specimens were aged for 200, 500, 1000 hours at 85, 125, 150°C to investigate the influence of the fatigue strength due to the growth of the intermetallics. Some specimens were exposed to the high-humidity/temperature environment (85%/85°C).

Figure 11.15 shows the results of the solder joints without high temperature aging. As shown in Figure 11.15, the specimen has been characterized and named by solder material [9Zn, 3Bi, 0.5Cu, Pb], reflow environment [air, N_2], plating materials on the substrate side [Ni/Au, Cu], plating materials on the chip side [Ni/Au, Cu], aging temperature [85°C, 125°C, 150°C, 85°C (85%)], aging time [100 h, 200 h, 500 h, 1000 h]. Figure 11.15 shows the relation between the fatigue life N_f and inelastic equivalent strain range $\Delta\varepsilon_{in}$ of initial state (non-aging) for Sn-9Zn, Sn-8Zn-3Bi, Sn-3Ag-0.5Cu and Sn-Pb CSP solder joints. In Figure 11.15, the dotted line represents the fatigue life due to the mechanical shear fatigue test for Sn-Pb solder as a reference. The solid line expresses the result of Sn-3Ag-0.5Cu solder. It was found that Sn-9Zn and Sn-8Zn-3Bi solder, without aging, have similar fatigue strength to Sn-3Ag-0.5Cu solder.

Figure 11.16 shows the fatigue strength of a Sn-9Zn specimen. The results of the fatigue test are divided into two groups. The dotted line shows the average results of the mechanical shear fatigue test for Sn-Pb solder. The solid line shows the average results of specimen groups with good fatigue mode. Not-aged specimens are located on this line. In the case of the Cu plating specimen, the aged specimens have similar fatigue strength to the initial state (non-aging) specimens, even if heat treatment was carried out at 150°C for 1000 hours. If the specimens with Ni/Au plating treatment were aged at 85°C for 1000 hours or a much higher temperature condition, these fatigue strength decreased. Although fatigue strength becomes lower, they have the almost identical fatigue strength with Sn-Pb solder (dotted line). However, when the specimens were exposed in the high-humidity/temperature environment, the fatigue strength greatly decreased. This is because Zn is more oxidizable than Sn, Ag, and Cu under high-humidity condition, and the oxidized materials might have a critical effect on the fatigue strength.

Figure 11.17 shows the results of the mechanical fatigue test for Sn-8Zn-3Bi solder joints. The dotted line shows the average results of mechanical shear fatigue test for Sn-Pb solder. The solid line expresses the approximate line of fatigue strength for Sn-8Zn-3Bi solder. The aged Sn-8Zn-3Bi solder specimens with Cu plating treatment show the similar fatigue strength to the initial state (non-aging) specimen. When the specimens with Ni/Au

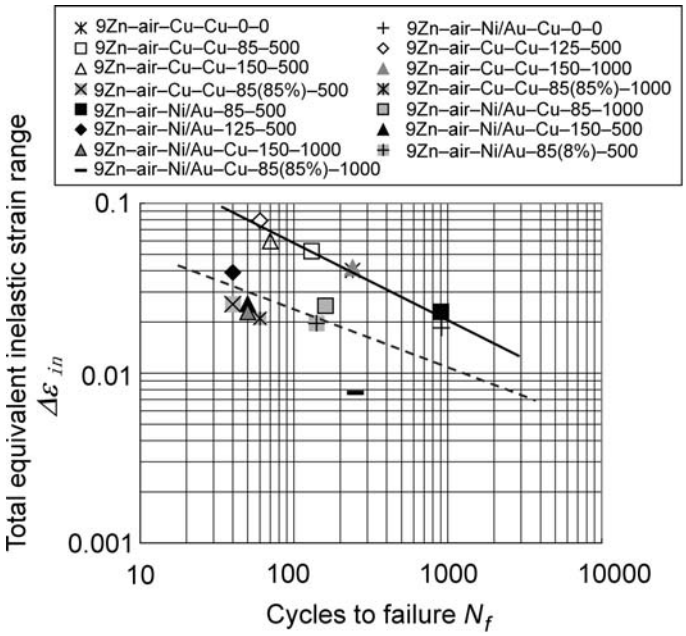


FIGURE 11.16. Fatigue life of Sn-9Zn solder joint.

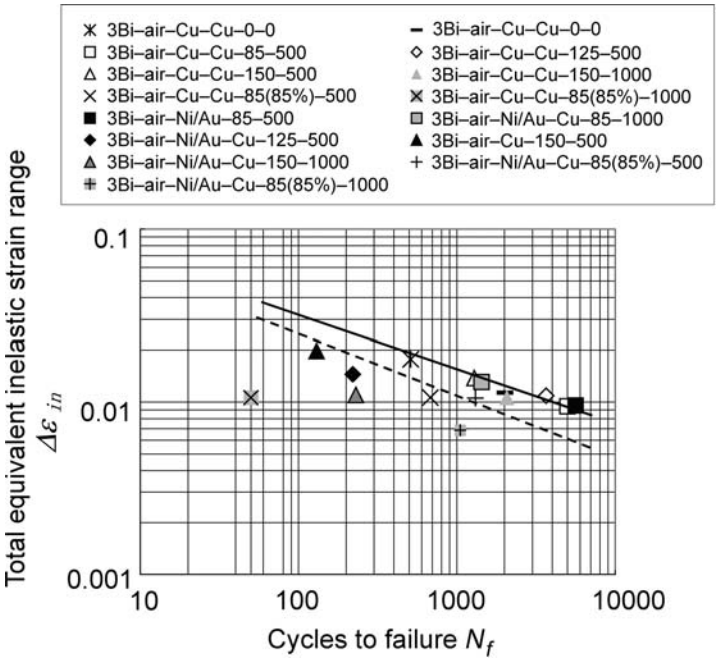


FIGURE 11.17. Fatigue life of Sn-8Zn-3Bi solder joint.

plating treatment were heat aged at 150°C for 1000 hours, their fatigue strength decreased more than that of the Sn-Pb solder. This means the Ni/Au can shelter the growth of the intermetallic layer in Sn-Zn(Bi) solder joints in almost all the cases. One should be very careful when using Sn-Zn(Bi) solder materials for electronics systems that have a high possibility to be used in high humidity environments.

11.7. CONCLUSION

The effect of material property of lead-free solder on the reliability of solder joints was addressed. There exist two kinds of major fatigue failure modes: the solder fatigue mode, and the interface fatigue mode. The interface fatigue mode results in a significant decrease in the fatigue life of solder joints. The mode transition from the solder to the interface is not only affected by the conditions of the reaction layer but is also controlled by the tensile strength of the solder material. The fatigue strength of 95.75Sn-3.5Ag-0.75Cu, Sn-Cu-Ni, Sn-2.9Ag-0.5Cu-3Bi lead free solder joints are not greatly affected by the intermetallic compounds formed during thermal cycles, as the Sn-Pb solder does. Their tensile strengths are lower than the critical limit, and the fatigue fracture is dominated by the solder fatigue mode. However, in the case, when the solder material contains much Bi (over 5%), the crack initiates easily along the interface between the intermetallic compound layer and the solder material. This is because the high-level tensile strength of Sn-3.5Ag-5Bi lead free solder causes severe stress distribution in the interface.

Based upon the obtained results, the possibility of giving guidelines for the new solder material development was suggested by showing the limit stress of fatigue strength of Sn-Cu intermetallic compound. It seems to be possible to obtain a limit stress for the intermetallic compound of Zn-Cu, as well as of Sn-Cu.

In the Sn-9Zn and Sn-8Zn-3Bi solder joints, when their Cu-pads were plated with Cu component, the fatigue strength does not get lower. Though the fatigue strength is reduced when the specimens are plated with Ni/Au and aged at 125°C or less, they have superior fatigue strength that the Sn-Pb solder.

When the Ni/Au plating specimens were aged over 150°C, the fatigue strength at the solder joints decreased greatly. This is caused by the growth of the interface fatigue crack. When the Sn-9Zn and Sn-8Zn-3Bi solder were exposed to high-humidity/temperature environments, their fatigue strength is decreased greatly. From the results, if these lead-free solder materials are used carefully with correct use environment and reflow conditions, practical applications of the Sn-Zn-(Bi) lead-free solder seem to be possible.

REFERENCES

1. Q. Yu and M. Shiratori, A study of mechanical and thermal stress behavior due to global and local thermal mismatch of dissimilar materials in electronic packaging, Proc. of the International Intersociety Electronic Packaging Conference (InterPack'95), EEP-Vol. 10, No. 1, 1995, pp. 389–394.
2. M. Shiratori, Q. Yu, and S.B. Wang, A computational and experimental hybrid approach to creep-fatigue behavior of surface-mounted solder joints, Proc. of the International Intersociety Electronic Packaging Conference (InterPack'95), EEP-Vol. 10, No. 1, 1995, pp. 451–457.
3. M. Shiratori and Q. Yu, Fatigue-strength prediction of microelectronics solder joints under thermal cyclic loading, Proc. of Intersociety Conference on Thermal Phenomena in Electronic Systems (I-Therm V), 1996, pp. 151–157.

4. M. Shiratori and Q. Yu, Life assessment of solder joint, advances in electronic packaging, Proc. of the Advances in Electronic Packaging 1997: Proceedings of the Pacific Rim/ASME International Intersociety Electronic & Photonic Packaging Conference (InterPack'97), EEP-Vol. 19, No. 2, 1997, pp. 1471–1477.
5. Q. Yu, M. Shiratori, and N. Kojima, Fatigue crack propagating evaluation of microelectronics solder joints, Proc. of the Advances in Electronic Packaging 1997: Proceedings of the Pacific Rim/ASME International Intersociety Electronic & Photonic Packaging Conference (InterPack'97), EEP-Vol. 19, No. 2, 1997, pp. 1445–1450.
6. Q. Yu and M. Shiratori, Fatigue-strength prediction of micro-electronics solder joints under thermal cyclic loading, IEEE Transactions on Components, Packaging, and Manufactuting Technology, Part. A, 20(3), pp. 266–273 (1997).
7. Q. Yu, M. Shiratori, and Y. Ohshima, A study of the effects of BGA solder geometry on fatigue life and reliability assessment, Proc. of the 6th InterSociety Conference on Thermal and Thermomechanical Phenomena in Electronic System (Itherm'98), 1998, pp. 229–235.
8. Q. Yu and M. Shiratori, Thermal fatigue reliability assessment for solder joints of BGA assembly, Proc. of the Advances in Electronic Packaging 1999: Proceedings of the Pacific Rim/ASME International Intersociety Electronic & Photonic Packaging Conference (Interpack'99), 1999, pp. 239–246.
9. Y. Kaga, Q. Yu, and M. Shiratori, Thermal fatigue assessment for solder joints of underfill assembly, Proc. of the Advances in Electronic Packaging 1999: Proceedings of the Pacific Rim/ASME International Intersociety Electronic & Photonic Packaging Conference (Interpack'99), pp. 271–275 (1999).
10. M. Ito, Q. Yu, and M. Shiratori, Reliability estimation for BGA solder joints in organic PKG, Proc. of the Advances in Electronic Packaging 2001: Proceedings of the Pacific Rim/ASME International Intersociety Electronic & Photonic Packaging Conference (InterPACK'01), 2001, pp. 1–6.
11. H. Amano and Q. Yu, Effect of interfacial factors on fatigue lifetime of lead-free die-attach solder joint, Proc. of the Advances in Electronic Packaging 2001: Proceedings of the Pacific Rim/ASME International Intersociety Electronic & Photonic Packaging Conference (InterPACK'01), 2001, pp. 1–8.
12. Q. Yu, D.S. Kim, and M. Shiratori, The effect of intermetallic compound on thermal fatigue reliability of lead-free solder joints, Proc. of the Advances in Electronic Packaging 2001: Proceedings of the Pacific Rim/ASME International Intersociety Electronic & Photonic Packaging Conference (InterPACK'01), 2001, pp. 1–8.
13. Q. Yu, D.S. Kim, J.C. Jin, Y. Takahashi, and M. Shiratori, Fatigue strength evaluation for Sn-Zn-Bi lead—free solder joints, Proc. of the ASME International Mechanical Engineering Congress & Exposition (IMECE2002), Paper No. 39686, 2002.
14. D.S. Kim, Q. Yu, T. Shibutani, and M. Shiratori, Nonlinear behavior study on effect of hardening rule of lead free solder joint, Proc. of the Advances in Electronic Packaging 2003: Proceedings of the Pacific Rim/ASME International Intersociety Electronic & Photonic Packaging Conference (Interpack'03), No. 35250, 2003, pp. 1–7.
15. D.S. Kim, Q. Yu, T. Shibutani, N. Sadakata, and T. Inoue, Effect of void formation on thermal fatigue reliability of lead-free solder joints, Proc. of the Ninth Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems (Itherm 2004), 2004, pp. 325–329.
16. M. Kitano, W. Kumazawa, and S. Kawai, A new evaluation method for thermal fatigue strength of solder joint, ASME, Advances in Electronic Packaging EEP-Vol. 1-1, p. 301 (1992).

12

Lead-Free Solder Materials: Design For Reliability

John H.L. Pang

School of Mechanical and Aerospace Engineering, Nanyang Technological University, Republic of Singapore

12.1. INTRODUCTION

The main driving force for the paradigm shift to lead-free (Pb-free) solders are the legislations in the European Union (EU) to ban the usage of hazardous materials such as lead (Pb) in electrical and electronic equipment by 1st July 2006. The EC Directives on Restriction of Hazardous Substances (ROHS) and Waste of Electrical and Electronic Equipment (WEEE) will ban usage of Pb in electronic products by 1st July 2006 [1,2]. The global electronic industry has rallied for implementation of Pb-free solder manufacturing [3]. The National Electronics Manufacturing Initiative (NEMI) organization [4] provides guidelines towards Pb-free soldering solutions. NEMI recommends Pb-free 95.5Sn-3.9Ag-0.6Cu solder alloy for solder reflow process and 99.3Sn-0.7Cu solder alloy for wave soldering process. Pb-free solder technology involves multi-disciplinary knowledge in Pb-free solder alloys, material properties, compatibility with IC components and PCB materials, soldering and surface mount technology (SMT), and Pb-free solder joint reliability. After more than four decades of user knowledge of lead-based or Pb-based solders in electronics manufacturing the global electronic industry is making a paradigm shift to lead-free or Pb-free solders and soldering technology. For solder reflow process, consolidation on Sn-xAg-yCu solder alloy falls in the range of Sn-(3–4)Ag%wt-(0.5–0.8)Cu%wt. The control of such small content in percentage of (3–4)Ag%wt and (0.5–0.8)Cu%wt in the solder material and soldering process is a challenge for verification and certification. The definition of Pb-free solder should also quantify the tolerance of impurities of Pb (0.1–0.2)%wt allowed in Pb-free soldered assemblies.

In this chapter a Design-for-Reliability (DFR) methodology [5,6] for evaluation of Pb-free solder performance in electronic packaging assemblies will be described. The methodology provides a comprehensive knowledge base for Pb-free solder materials [7–10], constitutive models, fatigue life prediction, and finite element modeling of reliability test methods [11–20]. Materials, mechanics, process and reliability characterizations are needed for design-for-reliability (DFR) assessment of Pb-free solder performance. Value

added capabilities of DFR strategies with comprehensive knowledge base will help industry understand Pb-free solder joint reliability problems [21–33].

12.2. MECHANICS OF SOLDER MATERIALS

Solder materials are usually subjected to thermal cycling reliability tests over temperature cycling in the range between 0.5 to 0.8 times of the melting temperature, T_m. The coefficient of thermal expansion (CTE), mismatch deformations in the solder joints are non-linear plastic and creep strains. The inelastic strain and strain energy density in the solder joints can be numerically computed by finite element analysis (FEA). Materials testing and characterization of the mechanical behavior of solder is needed to enable FEA applications. The constitutive models used to describe the solder material deformation can be categorized into the elastic-plastic-creep analysis approach and the viscoplastic analysis approach.

Elastic-Plastic-Creep Analysis Elastic-plastic-creep approach to the prediction of the Pb-free solder joint deformation computes the total strain as the sum of the phenomenological components of elastic strain, time independent plastic strain and time dependent creep strain as shown below.

$$\varepsilon_{ij} = \varepsilon_{ij}^{e} + \varepsilon_{ij}^{p} + \varepsilon_{ij}^{c}, \tag{12.1}$$

where ε_{ij}, ε_{ij}^{e}, ε_{ij}^{p} and ε_{ij}^{c} are the total, elastic, time-independent plastic and time-dependent creep strains respectively [34,35].

The steady state creep strain rate for the intermediate-to-high stress power law breakdown region can be described by a hyperbolic-sine expression [36,37]:

$$\dot{\varepsilon} = C[\sinh(\alpha\sigma)]^{n} \exp\left(-\frac{Q}{kT}\right), \tag{12.2}$$

where $\dot{\varepsilon}$ is the creep strain rate as a function of stress, σ, and temperature, T. The material constants, C is a coefficient, n is the exponent, α prescribes the stress level at which the power dependence break down and Q is the activation energy and k is the Boltzmann's constant.

Elastic-plastic-creep approach is a phenomenological way to describe the constitutive behavior of solder alloy. The elastic, plastic and creep strains represent different deformation. However, for thermal-mechanical cyclic loading, the time-independent plastic strain and time-dependent creep strain is difficult to measure, but these strain can be readily computed from the elastic-plastic-creep analysis using FEA.

Viscoplastic Analysis Viscoplastic strain analysis approach combines the creep and plastic strain components as a unified state variable inelastic strain. The total strain,

$$\varepsilon_{ij} = \varepsilon_{ij}^{e} + \varepsilon_{ij}^{in}, \tag{12.3}$$

where ε_{ij}^{in} is the inelastic strain.

One commonly used visco-plastic model for solder is the Anand model, which was reported by Anand et al. [38,39]. The Anand model consists of two coupled differential

equations that relate the inelastic strain rate to the rate of deformation resistance. The strain rate equation is,

$$\frac{d\varepsilon_{in}}{dt} = A\left[\sinh\left(\xi\frac{\sigma}{s}\right)\right]^{\frac{1}{m}} \exp(-Q/RT), \tag{12.4}$$

and the rate of deformation resistance equation is

$$\dot{s} = \left\{h_0(|B|)^{\alpha}\frac{B}{|B|}\right\}\frac{d\varepsilon_p}{dt}, \tag{12.5}$$

where,

$$B = 1 - \frac{s}{s^*}, \tag{12.6}$$

and

$$s^* = \hat{s}\left[\frac{1}{A}\frac{d\varepsilon_p}{dt}\exp(-Q/RT)\right]. \tag{12.7}$$

Hence, $d\varepsilon_{in}/dt$ is the effective inelastic strain rate, σ is the effective true stress, s is the deformation resistance, T is the absolute temperature, A is pre-exponential factor, ξ is stress multiplier, m is strain rate sensitivity of stress, Q is activation energy, R is universal gas constant, h_0 is hardening/softening constant, $\hat{s}$ is coefficient for deformation resistance saturation value, n is strain rate sensitivity of saturation value, and α is strain rate sensitivity of hardening or softening.

12.2.1. Fatigue Behavior of Solder Materials

Fatigue behavior of solders is an important property for fatigue life prediction based on thermal cycling tests for soldered assemblies. Typically, thermal cycling test failures are precipitated over 1000 to 10,000 cycles, which is related to low cycle fatigue failure mechanism. There are two types of approaches to evaluate the fatigue life: the total life approach and crack initiation and propagation approach. For the total life approach, some fatigue parameters such as strain or strain energy density are used to predict the total fatigue life. However, the crack initiation and propagation approach predicts the fatigue life with a pre-existing crack, the fatigue life is the number of cycles to propagate this crack to its critical size.

Total Life Approach Total life approach can be divided into three major methods, based on the driving force parameter used to characterize the fatigue damage process. These three parameters are (a) plastic strain range, (b) creep strain range and, (c) inelastic energy. The strain range based fatigue approach can be further divided into plastic strain range and creep strain range methods. Plastic strain deformation focuses on the time-independent plastic effect, while creep strain accounts for the time-dependent effects. The energy-based fatigue models employ the stress–strain hysteresis energy of the solder specimens.

Plastic Strain Range Fatigue Models The Coffin-Manson fatigue model [40,41] is perhaps the best-known and most widely used approach today. The total number of cycles to failure, N_f, is depicted as being dependent on the plastic strain range, $\Delta\varepsilon_p$, the fatigue ductility coefficient, ε'_f, and the fatigue ductility exponent, c:

$$\frac{\Delta\varepsilon_p}{2} = \varepsilon'_f(2N_f)^c. \tag{12.8}$$

The fatigue ductility coefficient, ε'_f, is approximately equal to the true fracture ductility, ε_f. The fatigue ductility exponent, c, varies between -0.5 and -0.7 [42]. Comprehensive fatigue test data have been reported for 63Sn-37Pb solder [43]. For DFR applications, FEA can be used to determine the plastic strain, which is then used to predict the fatigue life.

Solomon's low cycle fatigue model [44] relates the plastic shear strain range to fatigue life:

$$\Delta\gamma_p N_p^{\alpha} = \theta, \tag{12.9}$$

where $\Delta\gamma_p$ is the plastic shear strain range. N_p is the number of cycles to failure, θ is the inverse of the fatigue ductility coefficient, and α is a material constant.

Creep Strain Range Fatigue Models Creep strain based fatigue models account for the cyclic creep deformation in the solder joints. Early attempts at modeling creep were made by isolating the elastic and plastic deformation mechanisms. Creep, as mentioned previously, can be separated into two possible mechanisms, matrix and grain boundary creep. Knecht and Fox [45] proposed a matrix creep fatigue model relating the solder matrix creep shear strain range to the fatigue life:

$$N_f = \frac{C}{\Delta\gamma_{mc}}, \tag{12.10}$$

where the number of cycles to failure, N_f, is a function of the matrix creep strain, $\Delta\gamma_{mc}$, and C, is a constant.

Creep-Fatigue Interaction Model By applying Miner's linear superposition principal, both plastic and creep strain can be accounted for in strain-based fatigue model. Pang et al. [46] combined the Solomon's plastic shear strain range model with Knecht and Fox's matrix creep shear strain range model:

$$\frac{1}{N_f} = \frac{1}{N_p} + \frac{1}{N_c}, \tag{12.11}$$

where N_p is the number of cycles to failure due to plastic strain fatigue and is obtained directly from the Solomon's fatigue model. N_c is the number of cycles to failure due to the creep strain fatigue and is obtained from the Knecht and Fox's creep fatigue model. The model has been applied to ceramic ball grid array (BGA) and underfilled flip chip assemblies subjected to thermal cycling loading [46].

Total Shear Strain Range Fatigue Model The Engelmaier model [47] given in Equation (12.12), modified Solomon's model by incorporating the effects of mean solder temperature and frequency in the fatigue model. The total number of cycles to failure is related

to the total shear strain, $\Delta\gamma_t$, the fatigue ductility coefficient, ε'_f, and the variable exponent, c, which is a function of mean temperature and frequency:

$$N_f = \frac{1}{2}\left[\frac{\Delta\gamma_t}{2\varepsilon'_f}\right]^{1/c}, \tag{12.12}$$

where $c = -0.442 - 6 \times 10^{-4} T_s + 1.74 \times 10^{-2} \ln(1 + f)$. T_s is the mean cyclic solder joint temperature in °C, and f is the cyclic frequency in cycles/day.

Energy-Based Models Energy-based fatigue models are used to predict fatigue failure based on a hysteresis energy or volume-weighted average plastic work density [48–50],

$$N_f = \left(\frac{\Delta\overline{W}_{total}}{W_0}\right)^{1/k}, \tag{12.13}$$

where N_f is the mean cycles to failure, $\Delta\overline{W}_{total}$ is the total strain energy, W_0 and k are fatigue coefficients.

Darveaux et al. [37] used strain energy to predict crack initiation and crack growth, the equations are given below:

$$N_0 = K_1 \Delta W^{K_2}, \tag{12.14}$$

$$\frac{da}{dN} = K_3 \Delta W^{K_4}, \tag{12.15}$$

where ΔW is strain energy, K_1, K_2, K_3, and K_4 are material constants. When used with finite element analysis, the inelastic strain energy density, ΔW is normalized by the volume of the element:

$$\Delta W_{ave} = \frac{\sum \Delta W \cdot V}{\sum V}, \tag{12.16}$$

where ΔW_{ave} is the average inelastic strain energy density, V is the volume of each element.

The energy-based solder fatigue results for Sn-Pb solder reported Shi et al. [50] was used by Pang et al. [51,52].

12.3. DESIGN FOR RELIABILITY (DFR)

The Design-For-Reliability knowledge triangle [53] is shown in Figure 12.1 and involves traditionally experimental methods for reliability tests. With advances made in numerical modeling techniques particularly in finite element analysis (FEA) virtual simulation or upfront reliability modeling of solder joint reliability tests and service conditions are increasingly employed. Post-processing the FEA modeling and simulation results and applying fatigue life prediction models provide a failure prediction methodology within the DFR process.

The DFR methodology is illustrated in Figure 12.2 and involves; materials characterizations, fatigue life prediction models, FEA modeling and reliability tests.

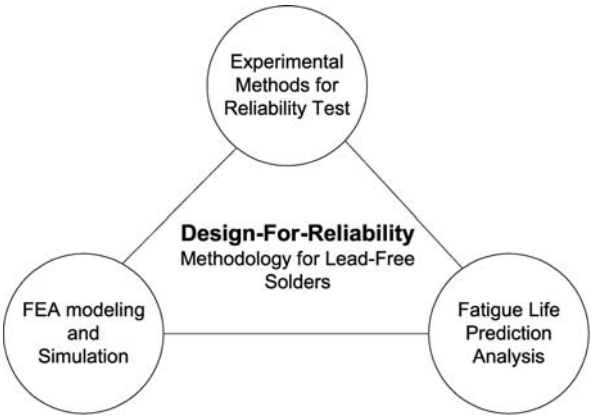


FIGURE 12.1. Design for reliability (DFR) knowledge triangle.

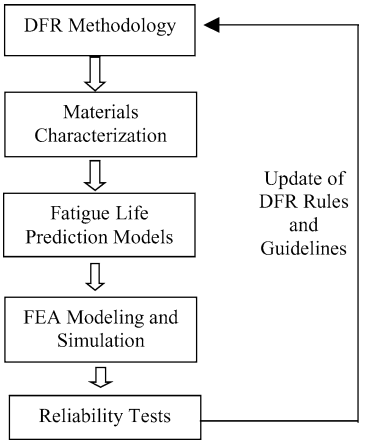


FIGURE 12.2. Design for reliability (DFR) methodology.

The DFR methodology applies to solder joint reliability performance evaluation for Pb-free and Pb-based electronic solders. The research scope for materials characterization and modeling, fatigue test and life prediction modeling, FEA modeling, and reliability testing.

Tensile Test Program The tensile test method was employed to measure the mechanical properties of 95.5Sn-3.8Ag-0.7Cu and 99.3Sn-0.7Cu solder alloys. Dog bone-shaped bulk solder specimens for the uniaxial tensile test were machined from solder bars. The specimen has a total length of 65 mm, a gage length of 15 mm, and a diameter of 3 mm (see Figure 12.3), which follows the ASTM standard [54,55]. The tensile tests were carried out on a universal testing machine, at four different temperatures (−40°C, 25°C, 75°C and 125°C). At each temperature, three different displacement rates were used (0.5 mm/min, 5 mm/min and 50 mm/min).

Creep Test Program The tensile test specimen for 95.5Sn-3.8Ag-0.7Cu and 99.3Sn-0.7Cu solder alloy was used also for the creep test. The creep tests were carried out on a universal

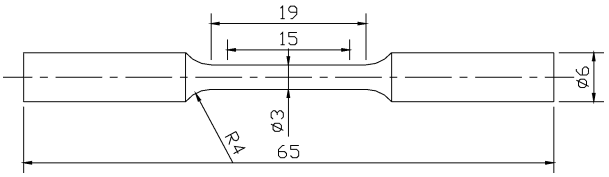


FIGURE 12.3. Geometry of tensile and creep specimen (in mm).

testing machine, which has the ability to hold the load constant for creep test [56]. Tests were carried out at four different temperatures (−40°C, 25°C, 75°C and 125°C) with constant stress set at eleven different values (3, 4, 5, 7, 10, 15, 20, 25, 30, 35, 40, 50, 60 and 70 MPa).

Fatigue Test Program The low cycle fatigue test specimens for 95.5Sn-3.8Ag-0.7Cu and 99.3Sn-0.7Cu solder alloy were machined from the solder bar. The configuration of the specimens has a diameter of 6 mm at the two ends, a center diameter of 3 mm, and a gauge length of 4.5 mm with a radius of curvature of 50 mm to reduce stress concentration effects due to sharp corners. The geometry of the specimen is following the ASTM standard [57]. The low cycle fatigue tests were conducted on a micro-force materials test system. The sine waveform was employed for all fatigue tests. The total strain was calculated from the crosshead displacement divided by the gauge length. The tests were carried out at three different frequencies (10^{-3}, 10^{-2} and 1 Hz) and at three different temperatures (25, 75 and 125°C) with total strain set at four different values (2, 3.5, 5 and 7.5%). The fatigue failure was defined as 50% reduction of maximum tensile load.

12.4. CONSTITUTIVE MODELS FOR LEAD FREE SOLDERS

In this section, documentation of the materials testing, characterization and modeling work on the Pb-free solder materials for 95.5Sn-3.8Ag-0.7Cu and 99.3Sn-0.7Cu is reported. The tensile, impact and creep properties are critical information needed for formulating Pb-free solder materials constitutive models which can be implemented in finite element analysis programs.

12.4.1. Tensile Test Results

Pb-free 95.5Sn-3.8Ag-0.7Cu Solder Alloy Results The effects of strain rate and temperature on the 95.5Sn-3.8Ag-0.7Cu tensile properties were evaluated and the stress–strain curves for the three different strain rates at a constant temperature of 125°C are shown in Figure 12.4. It can be seen that the mechanical properties of 95.5Sn-3.8Ag-0.7Cu is strongly dependent on the test temperature and strain rate parameters [10,11]. The result of the effect of temperature on apparent elastic modulus is shown in Figure 12.5. It can be seen that the curves demonstrate linear relationship between the apparent elastic modulus and the temperature and decrease with the increase of temperature. The plot shows approximately straight lines with a constant slope. The apparent elastic modulus has a linear function of logarithmic strain rate. The yield stress was computed from the stress–strain curve at 0.2 percent plastic strain. The effects of test temperature and strain rate on the yield stress were given in Figure 12.6. There is a linear relationship between the yield

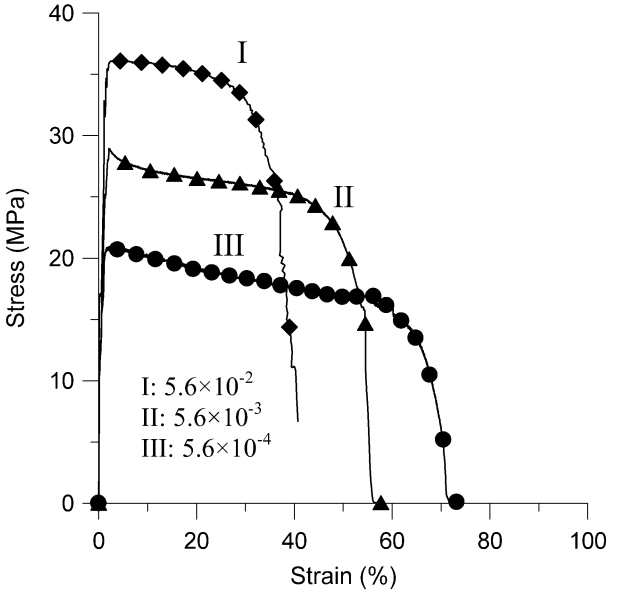


FIGURE 12.4. Tensile test results at 125°C.

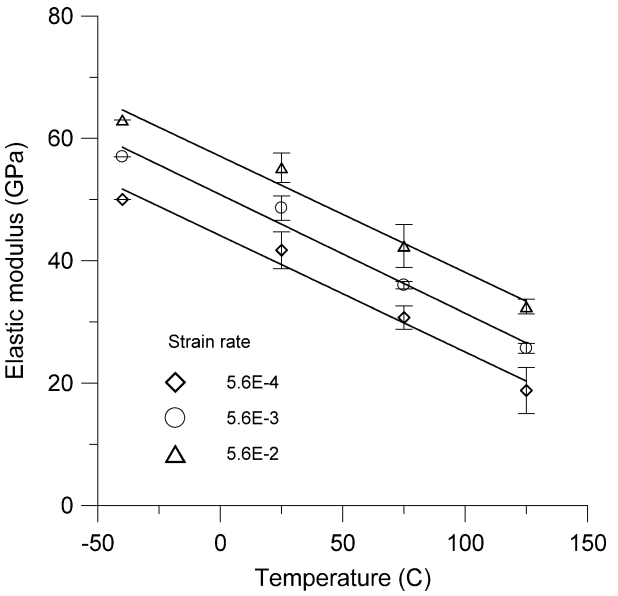


FIGURE 12.5. Effect of temperature on apparent elastic modulus.

stress versus temperature plot, and yield stress versus logarithmic strain rate. The yield stress increases with faster strain rate, and decreases with increase in temperature. The corresponding results for the ultimate tensile stress (UTS) are given in Figure 12.7. The UTS results follow has a similar trend as the yield stress results. UTS increases with increase in strain rate, and decreases with increase in temperature.

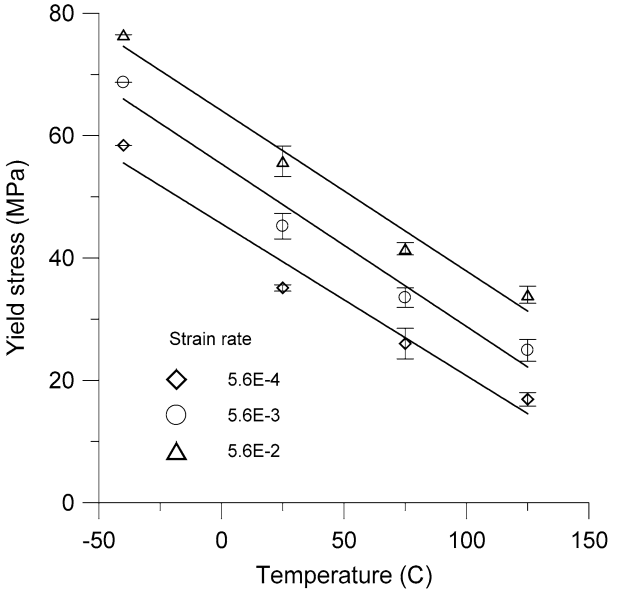


FIGURE 12.6. Effect of temperature on apparent yield stress.

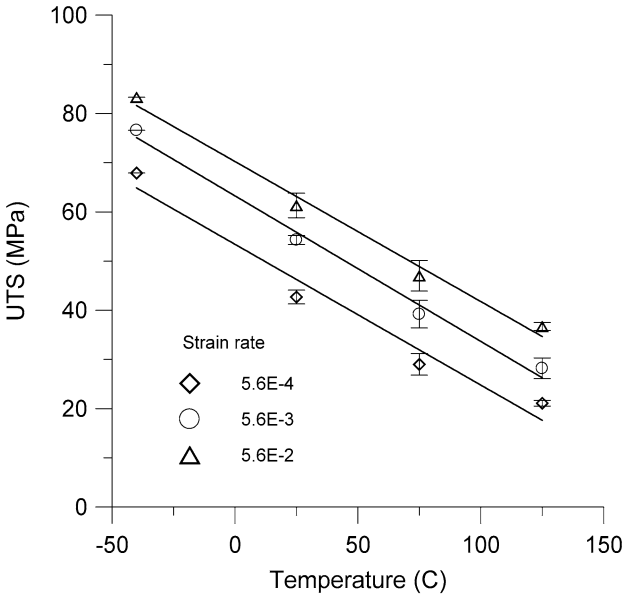


FIGURE 12.7. Effect of temperature on apparent UTS.

Comparison with 63Sn-37Pb Solder Data The comparison of apparent elastic modulus, yield stress and UTS for Sn-3.8Ag-0.7Cu, Sn-0.7Cu and 63Sn-37Pb at room temperature and high temperature are shown in Figures 12.8–12.10, respectively. It is noted that Sn-3.8Ag-0.7Cu has the highest mechanical properties among these three solders. While Sn-0.7Cu solder has slightly higher mechanical properties than Sn-37Pb solder. The differ-

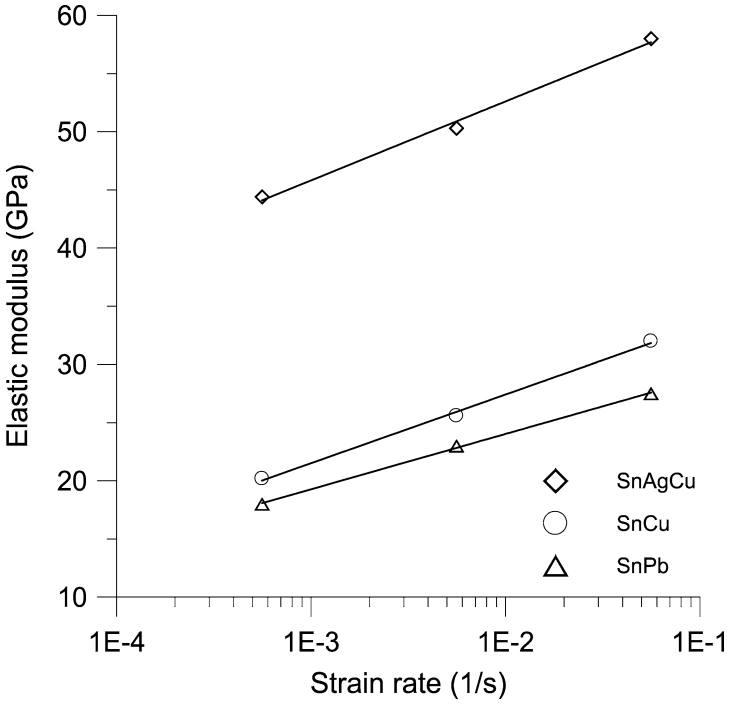


FIGURE 12.8. Comparison of apparent elastic modulus at 25°C.

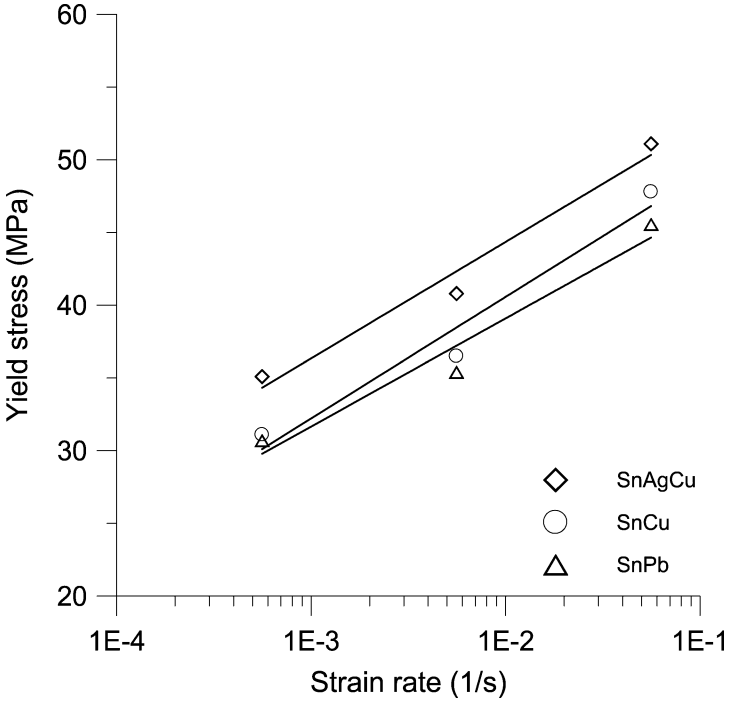


FIGURE 12.9. Comparison of apparent yield stress at 25°C.

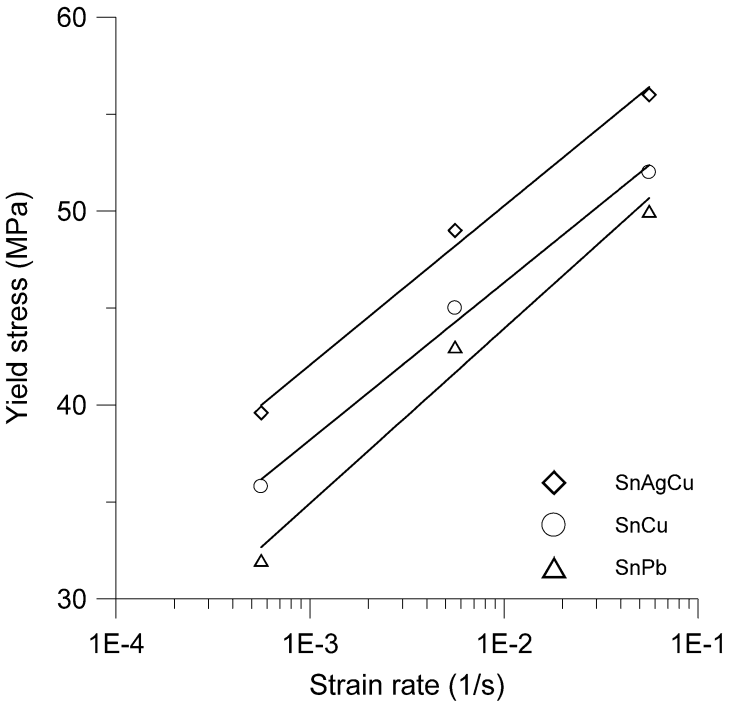


FIGURE 12.10. Comparison of apparent UTS at 25°C.

ences in apparent elastic modulus are significant for the three solders. At room temperature, the Sn-3.8Ag-0.7Cu has a modulus of 50 GPa, compared to 30 GPa for Sn-0,7Cu and Sn-37Pb solders. However, the difference in the apparent yield stress and UTS is not as significant, the differences for the three solders are within 15 MPa for all test conditions.

Temperature and Strain Rate Effects on Tensile Properties It is noted that the tensile properties of Sn-3.8Ag-0.7Cu and Sn-0.7Cu are dependent on temperature and strain rate. Linear regression was employed to quantify the temperature and strain rate dependent mechanical properties [10,11] of 95.5Sn-3.8Ag-0.7Cu and 99.3Sn-0.7Cu Pb-free solders.

The apparent elastic modulus of 95.5Sn-3.8Ag-0.7Cu is:

$$E(T,\dot{\varepsilon})_{SnAgCu} = (0.00074T + 6.44)\log(\dot{\varepsilon}) + (-0.1932T + 65.935). \tag{12.17}$$

The apparent yield stress and UTS were curve-fitted to:

$$\sigma_y(T,\dot{\varepsilon})_{SnAgCu} = (-0.2053T + 76.61)(\dot{\varepsilon})^{(6.54\times10^{-4}T+0.074)}, \tag{12.18}$$

$$\mathrm{UTS}(T,\dot{\varepsilon})_{SnAgCu} = (-0.2161T + 81.32)(\dot{\varepsilon})^{(6.27\times10^{-4}T+0.0619)}. \tag{12.19}$$

The closed-form equations of apparent elastic modulus, yield stress and UTS are within 5% of the test results.

12.4.2. Creep Test Results

Steady State Creep Strain Rate for Pb-free Solders By plotting all steady strain rates against the applied normal stress of Sn-3.8Ag-0.7Cu and Sn-0.7Cu solder alloys, creep test results were clearly shown in Figures 12.11 and 12.12. When the temperature and applied normal stress increase, the steady strain rate will increase. The creep properties of Sn-0.7Cu solder is lower than those of the Sn-3.8Ag-0.7Cu solder alloy [14,17].

Comparison with Tin-Lead Solder Creep Data The comparison of steady state creep strain rate for Sn-3.8Ag-0.7Cu, Sn-0.7Cu and 63Sn-37Pb [56] at room temperature and high temperature were shown in Figure 12.13. It can be noticed that lead free solder has the better creep resistance than tin-lead solder, while the creep property of Sn-3.8Ag-0.7Cu is better than Sn-0.7Cu solder.

Creep Models for Steady State Creep Strain Rate Creep of a material is often characterized by its steady state creep strain rate, $\dot{\varepsilon}$, which can be simply expressed as a power law relationship,

$$\dot{\varepsilon} = A\sigma^n e^{\frac{-Q}{kT}}, \tag{12.20}$$

where A is material constant, σ is applied stress, n is stress exponent, Q is creep activation energy, k is Boltzmann's constant, T is absolute temperature. The activation energy, Q, and the stress exponent, n, changes with the dominant creep mechanism, and may have different values at different regimes of the applied stress. To describe the steady state creep

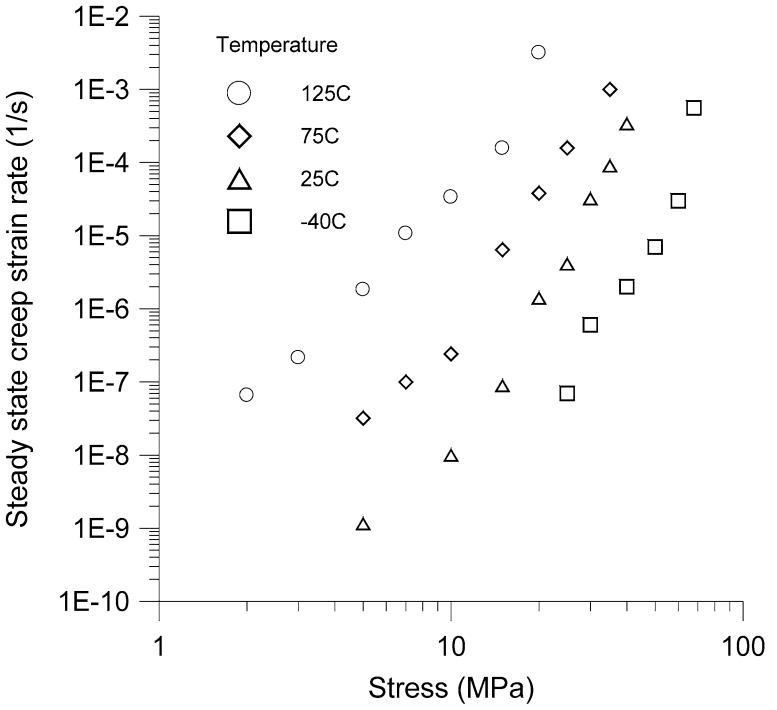


FIGURE 12.11. Steady state creep behavior of Sn-Ag-Cu.

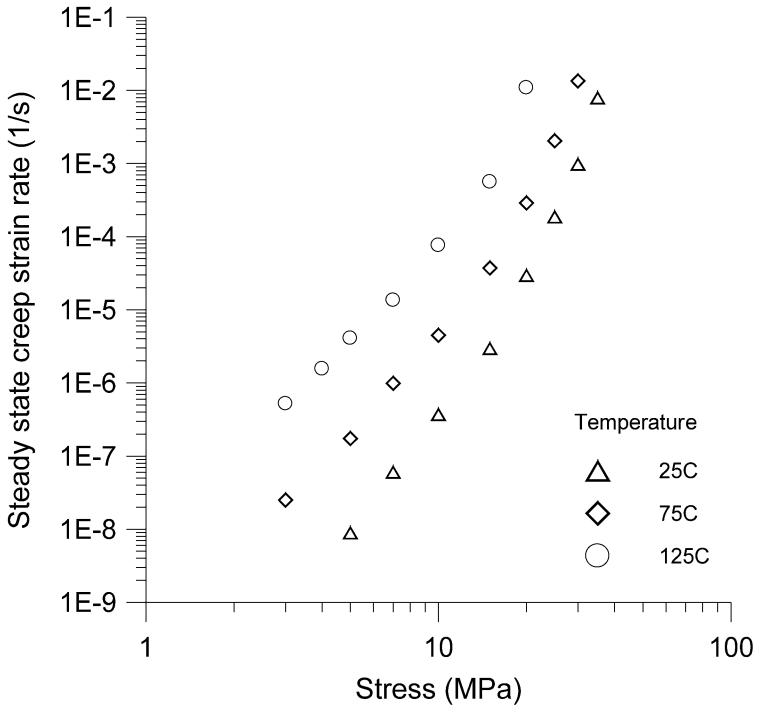


FIGURE 12.12. Steady state creep behavior of Sn-Cu.

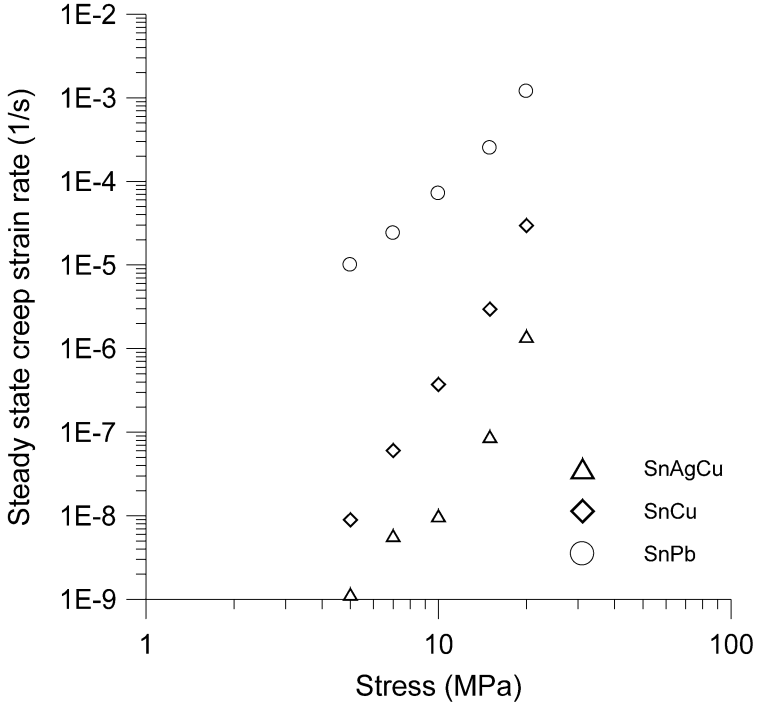


FIGURE 12.13. Comparison of creep data at 25°C.

strain rate at different regimes of stress and temperature, hyperbolic-sine model is widely used. The expressions of the models is,

$$\dot{\varepsilon} = C[\sinh(\alpha\sigma)]^n \exp\left(-\frac{Q}{kT}\right). \tag{12.21}$$

The results of steady state creep strain rate dependent on normal stress and temperature for 95.5Sn-3.8Ag-0.7Cu and 99.3Sn-0.7Cu lead free solders were use to curve fit to the model described in Equation (12.21). The parameters for Sn-3.8Ag-0.7Cu solder can be determined as:

$$\dot{\varepsilon}_{SnAgCu} = 501.3[\sinh(0.0316\sigma)]^{4.96} \exp\left(-\frac{5433.5}{T}\right). \tag{12.22}$$

TABLE 12.1.
Comparison of creep model constants for Sn-Ag-Cu.

	Material	Specimen	Stress exponent n	Activation energy Q (kJ/mol)
Pang [14]	95.5Sn3.8Ag0.7Cu	Tension bulk specimen with gauge diameter of 3 mm and gauge length of 15 mm	4.96	45.2
Lau et al. [21]	95.5Sn3.9Ag0.6Cu	Compression specimen with diameter of 10 mm and length of 19 mm	4.2	45
Schubert [30]	95.5Sn3.8Ag0.7Cu	Flat bulk specimen with 3 × 3 mm cross section and gauge length of 30 mm	6.4	54

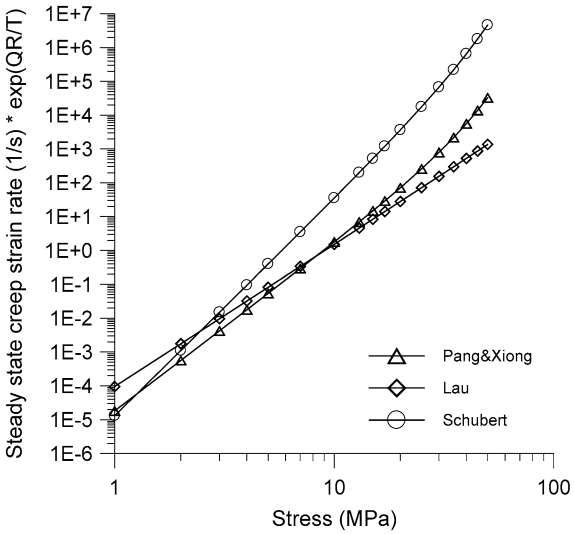


FIGURE 12.14. Comparison of creep behavior for Sn-Ag-Cu solder.

Lau et al. [21] and Schubert [30] also reported their creep tests data for Sn-Ag-Cu solder, and their curve-fitted results for the hyperbolic-sine creep model:

$$\text{Lau}_{(95.5\text{Sn-}3.9\text{Ag-}0.6\text{Cu})} \quad \dot{\varepsilon} = 4.41 \times 10^5 [\sinh(0.005\sigma)]^{4.2} \exp(-5412/T), \tag{12.23}$$

$$\text{Schubert}_{(95.5\text{Sn-}3.8\text{Ag-}0.7\text{Cu})} \quad \dot{\varepsilon} = 2.78 \times 10^5 [\sinh(0.02447\sigma)]^{6.41} \exp(-6496/T). \tag{12.24}$$

A comparison of the Sn-Ag-Cu solder material, specimen type, stress exponent and activation energy from Lau et al. [21], Schubert [30], and the Pang's [14] creep test result is given in Table 12.1. It is noted that the stress exponent n varies from 4.2 to 6.4, while the activation energy, Q, varies from 45 kJ/mol–54 kJ/mol. A comparison of the steady state creep model using the Equations (12.22), (12.23) and (12.24) is shown in Figure 12.14. From the figure, it can be seen that the author's creep data agree well with data reported by Lau et al. [21]. Satisfactory agreement with Schubert's data [30] was seen at lower stresses, but significant difference is noted at higher stresses above 10 MPa.

12.5. LOW CYCLE FATIGUE MODELS

Cyclic Stress–Strain Behavior The cyclic stress–strain hysteresis loop for 99.3Sn-0.7Cu and 95.5Sn-3.8Ag-0.7Cu solders at the tenth cycle (2% total strain range, at temperature of 125°C and frequency of 1 Hz) are shown in Figure 12.15. For the same total strain range, 99.3Sn-0.7Cu solder has a larger plastic strain range and much smaller stress range than Sn-3.8Ag-0.7Cu solder.

Effect of Temperature on Low Cycle Fatigue Behavior Test result of Sn-0.7Cu solder at different temperatures will show the effect of temperature on low cycle fatigue. By plotting the hysteresis loop of different temperature at strain range of 3.5% and frequency of 1 Hz, the material response at different temperature can be evaluated and compared (Figure 12.16). While the Sn-3.8Ag-0.7Cu solder has the similar trends with Sn-0.7Cu solder. The relationship between the fatigue life and plastic strain obtained from tests of 0.01 Hz at three different temperatures (25, 75 and 125°C) is shown in Figure 12.17.

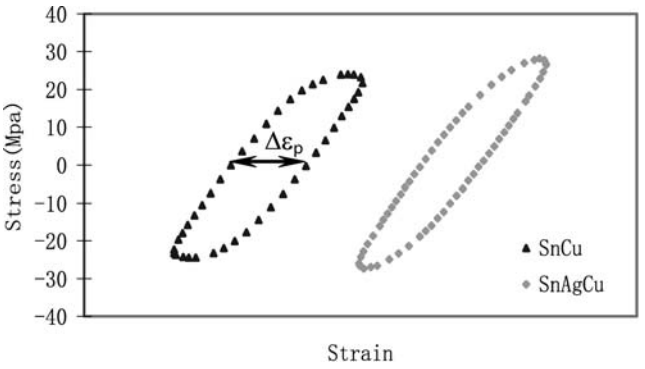


FIGURE 12.15. Cyclic stress–strain hysteresis loops.

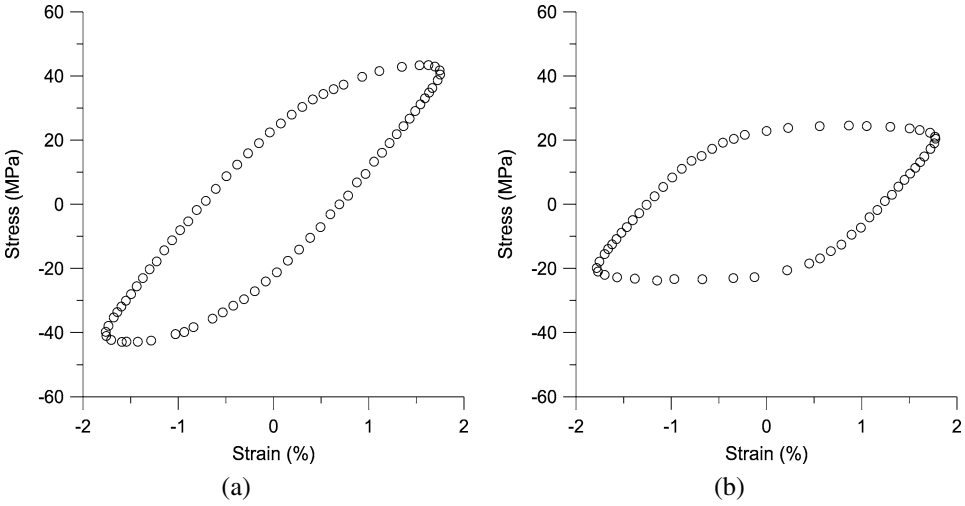


FIGURE 12.16. Hysteresis loop at 1 Hz and strain range of 3.5%. (a) 25°C, (b) 125°C.

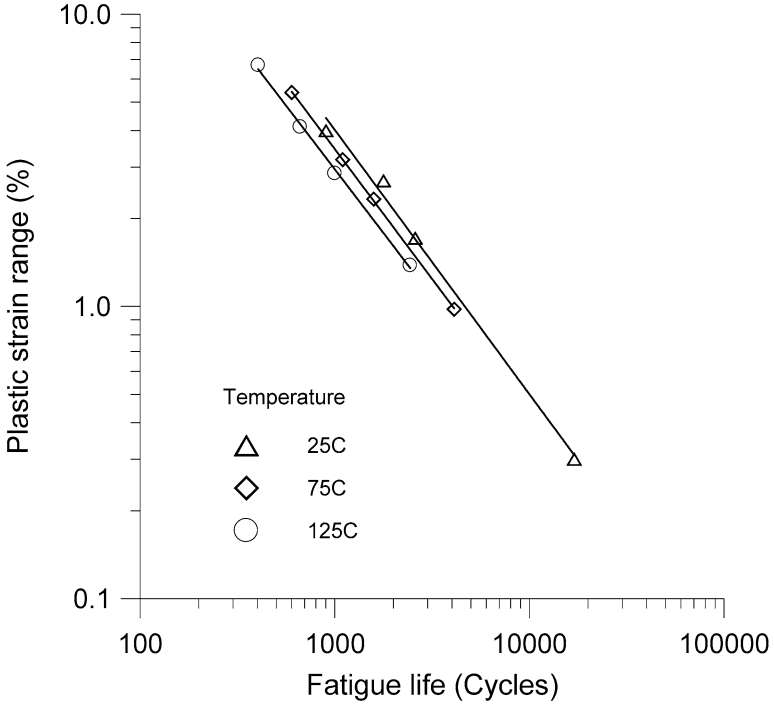


FIGURE 12.17. Plastic strain versus fatigue life at 0.01 Hz for different temperatures.

Effect of Frequency on Low Cycle Fatigue Behavior The fatigue test result of Sn-0.7Cu solder show the frequency on low cycle fatigue [29]. By plotting the hysteresis loop of different frequency at 125°C and strain range of 7.5%, the material response at different temperatures can be evaluated and compared in Figure 12.18. Sn-3.8Ag-0.7Cu solder has

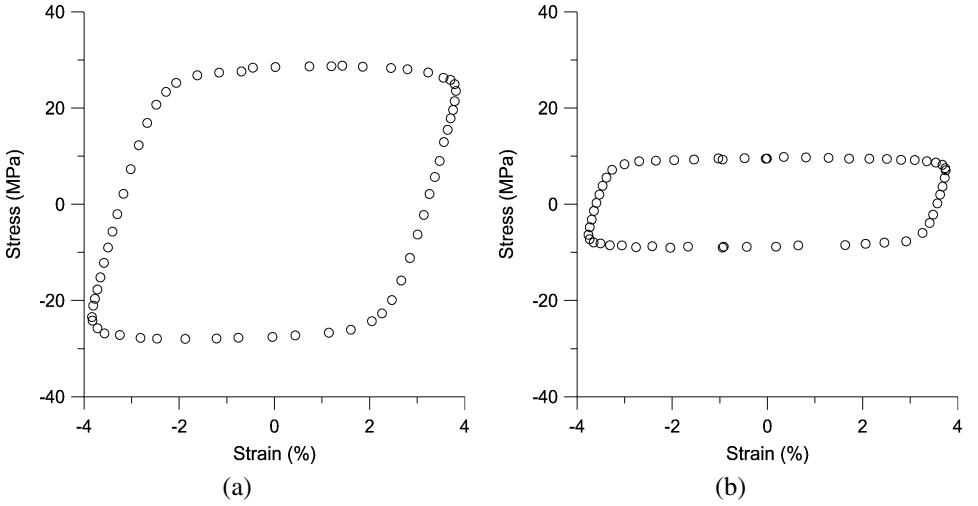


FIGURE 12.18. Hysteresis loop at 125°C and strain range of 7.5%. (a) 1 Hz, (b) 0.001 Hz.

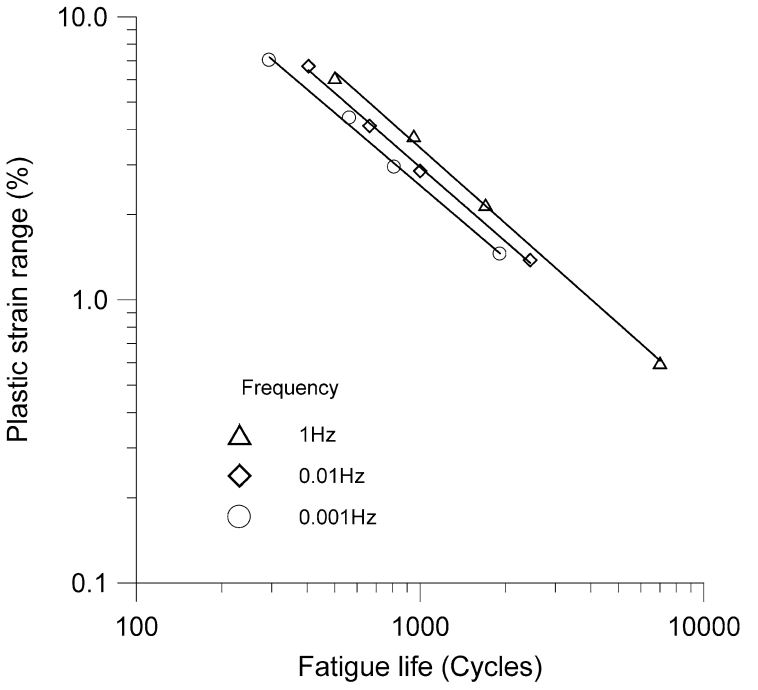


FIGURE 12.19. Plastic strain versus fatigue life at 125°C for different frequencies.

the similar trend. The fatigue tests results at the same temperature and total strain, but at different frequencies, show that the fatigue life of Pb-free solders decrease with slower frequency as shown in Figure 12.19. This can be attributed to the increasing effect of creep-fatigue damage as test frequency reduces. When frequency decreases, the time for creep

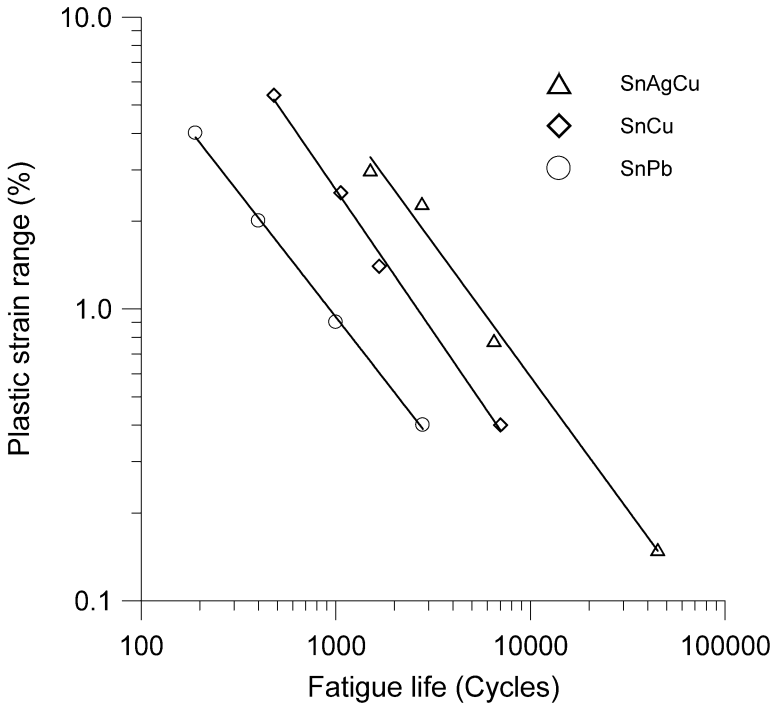


FIGURE 12.20. Comparison of fatigue resistance at room temperature and 1 Hz.

exposure is longer as the cycle time increases but this lead to short fatigue life in terms of cycles to failure.

Comparison with 63Sn-37Pb Solder Fatigue Data Shi et al. [43] reported the fatigue properties for 63Sn-37Pb solder at different temperatures and frequencies. The comparison of fatigue resistance for Sn-3.8Ag-0.7Cu, Sn-0.7Cu and 63Sn-37Pb at room temperature and 1 Hz frequency as shown in Figure 12.20. It can be seen that Pb-free solders have better fatigue resistance than Sn-Pb solder. The fatigue performance of Sn-Ag-Cu solder is much better than Sn-Cu solder.

The Coffin-Manson model has been widely used to relate low cycle fatigue life (N_f) of metallic materials with the plastic strain range ($\Delta\varepsilon_p$), as shown below:

$$N_f^m \Delta\varepsilon_p = C, \tag{12.25}$$

where m and C are material constants.

The Morrow's energy-based model was also use to predict low cycle fatigue lives. The model predicts fatigue life in terms of inelastic strain energy density (W_p), given below:

$$N_f^n W_p = A, \tag{12.26}$$

where n is fatigue exponent, and A is material ductility coefficient. The inelastic energy density was determined from the hysteresis loop. Based on the low cycle fatigue test result

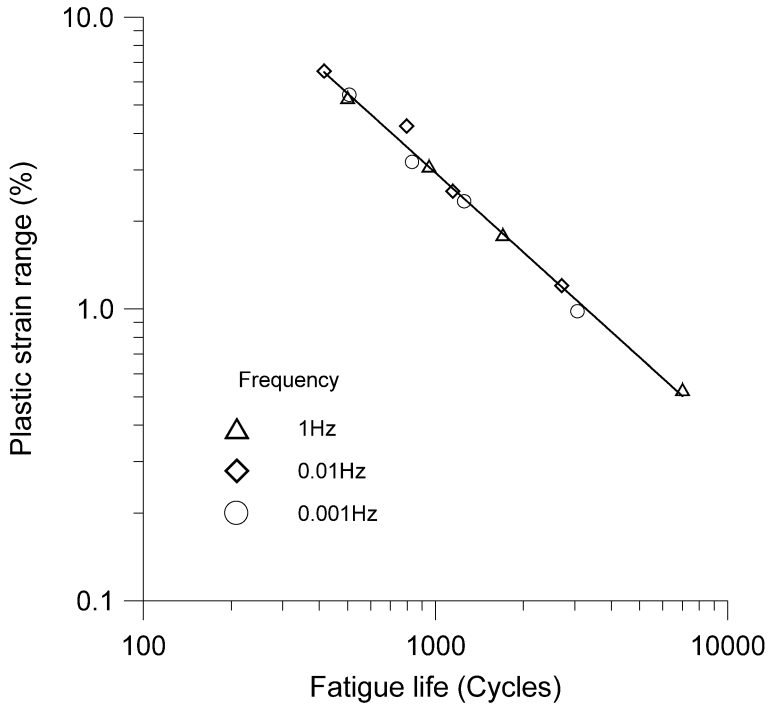


FIGURE 12.21. Plastic strain versus frequency-modified fatigue life at 125°C for Sn-Ag-Cu ($k = 0.95$).

at different temperatures and frequencies, the m, C, n and A at different conditions can be determined for 95.5Sn-3.8Ag-0.7Cu and 99.3Sn-0.7Cu lead free solders [14,16].

To model frequency dependent low cycle fatigue behavior of Sn-3.8Ag-0.7Cu and Sn-0.7Cu lead free solders, the frequency-modified Coffin-Manson relationship can be introduced to describe the low cycle fatigue behavior of Sn-0.7Cu solder, given below.

$$[N_f \nu^{(k-1)}]^m \Delta\varepsilon_p = C, \tag{12.27}$$

where ν is frequency, k, is frequency exponent.

Based on the fatigue test results at different frequencies at 125°C, the frequency exponent, k, can be determined to be 0.95 and 0.91 for Sn-3.8Ag-0.7Cu and Sn-0.7Cu solder, respectively. Then the frequency modified fatigue life $N_f \nu^{(k-1)}$ can be calculated [14,16]. When the plastic strain is plotted against the frequency-modified fatigue life, all the fatigue life data obtained at different frequencies were found to fit well into a single line, as shown in Figure 12.21. This means that the frequency-modified Coffin-Manson model can be used to model the effect of frequency for the low cycle fatigue behavior of Sn-3.8Ag-0.7Cu solder.

Similarly, the frequency-modified Morrow model can be used to model the frequency effect, as given below:

$$[N_f \nu^{(h-1)}]^n W_p = A, \tag{12.28}$$

where ν is the frequency, h is frequency exponent. Based on fatigue test of Sn-3.8Ag-0.7Cu and Sn-0.7Cu solder at different frequencies, the frequency exponent, h, can be

determined to be 0.82 and 0.762, respectively. Figure 12.22 show the inelastic strain energy density plotted against the frequency-modified fatigue life at 125°C. It can be seen that all fatigue life data at different frequencies falls on a single line. Hence, the frequency-modified Morrow model can be used to model the effect of frequency for the low cycle fatigue behavior of Pb-free Sn-Ag-Cu solder.

12.6. FEA MODELING AND SIMULATION

Finite Element Analysis (FEA) modeling for solder joint reliability analysis subject to thermal cycling loading is an important contribution to the DFR methodology. Studies employ 3D Strip, 3D Octant, and 2D Plane Strain or 2D Plane Stress models have been reported [35]. FEA models for analyzing flip chip on board assembly are illustrated in Figures 12.23 to 12.25. The FEA models investigated are the 2D-Plane Strain, 2D-Plane Stress, 3D-1/8th symmetry and 3D-Strip models. The different stress and strain responses generated are shown in Figure 12.26. Studies on the modeling parameters such as the loading profile, temperature range, stress free state, linear versus viscoelastic effects in polymer materials have been reported [52]. The results show that the 2D-Plane Strain and 2D-Plane Stress models gave the highest and lowest solder joint strains respectively. The 3D-Strip and 3D-1/8th symmetry model result, fall in between the 2D-Plane Strain and 2D-Plane Stress model results. The 3D-1/8th symmetry

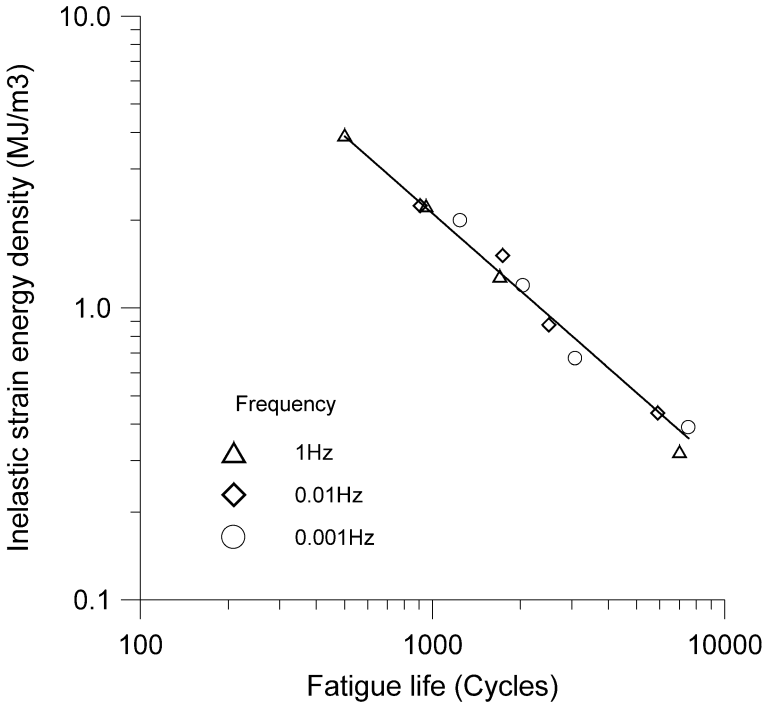


FIGURE 12.22. Inelastic strain energy density versus frequency-modified fatigue life at 125°C for Sn-Ag-Cu ($h = 0.82$).

model agrees better with the 2D-Plane Strain model, while the 3D-Strip model agrees better with the 2D-Plane Stress model results. Finite element analysis modeling may be used to study different thermal cycling (TC) and thermal shock (TS) loading analysis [52].

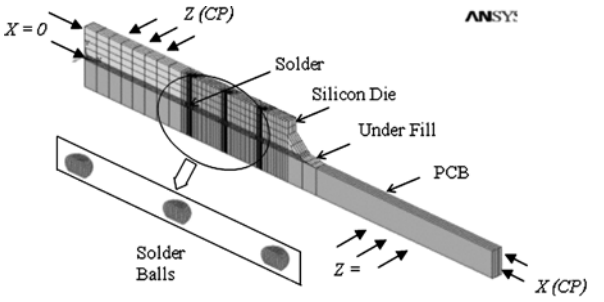


FIGURE 12.23. 3D-Strip model.

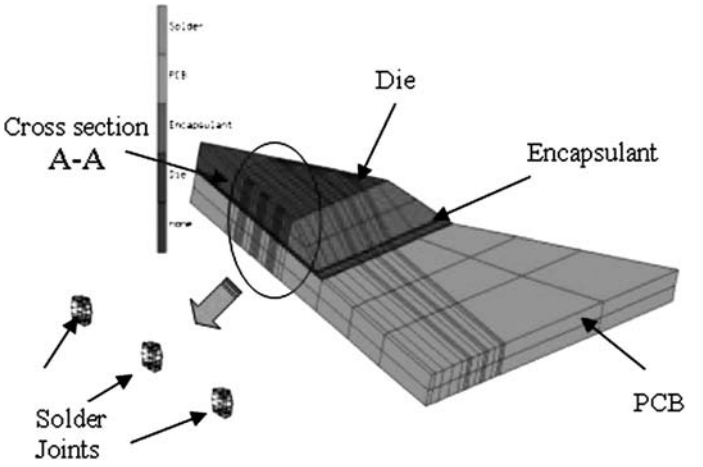


FIGURE 12.24. 3D-1/8th model.

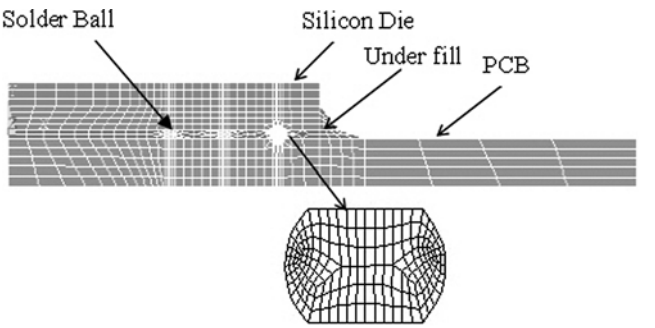


FIGURE 12.25. 2D Plane Strain or Plane Stress Model.

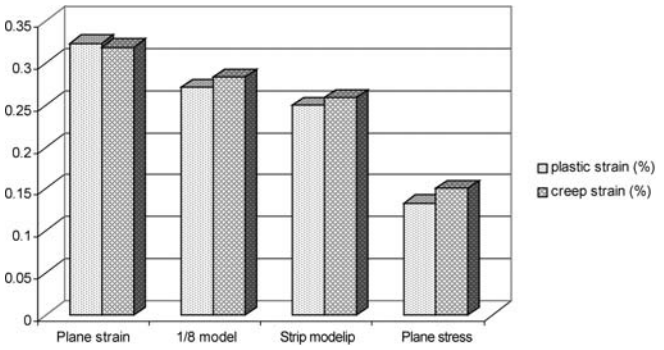


FIGURE 12.26. Comparison of 2D and 3D FEA model results.

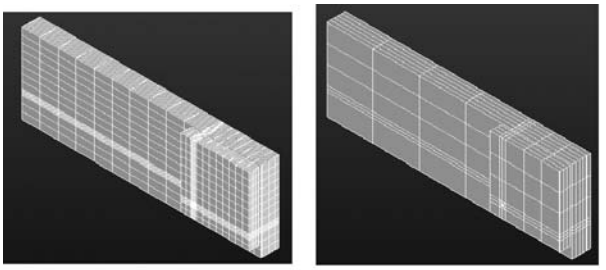


FIGURE 12.27. Global FE model for FCOB assembly.

Global-local modeling techniques [5,19] are needed for board-level finite element modeling and where full 3D geometric details of the PCB and several electronic packaging assemblies need to be modeled. An example to illustrate the submodeling technique and appropriate cut boundary, is shown in Figures 12.27 to 12.29 for a flip chip on board assembly. The silicon die is connected to the PCB with and without underfill encapsulation. The size of the die is 8.5 mm $\times$ 8.5 mm, with a thickness of 0.65 mm. The solder joint can be modeled as Sn-Pb or Pb-free solder material and has a standoff height of 0.1 mm.

Submodeling is known as a cut-boundary displacement method or the specified boundary displacement method. The cut boundary is the boundary of the submodel, which represents a cut through the coarse global model. Displacements calculated on the cut boundary of the coarse global model are transferred as boundary conditions to the fine submodel or local model. Submodeling technique can reduce, or even eliminate, the need for complicated transition regions in solid finite element models.

Due to the symmetry, only one quarter of the assembly was modeled. The 3-D quarter meshes for fine and coarse models are shown in Figure 12.27. Two types of cut boundary conditions may be employed as shown in Figure 12.28 for selected volume with silicon-to-solder-to-board materials, or as shown in Figure 12.29 for the selected solder joint volume only.

The cut boundary in Figure 12.28, include the IC chip and PCB volume subjected to thermal cycling loading. The thermal cycling loading profile (-40°C to $+125^{\circ}$C) used is

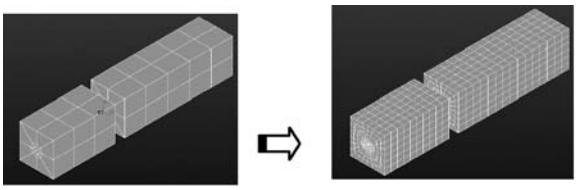


FIGURE 12.28. Cut boundary for coarse and fine submodel.

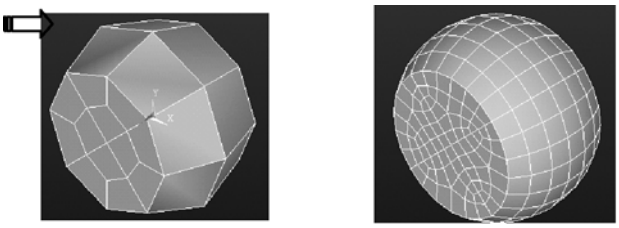


FIGURE 12.29. Cut boundary for coarse and fine solder submodel.

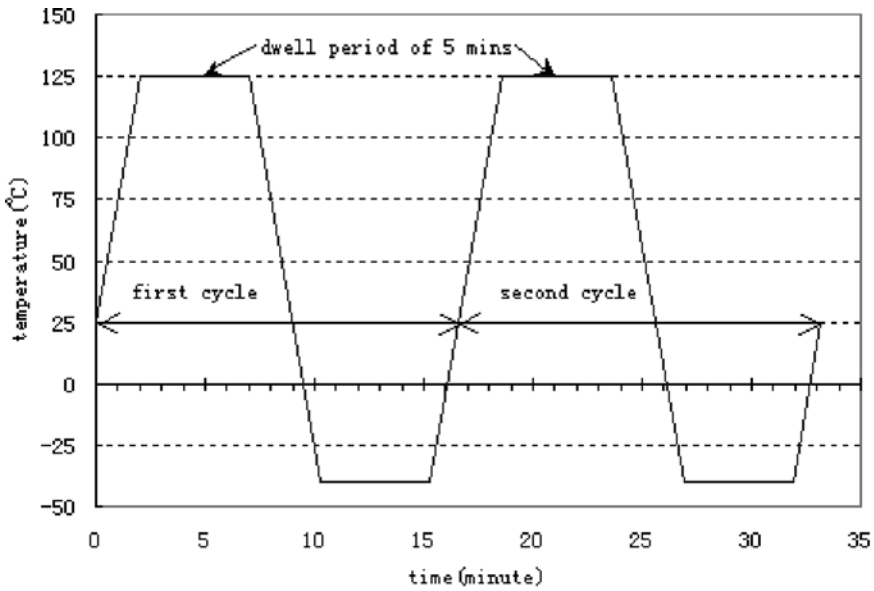


FIGURE 12.30. Thermal cycling loading profile (−40 to +125°C).

shown in Figure 12.30. Two cases were considered, the FCOB without underfill case and FCOB with underfill case.

For solder joint, Anand's viscoplastic model was used in the FEA model. The plastic work density is the important parameter to use for fatigue life prediction of solder joint. The plastic work density is shown in Figure 12.31 for the non underfill case, and Figure 12.32 for the underfilled FCOB case.

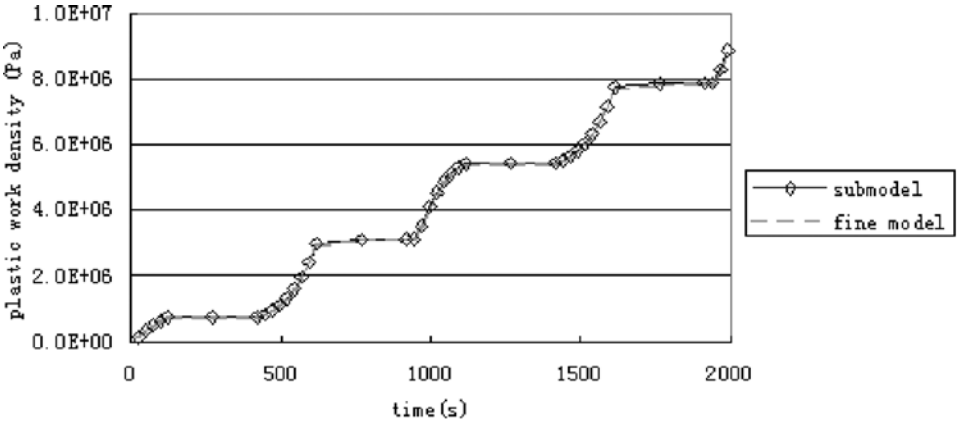


FIGURE 12.31. Plastic work density for non-underfill FCOB.

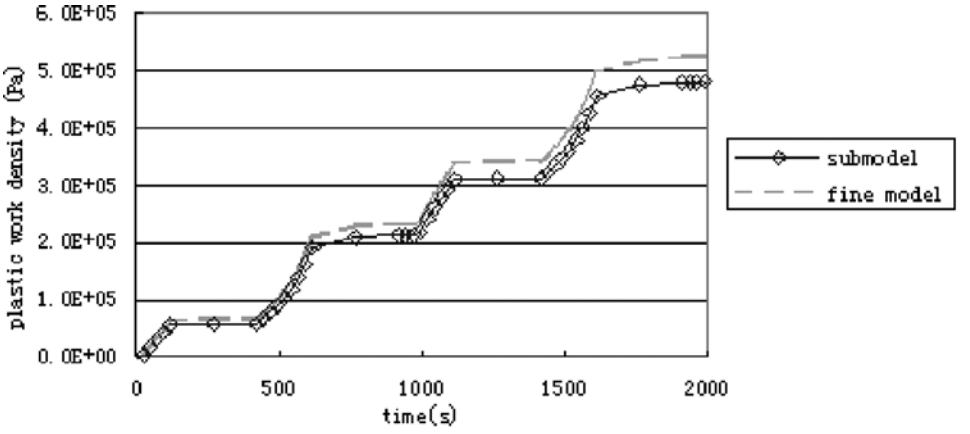


FIGURE 12.32. Plastic work density for underfill FCOB.

The submodeling technique can be used in FEA modeling for reliability analysis. The advantage of the submodeling technique include reduction in computational time, hard disk space, smaller virtual memory requirements with less elements used.

Comparison between lead-based and lead-free solders using viscoplastic Anand model properties for different solders were reported earlier [5]. Lead-based solders selected include high lead solders such as 92.5Pb-5Sn-2.5Ag, 97.5Pb-2.5Sn, near eutectic Sn-Pb solders such as 60Sn-40Pb and 62Sn-36Pb-2Ag, lead free solders for 96.5Sn-3.5Ag and the author's proprietary viscoplastic data for Sn-3.8Ag-0.7Cu and Sn-0.7Cu solder. The nine constants for Anand model of different solder alloys are shown in Table 12.2. A technique was used to calculate the averaged plastic work density accumulated per cycle for the top interface elements.

$$\Delta W_{ave} = \frac{\sum \Delta W \cdot V}{\sum \Delta V},$$

TABLE 12.2.
Anand model constants for different solder alloys.

Parameters	62Sn36Pb2Ag	92.5Pb5Sn2.5Ag	60Sn40Pb	96.5Sn3.5Ag	97.5Pb2.5Sn	SAC*	Sn-Cu*
S0 (MPa)	12.41	33.07	56.33	39.09	15.09	S01	S02
Q/k (K)	9400	11010	10830	8900	15583	k1	k2
A	4.00E6	1.05E5	1.49E7	2.23E4	3.25E12	A1	A2
ξ	1.5	7	11	6	7	ξ1	ξ2
m	0.303	0.241	0.303	0.182	0.143	m1	m2
ho (MPa)	1379	1432	2640.75	3321.15	1787.02	ho1	ho2
$s\hat{}$ (MPa)	13.79	41.63	80.42	73.81	72.73	s^1	s^2
n	0.07	0.002	0.0231	0.018	0.00437	n1	n2
a	1.3	1.3	1.34	1.82	3.73	a1	a2

*Proprietary data for Sn-3.8Ag-0.7Cu and Sn-0.7Cu solders by author.

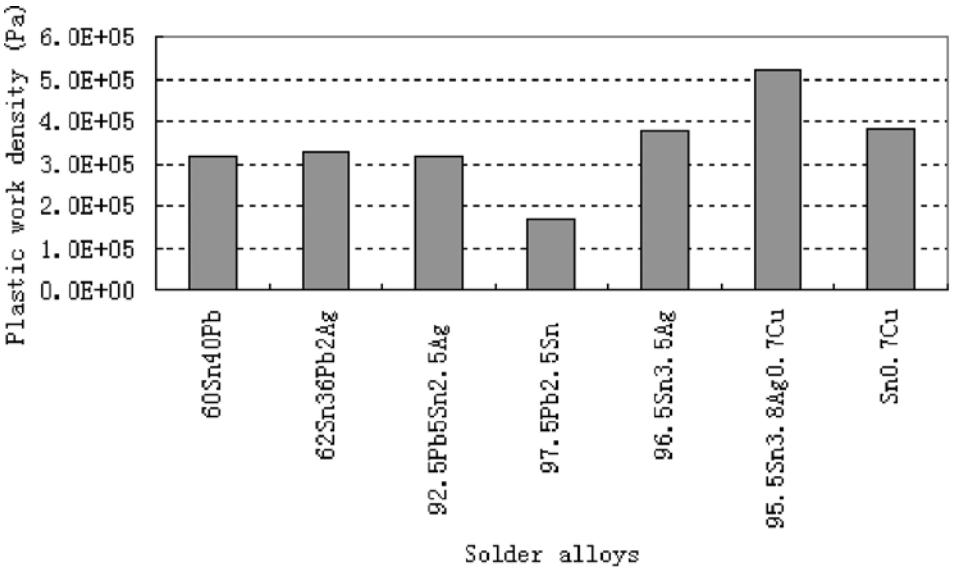


FIGURE 12.33. Plastic work density for different solder alloys.

where ΔW is the viscoplastic work density per cycle of each element, and V is the volume of each element

Plastic work density, was computed and compared for different solder material models shown in Figure 12.33.

From Figure 12.33, it is noted that the plastic work density accumulated per cycle is similar for near eutectic solders for 60Sn-40Pb and 62Sn-36Pb-2Ag. The high Pb solder, 97.5Pb-2.5Sn, has the lowest plastic work density. Pb-free solders have much higher plastic work density due to their higher strength. Using the plastic work density calculated for Sn-3.8Ag-0.7Cu and Sn-0.7Cu solder, in the energy-based fatigue life models developed, the fatigue life was calculated for Sn-3.8Ag-0.7Cu and Sn-0.7Cu and given in Figure 12.34. The fatigue life for Sn-3.8Ag-0.7Cu is 1256 cycles and for Sn-0.7Cu is 543 cycles. The effect of no underfill and with underfill for Sn-3.8Ag-0.7Cu solder joints is more than one order in magnitude.

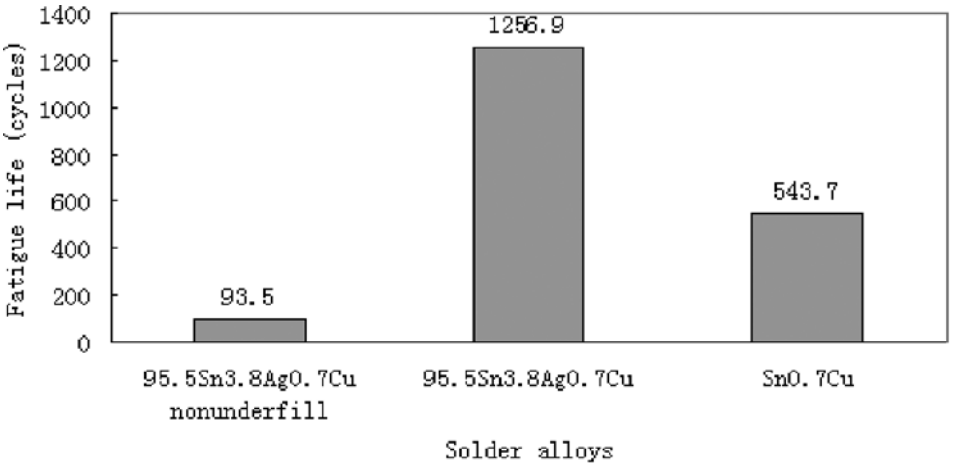


FIGURE 12.34. Fatigue life for Sn-0.7Cu and 95.5Sn-3.8Ag-0.7Cu.

12.7. RELIABILITY TEST AND ANALYSIS

Quality and reliability are two important product attributes that contribute to the success of electronic products. A design-built-test approach in quality and reliability testing of soldered IC and PCB assemblies is widely used by industry in the electronic packaging product development cycle. The electronic packaging development trends are pushing further in miniaturization, lower cost, improved reliability and shorter time-to-market. In the implementation of new Pb-free solders in the electronic packaging design-built-test cycle, the lack of user knowledge-base in Pb-free solder material and properties, mechanics and failure characterizations, reliability testing, and computer modeling and simulation of solder joint reliability is a major concern.

Electronic products may be subjected to different types of loading, such as thermal cycling (TC) or thermal shock (TS), vibration and drop impact loading. Failure mechanisms for electronic solders include creep, low-cycle fatigue, high-cycle fatigue and brittle fracture. In SMT assemblies, the solder joint in the electronic assembly is often the weakest link and solder joint reliability becomes even more important with further minimization. The solder joint is particularly prone to fatigue failure due to temperature cycling and mechanical loading.

In real service condition, the failure of solder joints in electronic assemblies is expected to precipitate after the warranty period for electronic products. It is not practical to obtain reliability test data for service condition, as it would take too long to test and collect failure data. Reliability testing often employs accelerated stress testing (AST). Reliability tests are aimed at revealing and understanding the physics of failure. Another objective of accelerated reliability tests is to accumulate representative failure statistics. Accelerated tests use elevated stress levels and/or higher stress-cycle frequency to precipitate failures over a much shorter time. Typical accelerating stresses are temperature, mechanical load, thermal cycling, and vibration. Such tests can facilitate physics of failure reliability tests that are cost effective, shorten the product process and improve long-term product reliability. Highly Accelerated Life Tests (HALT) is carried out to obtain, as soon as possible, the preliminary information about the reliability of products, and the principal physics of their failures.

Life distribution model is a statistic model that is a theoretical population model used to describe the lifetime of a component/product/system and is defined as the cumulative distribution function $F(t)$ for the population. The life distribution model $F(t)$ has two useful interpretations:

$F(t)$ is the probability a random unit drawn from the population fails by t hours (or cycles).

$F(t)$ is the fraction of all units in the population which fails by t hours (or cycles).

It is defined as:

$$F(t) = F(s \leq t) = \int_{-\infty}^{t} f(s)ds, \tag{12.29}$$

where $f(t)$ is probability density function.

The Weibull life distribution model is widely used for solder joint reliability investigations. For two-parameter Weibull distribution, the failure density function can be written as

$$f(t) = \frac{\beta}{\eta}\left(\frac{t}{\eta}\right)^{\beta-1} \exp\left[-\left(\frac{t}{\eta}\right)^{\beta}\right], \tag{12.30}$$

where β is the a shape factor, and measures how the failure frequency is distributed around the average lifetime. η is called the lifetime parameter, because it gives the time at which 63.2% of the devices failed. So the failure function can be obtained.

$$F(t) = \int_{0}^{t} f(\tau)d\tau = 1 - \exp\left[-\left(\frac{t}{\eta}\right)^{\beta}\right]. \tag{12.31}$$

The reliability function corresponding to this failure function can be obtained

$$R(t) = 1 - F(t) = \exp\left[-\left(\frac{t}{\eta}\right)^{\beta}\right], \tag{12.32}$$

and the failure rate can be calculated.

$$h(t) = \frac{f(t)}{R(t)}. \tag{12.33}$$

When the life distribution model is available, the mean time to failure can be calculated by the following equation.

$$\text{MTTF} = \int_{0}^{\infty} tf(t)dt = \eta \cdot \Gamma\left(\frac{1}{\beta} + 1\right). \tag{12.34}$$

The two-parameter Weibull probability plot is used to fit to the cumulative failure data. The slope parameters, β, is the measure of how quickly the failures are accumulating over failure time or cycles. The characteristic life parameter, η, is the test time or life cycle

corresponding to 62.3% cumulative failure point. For solder joint reliability studies the First-Time-To-Failure (FTTF) and Mean-Time-To-Failure (MTTF) is often reported for failure data. FEA modeling and predicted solder fatigue life is often compared to the MTTF. Caution is needed in interpreting the predicted fatigue life (N_f) to the Weibull cumulative distribution result (FTTF and MTTF).

12.8. CONCLUSIONS

A DFR methodology for evaluation of Pb-free solder performance in electronic packaging assemblies has been developed.

Comprehensive mechanics of materials modeling knowledge-base has been developed for 95.5Sn-3.8Ag-0.7Cu and 99.3Sn-0.7Cu solder. This has enabled non-linear finite element analysis of Pb-free solders for elastic-plastic-creep analysis or viscoplastic analysis.

Low cycle fatigue models have been developed for fatigue life prediction analysis of Pb-free solders. Solder Fatigue database has been developed for post-processing solder fatigue life prediction after FEA computation of inelastic strain or inelastic energy density fatigue failure parameters.

Global-local FEA submodeling techniques have been developed, calibrated and applied successfully to thermal cycling simulation test cases.

Reliability test and analysis methodologies and test cases for reliability test methods and used for validation of FEA modeling of Pb-free solder fatigue performance.

ACKNOWLEDGMENTS

The author acknowledge with appreciation the funding for the Pb-free solder research by the Ministry of Education (MOE) and the Science and Research Council of A*STAR if Singapore. The School of Mechanical and Aerospace Engineering of Nanyang Technological University for supporting the research labs and facilities. Research colleagues and students who worked on the Pb-free solder research program.

REFERENCES

1. European Parliament, Directive 2002/95/EC on the restriction of the use of certain hazardous substances in electrical and electronic equipment, Official J. Eur. Union, pp. L37/19–L37/23, February 2003.
2. European Parliament, Directive 2002/96/EC on waste of electrical and electronic equipment, Official J. Eur. Union, pp. L37/24–L37/38, February 2003.
3. J.H.L. Pang, S.C.K. Wong, and Z.P. Wang, Guest Editors, Lead-free and lead-bearing solders, Soldering and Surface Mount Technology, 14(3) (2002).
4. NEMI lead free interconnect project: statement of work, National Electronics Manufacturers Initiative, Herndon, VA, May 9, 2000 (www.nemi.org/PbFreePUBLIC/index.html/).
5. J.H.L. Pang, T.H. Low, B.S. Xiong, and F.X. Che, Design for reliability (DFR) methodology for electronic packaging assemblies, Proceedings of 2003 Electronics Packaging Technology Conference, 2003, pp. 470–478.
6. S.K. Sitaraman and J.H.L. Pang, Fundamentals of design for reliability, in R.R. Tummala, Ed., Fundamentals of Microsystems Packaging, Chapter 5, McGraw-Hill, 2001.
7. T. Siewert, S. Liu, D.R. Smith, and J.C. Madeni, Database for solder properties with emphasis on new lead-free solders: Properties of lead-free solders, Release 4.0, NIST and Colorado School of Mines, February 11, 2002.

8. M. Abtew and G. Selvaduray, Lead-free solders in microelectronics, materials science and engineering, reports, A Review Journal, 27, pp. 95–141 (2000).
9. Technical Reports for the Lead Free Solder Project: Properties Reports: Room Temperature Tensile Properties of Lead-Free Solder Alloys, Lead Free Solder Project CD-ROM, National Center for Manufacturing (NCMS), 1998.
10. J.H.L. Pang and B.S. Xiong, Mechanical properties for 95.5Sn-3.8Ag-0.7Cu lead free solder alloy, IEEE Transactions on Components and Packaging Technologies, 28(4), pp. 830–840 (2005).
11. H.L.J. Pang, B.S. Xiong, C.C. Neo, X.R. Zhang, and T.H. Low, Bulk solder and solder joint properties for lead free 95.5Sn-3.8Ag-0.7Cu solder alloy, Proceedings of 53nd Electronic Components and Technology Conference, 27–30 May, New Orleans, Louisiana, USA, 2003, pp. 673–679.
12. J.H.L. Pang, L. Xu, X.Q. Shi, W. Zhou, and S.L. Ngoh, Intermetallic growth studies on Sn-Ag-Cu lead-free solder joints, Journal of Electronic Materials, 33(10), pp. 1219–1226 (2004).
13. H.L.J. Pang, B.S. Xiong, and T.H. Low, Low cycle fatigue models for lead free solders, Thin Solid Film, 462-463, pp. 408–412 (2004).
14. H.L.J. Pang, B.S. Xiong, and T.H. Low, Creep and fatigue properties of lead free Sn-3.8Ag-0.7Cu solder, Proceedings of 54th ECTC, June 1–4, Las Vegas, Vol. 2, 2004, pp. 1333–1337.
15. J.H.L. Pang, T.H. Low, B.S. Xiong, X. Luhua, and C.C. Neo, Thermal cycling aging effects on Sn-Ag-Cu solder joint microstructure, IMC and strength, Thin Solid Film, 462-463, pp. 370–375 (2004).
16. J.H.L. Pang, B.S. Xiong, and T.H. Low, Low cycle fatigue of lead-free 99.3Sn-0.7Cu solder alloy, International Journal of Fatigue, 26, pp. 865–872 (2004).
17. J.H.L. Pang, B.S. Xiong, and T.H. Low, Comprehensive mechanics characterization of lead-free 95.5Sn-3.8Ag-0.7Cu solder, Micromaterials and Nanomaterials, (3), pp. 86–93 (2004).
18. J.H.L. Pang, B.S. Xiong, and F.X. Che, Modeling stress strain curves for lead-free Sn-3.8Ag-0.7Cu solder, IEEE proceedings of EuroSime 2004 Conference, May 9–12, Belgium, 2004.
19. F.X. Che, J.H.L. Pang, et al., Lead free solder joint reliability characterization for PBGA, PQFP and TSSOP assemblies, Proceedings of 2005 Electronic Components and Technology Conference, 55th ECTC, 2005, pp. 916–921.
20. J.H.L. Pang, F.X. Che, and T.H. Low, Vibration fatigue analysis for FCOB solder joints, Proceedings of IEEE, 2004 Electronic Components and Technology Conference, 54th ECTC, Vol. 1, 2004, pp. 1055–1061.
21. J. Lau, W. Dauksher, and P. Vianco, Acceleration models, constitutive equations, and reliability of lead-free solders and joints, IEEE Proceedings of 2003 Electronic Components and Technology Conference, 2003, pp. 229–236.
22. A. Schubert, R. Dudek, E. Auerswald, A. Gollhardt, B. Michel, and H. Reichl, Fatigue life models for SnAgCu and SnPb solder joints evaluated by experiments and simulation, 53th ECTC, New Orlearns, Louisiana, USA, May 27–30, 2003, Proceedings of IEEE ECTC Conference, 2003, p. 603.
23. J.W. Morris, Jr., H.G. Song, and H. Fay, Creep properties of Sn-rich joints, Electronic Components and Technology Conference, 2003, pp. 54–57.
24. Y. Kariya and M. Otsuka, Mechanical fatigue characteristics of Sn-3.5Ag-X(X = Bi, Cu, Zn and In) solder alloys, Journal of Electronic Materials, 27, p. 1229 (1998).
25. C. Kanchanomai, Y. Miyashita, and Y. Mutoh, Low cycle fatigue behavior of lead-free solder 96.5Sn/3.5Ag, Journal of Electronic Materials, 31, p. 142 (2002).
26. C. Kanchanomai, Y. Miyashita, and Y. Mutoh, Low cycle fatigue behavior of Sn-Ag, Sn-Ag-Cu and Sn-Ag-Cu-Bi lead-free solders, Journal of Electronic Materials, 31, p. 456 (2002).
27. J.H.L. Pang, P.T.H. Low, and B.S. Xiong, Lead-free 95.5Sn-3.8Ag-0.7Cu solder joint reliability analysis for micro-BGA assembly, Proceedings of IEEE, 2004 Inter Society Conference on Thermal Phenomena, ITherm 2004, Vol. 2, 2004, pp. 131–136.
28. J.H.L. Pang, A. Yeo, T.H. Low, and F.X. Che, Lead-free 96.5Sn-3.5Ag flip chip solder joint reliability analysis, Proceedings of IEEE, 2004 Inter Society Conference on Thermal Phenomena, ITherm 2004, Vol. 2, 2004, pp. 160–164.
29. J.H.L. Pang and R. Dudek, Lead free solder materials and reliability performance, Short Course Notes at 7th Electronics Packaging Technology Conference, 6th December 2005.
30. A. Schubert and J. Pang, Lead free solder materials and reliability performance, Short Course Notes at 4th Electronics Packaging Technology Conference, Lead-Free Workshop, 10th December 2002.
31. A. Syed and J. Pang, Solder joint reliability: materials, modeling and testing, Workshop notes, APACK 2001, 5 December 2001.
32. J.H.L. Pang and A. Yeo, Lead-free electronics manufacturing and reliability challenges, MTA 2003 Forum, Eco-engineering: Environmental-oriented Manufacturing symposium, September 2003, Singapore.

33. J.H.L. Pang, Lead free solder reliability, Short Course Notes, GlobalTRONICS Technology Conference, 2004, Singapore.
34. S. Knecht and R. Fox, Constitutive relationship and creep-fatigue life model for eutectic tin-lead solder, IEEE Transaction on Components, Hybrids, and Manufacturing Technology, 13(2), pp. 424–433 (1990).
35. J.H.L. Pang and D.Y.R. Chong, Flip chip on board solder joint reliability analysis using 2-D and 3-D FEA models, IEEE Transactions on Advanced Packaging, 24(4), pp. 499–506 (2001).
36. R. Darveaux, K. Banerji, and G. Dody, Reliability of plastic ball grid array assembly, in J. Lau, Ed., Ball Grid Array Technology, McGraw-Hill, Inc., New York, 1995.
37. R. Darveaux, Effect of simulation methodology on solder joint crack growth correlation, Electronic Components and Technology Conference, IEEE, 2000, pp. 158–169.
38. L. Anand, Constitutive equations for hot working of metals, J. Plasticity, 1, pp. 213–231 (1985).
39. S.B. Brown, K.H. Kim, and L. Anand, An internal variable constitutive model for hot working of metals, International Journal of Plastic, 5, pp. 95–130 (1989).
40. L.F. Coffin, A study of the effects of cyclic thermal stresses on a ductile metal, Trans. ASME, 76, p. 931 (1954).
41. S.S. Manson, Fatigue a complex subject-some simple approximations, Experimental Mechanics, 1965.
42. T.J. Kilinski, J.R. Lesniak, and B.I. Sandor, Modern approaches to fatigue life prediction of SMT solder joints, in J.H. Lau, Ed., Solder Joint Reliability Theory and Applications, New York, 1991.
43. X.Q. Shi, H.L.J. Pang, W. Zhou, and Z.P. Wang, Low cycle fatigue analysis of temperature and frequency effects in eutectic solder alloy, International Journal of Fatigue, pp. 217–228 (2000).
44. H.D. Solomon, Predicting thermal and mechanical fatigue lives from isothermal low cycle data, in J.H. Lau, Ed., Solder Joint Reliability, Van Nostrand Reinhold, New York, Chapter 14, 1991.
45. S. Knecht and L. Fox, Integrated matrix creep: application to accelerated testing and lifetime prediction, in J.H. Lau, Ed., Solder Joint Reliability, Van Nostrand Reinhold, New York, Chapter 16, 1991.
46. J.H.L. Pang, C.W. Seetoh, and Z.P. Wang, CBGA solder joint reliability evaluation based on elastic-plastic-creep analysis, Journal of Electronic Packaging, 122, pp. 255–261 (2000).
47. W. Engelmaier, Solder attachment reliability, accelerated testing, and result evaluation, in J.H. Lau, Ed., Solder Joint Reliability, Van Nostrand Reinhold, New York, Chapter 17, 1991.
48. J.D. Morrow, Cyclic Plastic Strain Energy and Fatigue of Metals, ASTM STP 378, American Society for Testing and Materials, Philadelphia, PA, 1964, p. 45.
49. H.U. Akay, N.H. Paydar, and A. Bilgic, Fatigue life prediction for thermally loaded solder joints using a volume-weighted averaging technique, J. Electronic Packaging, 119, pp. 228–235 (1997).
50. X.Q. Shi, H.L.J. Pang, W. Zhou, and Z.P. Wang, A modified energy-based low cycle fatigue model for eutectic solder alloy, Scripta Material, 41(3), pp. 289–296 (1999).
51. J.H.L. Pang, K.H. Ang, X.Q. Shi, and Z.P. Wang, Mechanical Deflection System (MDS) test and methodology for PBGA solder joint reliability, IEEE Transactions on Components and Packaging Technologies, 24(4), pp. 507–514 (2001).
52. J.H.L. Pang, D.Y.R. Chong, and T.H. Low, Thermal cycling analysis of flip chip solder joint reliability analysis, IEEE Transactions on Components and Packaging Technologies, 24(4), pp. 705–712 (2001).
53. J.H. Lau and Y.-H. Pao, Solder Joint Reliability of BGA, CSP, Flip Chip, and Fine Pitch SMT Assemblies, McGraw-Hill, 1997.
54. ASTM Standards, ASTM E8M: Standard Test Methods for Tension Testing of Metallic Materials [Metric], The American Society for Testing and Materials, Vol. 03.01, 1998.
55. X.Q. Shi, W. Zhou, H.L.J. Pang, and Z.P. Wang, Effect of temperature and strain rate on mechanical properties of 63Sn/37Pb solder alloy, Journal of Electronic Packaging, 121(September), pp. 179–185 (1999).
56. X.Q. Shi, Z.P. Wang, W. Zhou, H.L.J. Pang, and Q.J. Yang, A new creep constitutive model for eutectic solder alloy, Journal of Electronic Packaging, (March) (2002).
57. ASTM Standards, ASTM E606: Standard Practice for Strain-Controlled Fatigue Testing, The American Society for Testing and Materials, Vol. 03.01, 1998.

13

Application of Moire Interferometry to Strain Analysis of PCB Deformations at Low Temperatures

Arkady Voloshin

Department of Mechanical Engineering and Mechanics, Lehigh University, Bethlehem, PA 18015, USA

Abstract

Microelectronics packaging has been developing rapidly over the past ten years due to the demands for faster, lighter and smaller products. Printed circuit boards (PCBs) provide mechanical support and electrical interconnection for electronic devices. Many types of composite PCBs have been developed to meet various needs. Recent trends in reliability analysis of PCBs have involved developing of the structural integrity models for predicting lifetime under thermal environmental exposure; however the theoretical models need verification by the experiment.

The objective of the current application is the development of an optical system and testing procedure for evaluation of the thermal deformation of PCBs using moire interferometry. Due to the special requirements of the specimen and test condition, the existing technologies and setups were updated and modified. The discussions on optical methods, thermal loading chambers, and image data processing are presented. The proposed technique and specially designed test bench were employed successfully to measure the thermal deformations of PCB in the temperature range of $-40°C$ to $+160°C$. The video-based moiré interferometry was used for generating, capturing and analysis of the fringe patterns. The obtained information yields the needed coefficients of thermal expansion (CTE) for tested printed circuit boards.

13.1. INTRODUCTION

Printed circuit boards (PCBs) consist of one or more layers of metal, bonded onto insulating substrates that are fabricated from the glass-fiber-reinforced thermosetting resin. Thermal stresses in microelectronics interconnections are developed due to mismatch of the coefficient of thermal expansion (CTE) between the package and substrate. Accurate measurements of the PCB thermal deformation will provide actual values for the CTE which is of importance for reliability assessment of the microelectronics assembly.

Number of optical methods such as shadow moiré [1–3], differential interferometry [4], optical profilometry [5], Fizeau interferometry [6], electronic speckle pattern inter-

ferometry [7,8], digital speckle correlation [9,10], and moiré interferometry, described in details by D. Post et al. [11], are available for experimental evaluation of the deformations. Some of these, like shadow moiré, electronic speckle pattern interferometry and digital speckle correlation, do not provide the necessary deformation resolution, while others require specialize equipment and knowledge for use.

Compared with other, the moiré interferometry is an ideal optical technique for high resolution, non-contact and full-field deformation measurements. Some measurements of the CTE by using moire interferometry have been reported [12–14], and the test results were in good agreement with other methods, such as strain gage techniques. These reports were for the temperature range from $+20°C$ to around $+110°C$ only. However, higher temperatures occur in the reflow process and in some application, e.g., under the hood electronic devices. When the PCB substrate material is heated above T_g (glass transition temperature), it behaves as an elastomer. The manufacturers are trying to develop PCBs with high T_g. However, in the meantime, it will be helpful if the test data for the organic substrate can be provided near or above T_g. Low temperature also might be encountered in storage, shipping or operation of the electronics components. Almost all reliability testing standards for microelectronics require the temperature cycling down to $-40°C$ or lower, but few studies appear to have been done at these temperature ranges. A real time low temperature test of PCBs will provide useful information about their behavior. As a composite structure made of organic thermosetting resin, woven fiberglass and copper foil, used as inner-layer, the PCB's dimensional stability is significantly impacted by any temperature change. Initial surface irregularities and deformation, residual stress, free-edge effect and fixture mode of specimen may have a significant influence on the test results. Finally, several technical issues have to be resolved in the practice of specimen grating application, optical interferometry setup, experimental arrangement and data processing.

The present chapter discusses the technical concerns, describes the experimental procedure and presents results for four types of PCBs. The thermoelastic properties (coefficient of thermal expansion) were measured in both the warp and the fill directions. Samples were made of a common, commercially pressed core (2116) woven glass epoxy substrate sandwiched between copper cladding. Obtained results show the viability of methodology presented here for CTE measurements in the wide temperature range. The obtained data can be used in the models for predicting the reliability of the microelectronics packages.

13.2. OPTICAL METHOD AND RECORDING OF FRINGE PATTERNS

One of the most intriguing problems in experimental mechanics is the accurate measurement of relatively small deformations without applying too much change to the specimen's stiffness. This quest brought the development of the so called moire method. The idea is extremely simple: attach a very low stiffness grating to the sample surface and monitor its deformation due to the loads experienced by the sample. The direct monitoring of the grating is called a "grid" method, which has some merits, but suffers from low sensitivity [15]. But even the moire method, that is based on interference of the two gratings: "master" and "specimen," as late as up to 1960's had a limited resolution since the technique was limited to gratings of up to 30–40 lines/mm. Such systems allowed measurement of strains in the several percent range. To overcome this limitation one may pursue either increase in the number of lines per unit length or devise methods that will allow extracting information not only from the full fringe values, but also work with the fractional fringes.

The review of the moire history shows that researches pursued both avenues with significant success. The current systems allow measurement of strains on the order 20 microstrains, i.e., 500-fold improvement in the sensitivity of moire method.

13.2.1. Fractional Fringe Approach

Moiré pattern characterizes the displacement at every point on the specimen surface. By using digital image analysis and fractional fringe analysis techniques [16], the displacement can be computed at every location throughout the moire pattern. This is accomplished by utilizing a basic optical law [17] that relates light intensities in a moiré field to the corresponding displacements as

$$W(x, y) = W_0 + 1/2\pi \cdot f \cdot \arccos[(I(x, y) - I_0)/I_1], \tag{13.1}$$

where W_0 is the displacement at some starting point, f is the frequency of the grating, $I(x, y)$ is the light intensity at the point under consideration, I_1 is the intensity amplitude of the first harmonic term in the optical law, and I_0 is the average background light intensity.

Equation (13.1) utilizes only the first term in the series [17], since here one is dealing with nearly pure sinusoidal light intensity distribution due to the nature of the moiré grating. The values of I_1 and I_0 are determined by the image analyzer for each half fringe separately. Since the digital image analyzer has light intensity resolution of 256 gray levels, one can hope for effective fringe multiplication of 512, however in practice multiplication of 20 are more reliable due to inherent optical and electronic noise [18]. This approach is applicable to any moire pattern and is not dependent on the way the pattern produced. It has to be mentioned that the Equation (13.1) deals only with the fractional part of the fringe value. Its integer part W_i has to be established by either manual or automated fringe counting routine.

$$W_i = \frac{N}{f}, \tag{13.2}$$

where $1/f$ is a pitch, or distance between the grating lines at the master and N is the fringe number.

13.2.2. Grating Frequency Increase

As it is mentioned above, the alternate approach to increase the sensitivity of the moire method is to increase the frequency of the grating. But, here one comes to the problem of the diffraction effects. An effective way to overcome this is to produce a so called "phase grating" that effectively did not absorb light. This became possible due to availability of newly developed techniques used in the microelectronic industry. As result, gratings with frequency of 2240 lines/mm were produced already in 1973 [19]. To apply these gratings to the specimen surface, a replication technique was developed [20] that is used by nearly every practitioner of moire interferometry. Thus, the problem of high-frequency specimen grating was solved. The next challenge was creation of a high-frequency master grating.

13.2.3. Creation of a High-Frequency Master Grating

Moire interferometry allows operation with a very small grating pitch. The interference pattern formed by a pair of collimated coherent light beams, C′ and D′, interacting at angles $+\alpha$ and $-\alpha$, is used as a virtual master grating (Figure 13.1). The frequency of reference grating is

$$f = \frac{2}{\lambda} \sin \alpha, \tag{13.3}$$

where λ is the wavelength of the beam. The initial frequency of the specimen grating, f_s, is chosen to be half of the virtual reference grating f. The sensitivity of the measurements is controlled by f.

The $\pm 45°$ adjustable mirrors direct portions A′ and B′ of the collimated beam to angle $+\alpha$ and $-\alpha$ in the vertical plane. They form a virtual reference grating with its grating lines perpendicular to the y axis; it interacts with the corresponding specimen grating lines to form the V field. When light from A′ and B′ is blocked, beams from C′ and D′ form the U field (Figure 13.1). The system does not need a polarizing filter, however, a larger collimated input beam ($> \varnothing$ 100 mm) is required. The wavelength of the light 0.6328 μm and $\alpha = 49.4°$ results in $f = 2400$ line/mm. Thus, each fringe represents an in-plane displacement of $1/f = 416$ nm in this study.

When moire is used for deformation studies, the location of the zero-order fringe is arbitrary since any rigid-body translations are not importance for strain analysis. What is important is relative displacement and not the absolute value. Several rules have to be observed when assigning the values to visible moire fringes. Adjacent fringes differ by one fringe order, either up or down. However, in the areas of local maxima or minima the adjacent full fringes may be assigned the same order since the local extremum maybe located in between the full fringe values. Fringes of unequal orders can not intersect.

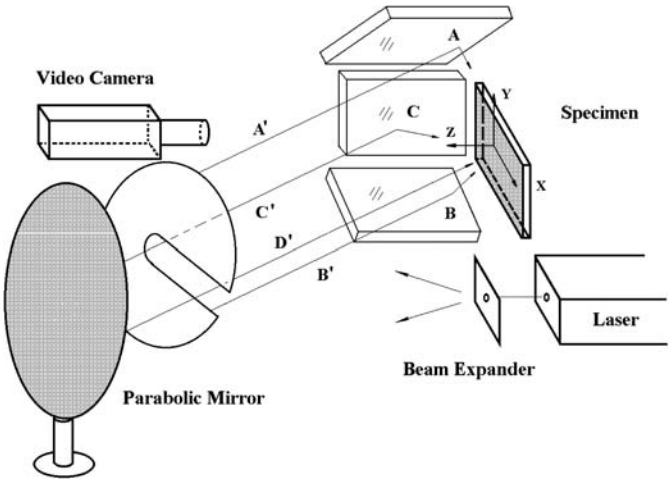


FIGURE 13.1. Four beam moiré interferometry.

Since we are looking for deformations, it is obvious that only derivatives of the displacements are of interest.

$$\varepsilon_x = \frac{\partial W_x}{\partial x} \quad \text{and} \quad \varepsilon_y = \frac{\partial W_y}{\partial y}. \tag{13.4}$$

In order to get the correct strain value sign, it is important to establish the direction of the positive strain gradient. After selecting the coordinate system, one has to assign fringe values in such a way that the derivative of the fringe number along the positive x or y direction will result in correct sign for the strain.

In cases where the sign of the gradient is not known a priori it may be established by an simple experiment. Apply the small displacement to the specimen while observing the fringe pattern. If the displacement is in the positive x direction, the fringe orders will all increase, i.e., fringes will move in the direction of the lower fringe value. For a negative fringe gradient, fringes will move in the same direction as displacement applied to the specimen.

13.2.4. Combination of the High Grating Frequency and Fractional Fringe Approach

The above combination allows significant increase in the deformation measurement sensitivity. The developed hardware-software combination resulted in a systems capable resolve displacement as low as 20 nm. This wining combination becomes a standard way to study the deformations in the systems that could never be approached even 20 years ago. The contemporary systems consist of the environmental chamber with the sample, mirror holders and camera stage. These elements were attached to a solid base plate, which makes the optical assembly stable and robust. The developed system produces exceptionally stable fringes under thermal and mechanical loading. SONY® XC-75 CCD video camera was used to capture the real time fringe patterns. An analog video output signal was saved on the video tape, later converted into an 8-bit digital signal and stored in a PC in the form of a 640×480 pixels picture for further processing. The linear scale for all measurements was in the range of 0.010 to 0.024 mm/pixel.

13.3. DATA PROCESSING

The moiré patterns were recorded during thermal loading and processed off-line. These patterns had sufficient number of fringes in order to use a simple fringe counting; there was no need to utilize any of the techniques for fractional fringe analysis [21]. The data processing was performed by a routine written in MATLAB®. Each moiré image was scanned along several selected sections; number of full fringes was automatically detected, multiplied by the pitch (416 nm) to represent the displacement along the scan line. The straight line was fitted to the obtained displacement; its slope represented the average strain along this section.

13.4. TEST BOARDS AND SPECIMEN GRATING

Etched or unetched copper (CU) panels and adhesive prepregs (type 2116) were laminated, pressed and cured to produce a rigid PCB board of the desired configuration. Four

types of PCB configurations were prepared for this study. The samples of the PCBs for testing were cut to the size of 34 × 28 mm. The warp and fill directions are corresponding to the x and y axes as shown in Figures 13.2–13.4. Each board contained various number of prepreg and CU layers as described below.

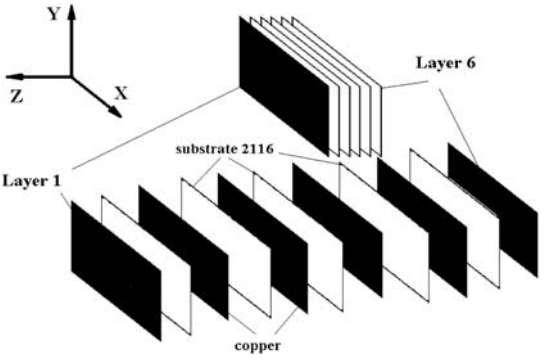


FIGURE 13.2. Configuration of board B.

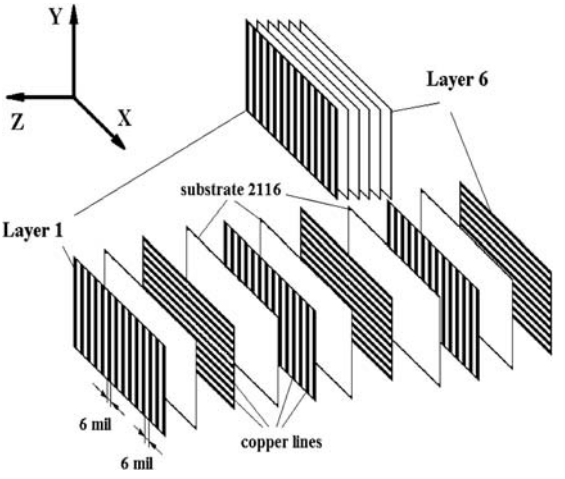


FIGURE 13.3. Configuration of board C.

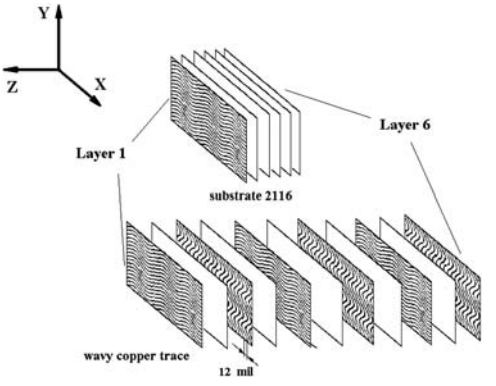


FIGURE 13.4. Configuration of board D.

Board A: 5 layers of the prepreg 2116, the final thickness was 1.14 mm;
Board B: 6 full CU layers and 5 layers of the prepreg 2116, final thickness was 1.36 mm;
Board C: 6 layers of straight copper traces (152.4 μm line width, and 152.4 μm space between lines and 5 layers of the prepreg 2116, adjacent copper layers have traces in alternating directions), final thickness was 1.25 mm;
Board D: 5 layers of the prepreg 2116 with wavy copper traces (152.4 μm line width, and 152.4 μm space between lines, adjacent copper layers have traces in alternating direction, the amplitude and wavelength of the waviness was 0.305 mm and 2.5 mm). Its final thickness was 1.25 mm.

The configurations of the last three types of test boards are shown in Figures 13.2–13.4.

A 1200 line/mm crossed-line grating was first replicated onto an 87 × 75 × 5.7 mm glass plate, which was coated with a reflective aluminum film by evaporation using Denton® DV-502A system. This grating was replicated on the specimen surface at room temperature. The specimen was then placed inside the environmental chamber. Three K type thermocouple probes were attached to the specimen to monitor its temperature during the test. The differences between the thermocouples were under 0.8°C.

13.5. ELEVATED TEMPERATURE TEST

The heat source was a resistance heater attached to the heat sink built into the chamber. Typical temperature profile is shown in the Figure 13.5 (the heat loading rate was about 2.9° C/min).

The mirrors were adjusted to provide a null field. However it is nearly impossible to get a true null field, this is especially complicated task for the four beam moiré interferometry. Thus, the initial moiré patterns were recorded and used later as a basis for analysis. The optical data was continuously recorded by a VCR during test. After the end of data acquisition, number of frames representing PCB thermal deformation at selected temperatures was selected; moiré fringe patterns were digitized by a frame grabber (MATROX Meteor II) and stored for data processing. Each frame was later scanned along vertical and horizontal sections to extract the pertinent displacement data for each increment of temperature.

If the specimen would have a uniform CTE and would be subjected to a homogeneous temperature field, the moiré fringe field would appear as uniformly spaced, parallel, straight

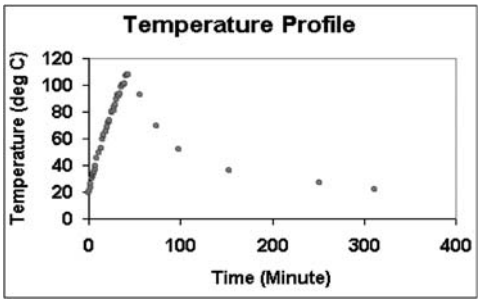


FIGURE 13.5. Typical heat loading history.

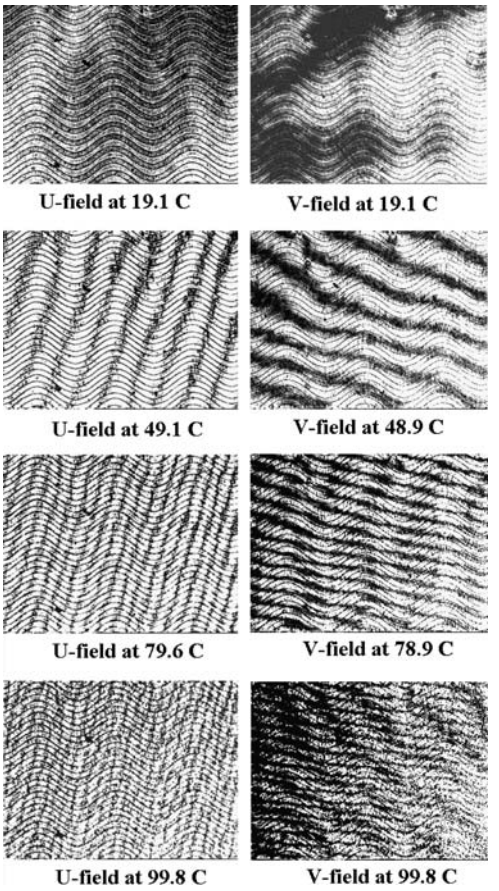


FIGURE 13.6. Temperature dependent moiré fringe patterns for board D.

lines—vertical lines for the horizontal or U-displacement field and horizontal lines for the vertical or V-displacement field. However, the null field and surface irregularity of PCB made the fringe patterns somewhat tilted.

Typical U and V-field fringe patterns for the board D are shown in Figure 13.6 for a range of temperatures. The scan lines used for analysis of the moiré fringe patterns are shown in Figure 13.7.

Scan lengths for U-field and V-field were 6.78 and 4.67 mm, respectively. Light intensity along the scan lines was acquired by a MATLAB routine and position of the full fringes were detected automatically. They were numbered consequently since the deformation was either uniform extension (heating) or uniform compression (cooling). The average temperature induced strain was obtained from a numerical derivative of the measured displacement vs. position. This process was repeated at about 10–15 different temperatures. The obtained data—measured average strain as a function of temperature—is shown in Figure 13.8 for each type of the measured boards. Slope of the data curve provides values for the apparent coefficients of thermal expansion in x and y directions. The first order curve was fitted to the data; the equation can be seen in each of the plots (Figure 13.8). The coefficient in front of "x" represents the slope of the line, i.e., the measured coefficient of thermal extension for the given test.

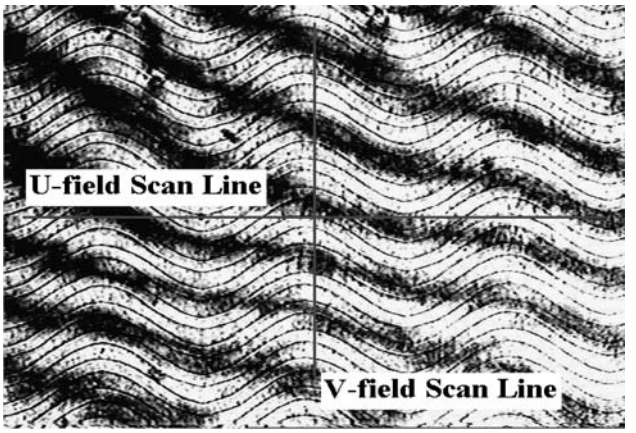


FIGURE 13.7. Scan lines location.

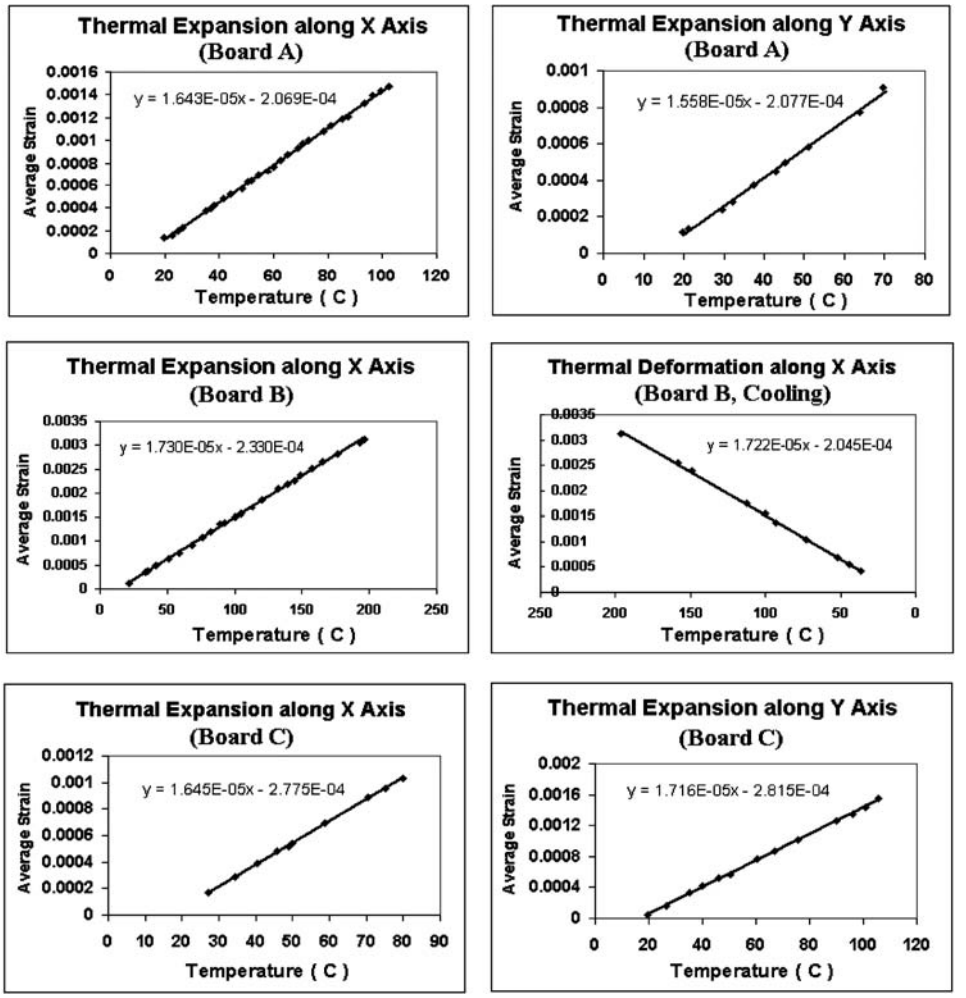


FIGURE 13.8. Temperature dependent strain.

TABLE 13.1.
Test results.

Test board		CTE (ppm/°C)	
		Warp direction	Fill direction
A	2116	17.2 ± 0.3	16.8 ± 0.3
B	2116 with full Cu	18.2 ± 0.4	18.0 ± 0.3
C	2116 with straight Cu traces	18.5 ± 0.3	18.4 ± 0.4
D	2116 with wavy Cu traces	17.6 ± 0.4	16.9 ± 0.4

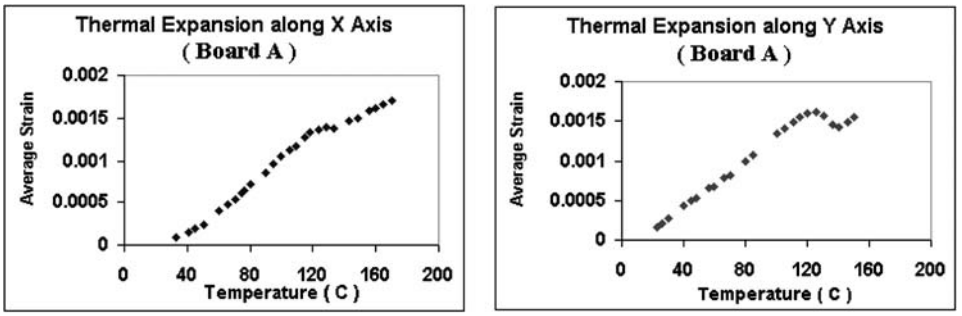


FIGURE 13.9. Test board A (25–160°C).

The presented data demonstrates that the thermal expansion of PCB boards is basically linear from the room temperature to +110°C. The data was recorded both during the heating up (+20 to +110°C) and cooling down (+110 to +20°C) of the board. The obtained results did not differ significantly, as may be seen from the data for "board B" (Figure 13.8). The CTEs were not significantly different in the wrap and fill directions, as may be seen in the Table 13.1.

The data presented in Table 13.1 was acquired for the temperature range of room to +1100°C. This range is still below the glass transition temperature, T_g. Above +1100°C the *in-situ* fringe patterns became not very clear, mainly due to moisture accumulation on the glass window of the environmental chamber.

To study the board behavior at higher temperatures, a single window glass was replaced by a double layer glass window. This modification allowed increasing the temperature in the chamber up of +160°C; at higher temperatures the specimen grating became damaged. The test results for the board A are presented in Figure 13.9. It is obvious that the glass transition temperature for this specimen is in the range of +120°C to +130°C. The CTE in the X and Y directions usually decreases above the T_g, as it is seen in Figure 13.9, mainly due to the resin modulus decrease. The effect of the fiber reinforcement becomes stronger since its modulus is relatively unaffected by the higher temperature.

13.6. LOW TEMPERATURE TEST

The sample preparation for low temperature test was done in the same manner as for the elevated temperature test. The specimen grating was replicated at the room temperature and the sample was placed in the environmental chamber. It was built to provide a range

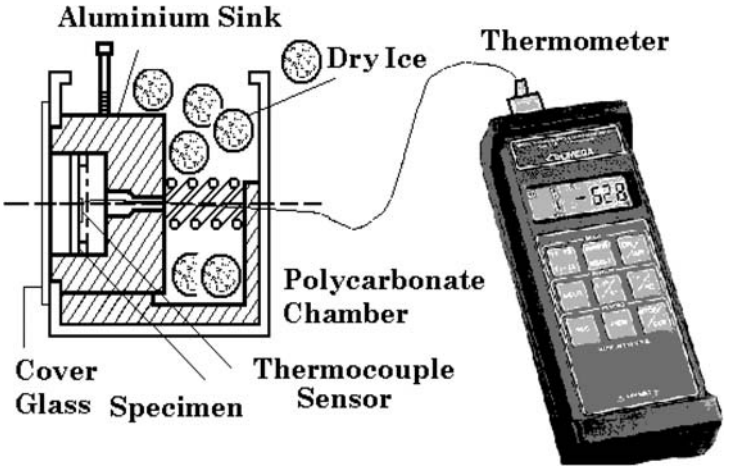


FIGURE 13.10. Cooling chamber.

of temperatures from the room to −40°C. Several cooling methods were evaluated for the applicability and convenience.

The liquid nitrogen (−196°C) can provide fast and efficient cooling loading, but boiling will result in the chamber tremble, which may be detrimental for the optical testing. Thermoelectric coolers (TECs) are solid state heat pumps that utilize the Peltier effect. However, high performance of these coolers can not be achieved without developing and building a custom TEC setup.

Dry ice, frozen carbon dioxide (CO_2), also can be used for cooling since its temperature is −78.5°C. Cooling by dry ice is safe, environmentally clean and easy to control. A specially designed chamber (Figure 13.10) capable of holding up to one kilogram of dry ice was built. This quantity was sufficient to cool the chamber and the sample to the needed temperature of −40°C.

The heat sink was made of aluminum. Kodak projector slide cover glass (102 × 83 × 1.2 mm) was used to build a four layer glass window. Multi-layer glass window was needed for two reasons: to prevent the fogging and to reduce the heat loss. The mist on the outside glass layer was wiped out by the Varitemp® heat gun.

Dry ice was placed into the chamber and the optical setup was adjusted to get the parallel null field fringe pattern. The data recording was initiated. It went from the room temperature to −40°C, the cooling phase took about 30–40 minutes (Figure 13.11, Mode I); the data recording continued until the sample reached the room temperature again (Mode II). During thermal loading process the temperature gradient was between 2°C/min to less than 0.2°C/min.

Fringe patterns and corresponding temperatures were recorded and processed off-line. The typical moiré patterns and the temperature dependent average strain for the Mode I are shown for all four boards (Figures 13.12–13.15).

The CTEs of all boards were basically constant, with values of 16–18 ppm/°C in the range of −40°C to +20°C. The results were in agreement with those recorded in the temperature range of +20°C to +110°C.

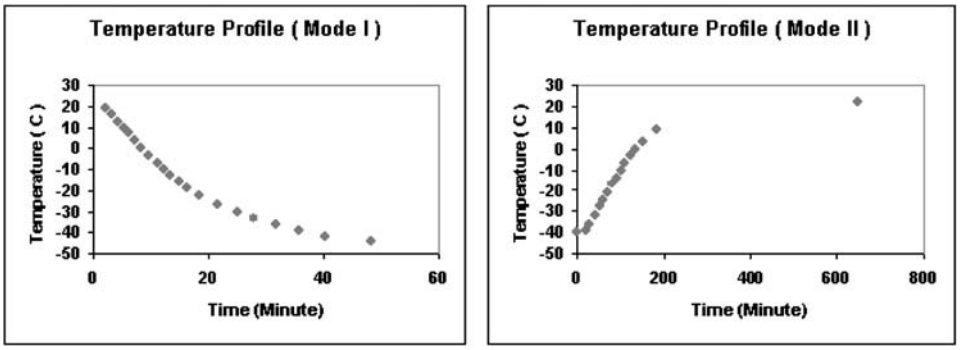


FIGURE 13.11. Two modes of cooling.

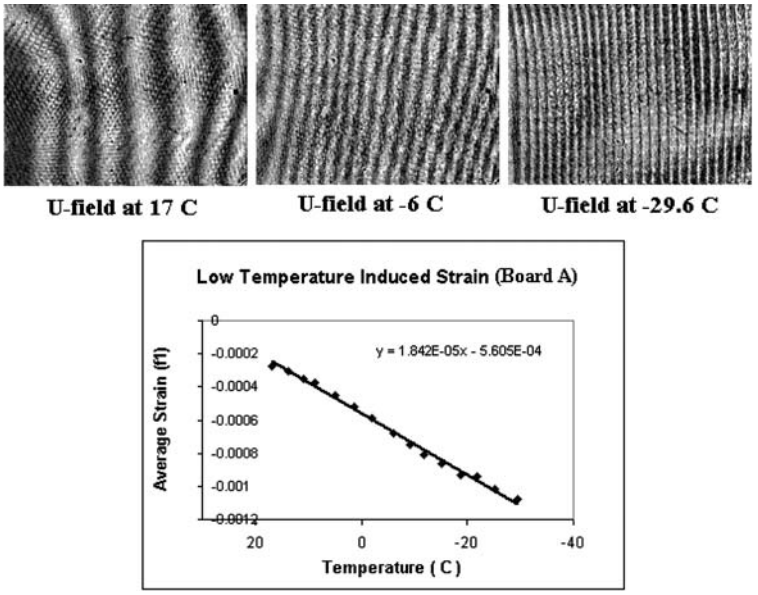


FIGURE 13.12. Cooling test result of board A.

13.7. CONCLUSIONS

Despite the wide proliferation of the finite element and other procedures capable to model the thermal behavior of the complex printed circuit boards, still here is a need to measure the deformations experienced by the PCB under various thermal loadings. To fully understand the nature of these deformations full-field methods are necessary. One of the most successful candidates for this task became the moire interferometry method for real-time deformation measurements. This method provide high special resolution, the displacement on the order of 20 nm are measured along the scan lines up to 10 mm long. It is applicable to study in-plane and out-of plane deformations. The main restriction is that the starting surface has to be reasonable flat, how flat depends on the required resolution.

A robust scheme of moiré interferometry for real-time observation of thermal deformations was described here. The scheme was implemented with both high and low tem-

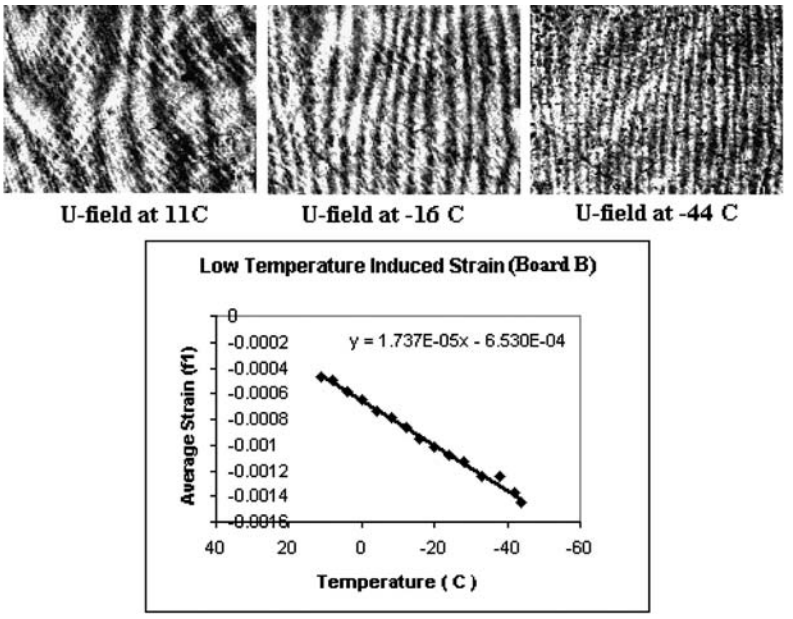


FIGURE 13.13. Cooling test result of board B.

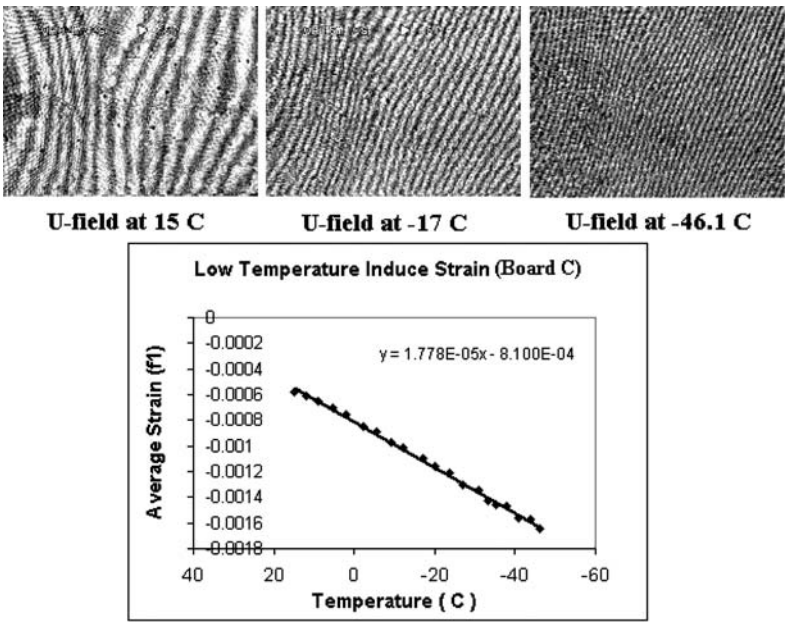


FIGURE 13.14. Cooling test result of board C.

perature environmental chambers that produce temperature profiles needed for the thermal testing. The in-plane deformations of the four types of PCBs were measured in the -40 to $+160$°C range. The test results show that the CTE of all boards are basically con-

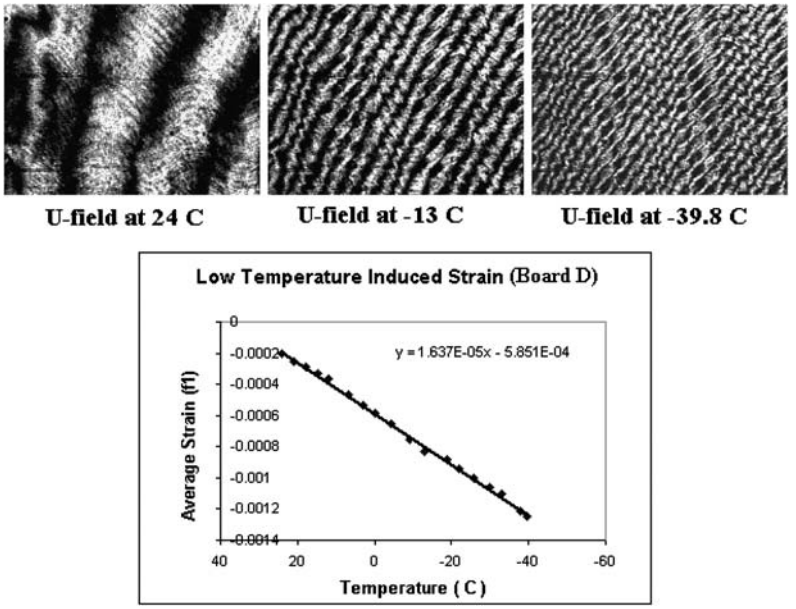


FIGURE 13.15. Cooling test result of board D.

stant from $-40°C$ to T_g, with values of 16–18 ppm/°C. It is the first work describing the *in-situ* measurements of the PCBs coefficient of thermal expansion in the low temperature range ($-40°C$).

The system described here was successful for the thermal-mechanical behavior measurement of PCB specimens in the wide temperature range. The obtained data may be used for verification of the numerical procedures used for prediction of the CTE for the complicated PCBs arrangements. They also may be useful for prediction of the PCB lifetime under various thermal environments.

The main drawback of this technique is that there are no commercially available systems that allow a user to place a sample, push a button and get the results. Significant preparation is required to get the moire patterns and not less effort is needed for proper interpretation of the recorded images. But, these difficulties are justified by the wealth of full-field data obtained from the real samples under real thermal loads.

ACKNOWLEDGMENT

This research was supported by the Pennsylvania Department of Community and Economic Development for the program “Miniaturized Electronics” (Contract No. 20-906-0015) which was administered through an Alliance of Visteon Systems, LLC, Lehigh and Penn State Universities. Special thanks go to Dr. Terence Clark and Larry Schmidt at Visteon for fruitful discussions, as well as to Dr. Terence Clark for program coordination.

REFERENCES

1. Y. Wang and P. Hassell, Measurement of thermally induced warpage of BGA packages/substrates using phase-stepping shadow moiré, Proc. of the 1st Electronic Technology Conference, 1997, pp. 283–289.
2. K. Verma, D. Columbus, B. Han, and B. Chandran, Real-time warpage measurement of electronic components with variable sensitivity, 1998 Electronic Components and Technology Conference, 1998, pp. 975–980.
3. P.B. Hassell, Advanced warpage characterization: location and type of displacement can be equally as important as magnitude, Proc. of Pan Pacific Microelectronics Symposium Conference, Feb. 2001.
4. S. Dilhaire, T. Phan, E. Schaub, and W. Claeys, High sensitivity and high resolution differential interferometer: micrometric polariscope for thermomechanical studies in microelectronics, Microelectronics Reliability, 37(10/11), pp. 1587–1590 (1997).
5. C.W. Tsai, C.H. Lee, and J. Wang, Deconvolution of local surface response from topography in nanometer profilometry with a dual-scan method, Optics Letters, 24(23), pp. 1732–1734 (1999).
6. K. Verma, D. Columbus, and B. Han, Development of real time/variable sensitivity warpage measurement technique and its application to plastic ball grid array package, IEEE Transactions on Electronics Packaging Manufacturing, 22(1), pp. 63–70 (1999).
7. T. Ahrens and M. Krumm, Deformation measurement at components, printed wiring boards and microelectronic assemblies to ensure the reliability of a system, Proceedings of EUPaC'96, 2nd European Conference on Electronic Packaging Technology, 1996, pp. 108–112.
8. K.-S. Kim, J.-H. Kim, J.-K. Lee, and S.-S. Jarng, Measurement of thermal expansion coefficients by electronic speckle pattern interferometry at high temperature, Journal of Materials Science Letters, 16(21), pp. 1753–1756 (1997).
9. Y.C. Chan and F. Yeung, Nondestructive detection of defects in miniaturized multilayer ceramic capacitors using digital speckle correlation techniques, IEEE Transactions on Components, Packaging, and Manufacturing, Part A, 18(3), pp. 677–684 (1995).
10. H. Lu, C. Yeh, and K. Wyatt, Experimental evaluation of solder joint thermal strain in a CSP using digital speckle correlation, 1998 InterSociety Conference on Thermal Phenomena, 1998, pp. 241–245.
11. D. Post, B. Han, and P. Ifju, High Sensitivity Moire, Spring-Verlag, New York, 1994.
12. T. Ratanawilai, B. Hunter, G. Subbarayan, and D. Rose, A comparison between moire interferometry and strain gages for effective CTE measurement in electronic packages, 2000 InterSociety Conference on Thermal Phenomena, 2000, pp. 246–252.
13. A.F. Bastawros and A.S. Voloshin, Transient thermal strain measurements in electronic packages, IEEE Transactions on Components, Hybrids and Manufacturing Technology, 13(4), pp. 961–966 (1990).
14. B. Han, Z. Wu, and S. Cho, Measurement of thermal expansion coefficient of flexible substrate by moire interferometry, Experimental Techniques, 25(3), pp. 22–25 (2001).
15. A.J. Durelli and W.F. Riley, Development in the grid method of experimental stress analysis, Proc. SESA, XIV(2), pp. 91–100 (1957).
16. A.S. Voloshin, C.P. Burger, R.E. Rowland, and T.S. Richard, Fractional moire strain analysis using digital imaging technique, Experimental Mechanics, 26, pp. 254–258 (1986).
17. C.A. Sciammarella, Basic optical law in the interpretation of moire pattern applied to the analysis of strain—Part I, Experimental Mechanics, 5, pp. 154–160 (1965).
18. A.F. Bastawros and A.S. Voloshin, Thermal strain measurements in electronic packages through fractional moiré interferometry, Journal of Electronic Packaging, 112, pp. 303–308 (1990).
19. K. Kato, F. Yamato, T. Murota, and T. Jimma, Improvement on method of measuring strain using interference fringes of diffracted beams at gratings on specimens, Bull. JSME, 16(100), pp. 1513–1523 (1973).
20. D. Post and W.A. Baracat, High sensitivity moire interferometry—a simplified approach, Experimental Mechanics, 21(3), pp. 100–104 (1981).
21. A.S. Voloshin, P.H. Tsao, and R.A. Pearson, In situ evaluation of residual stresses in an organic die-attach adhesive, Journal of Electronic Packaging, 120(3), pp. 314–318 (1998).

14

Characterization of Stresses and Strains in Microelectronics and Photonics Devices Using Photomechanics Methods

Bongtae Han

CALCE Electronic Products and Systems Center,
Mechanical Engineering Department,
University of Maryland, College Park, MD 20742, USA

14.1. INTRODUCTION

The traditional role of mechanical analysis in electronic packaging had been reliability assessment of microelectronics and photonics devices at the final stage of development. The shrinking product development cycle time, however, has changed the role of mechanical analysis from a problem solving (passive) mode to a predictive (active) mode, where the mechanical analysis is performed for (1) design optimization and (2) reliability prediction of a new technology product at its conceptual stage of development. This dependency of product development on mechanical analysis has fostered increasing activity in mechanical experimentation, both for specific studies and for guidance of numerical modeling.

As the components and structures involved in high-end devices are made smaller, the thermal gradient increases and the strain concentrations become more serious. Numerical analyses have been used extensively to estimate stresses and strains in device structures. Although one can model almost any kind of device for complex loading and boundary conditions, simplifications and uncertainties are inevitable. The models and results usually require verification by other means. Accordingly, advanced experimental techniques are in high demand to provide accurate solutions for deformation studies of microelectronics and photonics devices.

In recent decades, numerous optical methods for deformation measurements have matured and emerged as important engineering tools [1–3]. With these methods, the data are received as whole-field fringe patterns representing contour maps of equal displacements. Recently, several methods have been applied to microelectronics and photonics product development [4]. They include moiré interferometry [5–25], microscopic moiré interferometry [11,13,17,26,27], Twyman/Green interferometry [11,13,28,29], far infrared

Fizeau interferometry [30,31] and shadow moiré [32–39]. The first two provide contour maps of in-plane displacement fields for stress/strain analyses, and the next three map out-of-plane displacement fields for warpage analyses. This chapter presents recent developments of the methods as tools for product development of microelectronics and photonics devices and illustrates selected applications for design evaluation, failure analysis, and verification of numerical modeling.

14.2. STRESS/STRAIN ANALYSIS

14.2.1. Moiré Interferometry

The general scheme of moiré interferometry is illustrated in Figure 14.1 [1]. A high-frequency cross-line grating on the specimen, initially of frequency f_s, deforms together with the specimen. A parallel (collimated) beam, B_1, of laser light strikes the specimen and a portion is diffracted back, nominally perpendicular to the specimen, in the $+1$ diffraction order of the vertical lines of the specimen grating. Light from the mutually coherent collimated beam B_2 is diffracted back in its -1 order. Since the specimen grating is deformed as a result of the applied loads, these diffracted beams are no longer collimated. Instead, they are beams with warped wavefronts, where the warpages are related to the deformation of the grating. These two coherent beams interfere in the image plane of the camera lens, producing an interference pattern of dark and light bands, which is the N_x moiré pattern.

Similarly, mutually coherent collimated beams B_3 and B_4, centered in the vertical plane, are diffracted in $+1$ and -1 diffraction orders by the nominally horizontal lines of the deformed specimen grating. These two diffracted beams interfere to produce the N_y

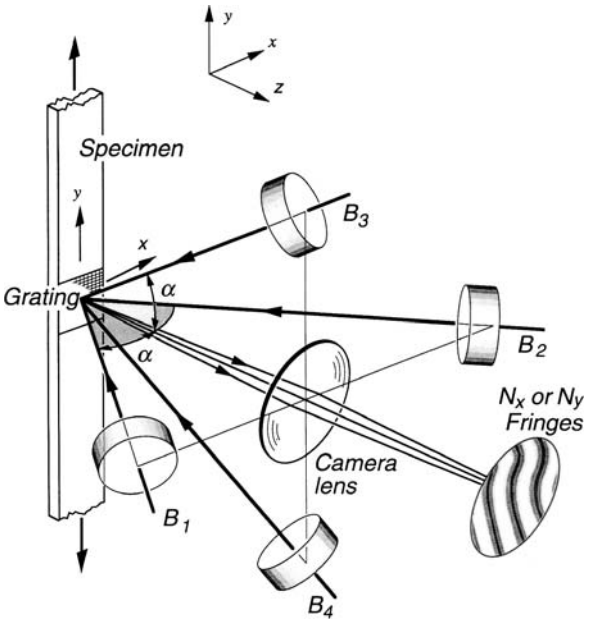


FIGURE 14.1. Schematic illustration of four-beam moiré interferometry to record the N_x and N_y fringe patterns, which depict the U and V displacement fields.

moiré pattern. In practice, beams B_1 and B_2 are blocked, so the N_y fringes are viewed alone. Alternately, B_3 and B_4 are blocked to view the N_x fringes.

These moiré patterns are contour maps of the U and V displacement fields, i.e., the displacements in the x and y directions, respectively, of each point in the specimen grating. The relationships, for every x, y point in the field of view, are

$$U(x, y) = \frac{1}{2f_s} N_x(x, y),$$
$$V(x, y) = \frac{1}{2f_s} N_y(x, y). \tag{14.1}$$

In routine practice of moiré interferometry, $f_s = 1200$ lines/mm (30,480 lines/in). In the fringe patterns, the contour interval is $1/2f_s$, which is 0.417 μm displacement per fringe order. The sensitivity is its reciprocal (i.e., the number of fringes generated per unit displacement), 2.4 fringes per μm displacement. For microscopic moiré interferometry, described below, sensitivity of 57.6 fringe contours per μm displacement has been achieved.

This is an excellent physical explanation. It is consistent with the mathematical derivation, which defines the intensity distribution in the moiré pattern and its relationship, Equation (14.1) to the fringe orders [1]. However, another physical explanation is equally compelling; perhaps it is more closely related to our experience and intuition, and perhaps it is more helpful for understanding the relationship between the deformation and the moiré pattern. It is very simple.

In the second physical description of moiré interferometry, a virtual reference grating is created by the two beams B_1 and B_2 in their zone of intersection. It interacts with the deformed specimen grating to form the moiré pattern. The frequency, f, of the virtual reference grating is determined by the angle α and wavelength λ by $(2\sin\alpha)/\lambda$. When the ± 1 diffraction orders are used, the frequency, f, of the virtual reference grating becomes $2f_s$ [1]. Then the relationships of Equation (14.1) can be rewritten as

$$U(x, y) = \frac{1}{f} N_x(x, y),$$
$$V(x, y) = \frac{1}{f} N_y(x, y). \tag{14.2}$$

14.2.2. Extension: Microscopic Moiré Interferometry

Special considerations arise for deformation measurements of tiny specimens or tiny regions of larger specimens. The relative displacements within a small field of view will be small (even if the strains are not small), so the number of moiré fringes might not be enough for an accurate analysis. Perhaps the most important consideration, therefore, is the need for increased displacement sensitivity–enhanced sensitivity beyond the high sensitivity discussed above.

In the method called *microscopic moiré interferometry*, sensitivity is increased progressively by two techniques [40,41]. The first is an *immersion interferometer*, whereby the specimen is coupled optically to the interferometer by a thin layer of immersion fluid. Beams corresponding to B_1–B_4 in Figure 14.1 propagate inside a medium of higher index of refraction, as illustrated in Figure 14.2. This strategy reduces the wavelength of the light propagating in the medium and thus increases the upper limit of frequency for the

virtual reference grating. The moiré interferometer illustrated in Figure 14.1 was implemented with $\lambda = 514$ nm, $\alpha = 54.3°$ and $n = 1.52$, where λ is the wavelength in air and n is the index of the fluid. Within the refractive medium shown in Figure 14.2, the wavelength was reduced to 338 nm, providing a virtual reference grating frequency of 4800 lines/mm. This is twice the frequency used (without immersion) for the conventional applications of moire interferometry shown in this chapter. By Equation (14.2), twice as many fringes are obtained for a given displacement, compared to designs without immersion.

The second technique to enhance the sensitivity is optical/digital fringe multiplication (O/DFM), whereby additional experimental data is obtained fringe shifting. An efficient algorithm is used to generate an enhanced contour map of the displacement field, whereby the map displays m times as many fringe contours as the original moiré pattern [42]. In practice, $m = 12$ has been achieved for microscopic moiré interferometry,

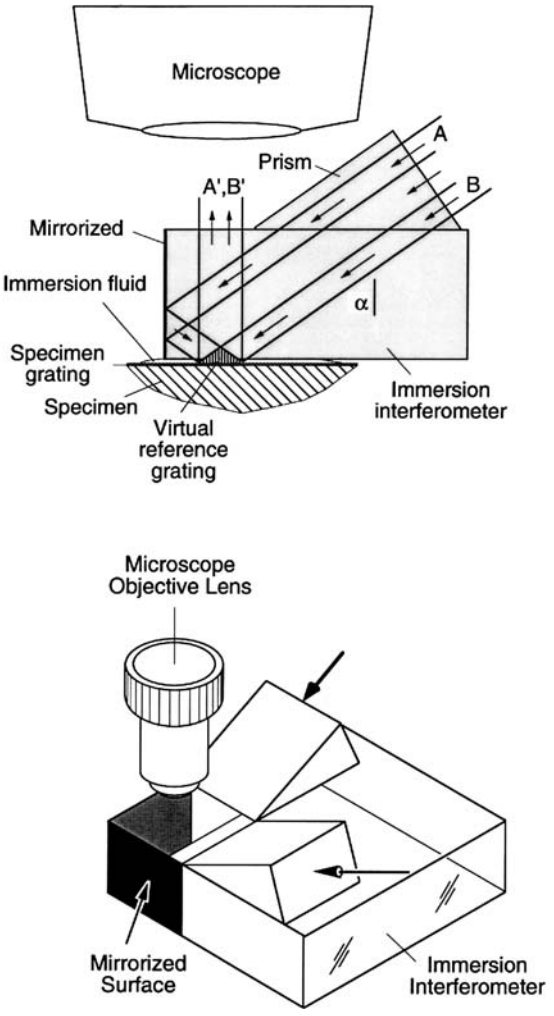


FIGURE 14.2. Mechanical and optical arrangement for microscopic moiré interferometry.

which with the doubled sensitivity of the immersion interferometer, provides increased sensitivity by a factor of 24.

14.2.3. Specimen Gratings

The bar-and-space gratings of geometrical moiré cannot be printed with very high frequencies. Instead, phase gratings are used, which means that the grating surface consists of a regular array of hills and valleys. For most analyses, the specimen grating is applied by the replication process illustrated by cross-sectional views in Figure 14.3 [1]. A special mold is used, which is a plate with a cross-line phase grating on its surface. The grating is overcoated with a highly reflective metallic film, usually evaporated aluminum. A small pool of liquid adhesive is poured on the mold, and the specimen is pressed into the pool to spread the adhesive into a thin film. Excess adhesive is cleaned off repeatedly as it flows out. The mold is pried off after the adhesive has hardened. The weakest interface is between the metallic film and the cross-line grating, so the film is transferred to the specimen. Thus, a thin, highly reflective cross-line grating is firmly attached to the specimen surface, such that it deforms together with the specimen surface.

The adhesive thickness is typically about 25 μm (0.001 in) for larger specimens–greater than 300 mm^2. For most analyses the thickness and stiffness of the grating is negligible. Various room temperature curing adhesives can be used, including epoxies, acrylics, urethanes, and silicone rubbers. Recent reports of success with instant cyanoacrylate cements have been circulated. Adhesives that cure by exposure to ultraviolet light have been used successfully. Techniques described below for replicating specimen gratings on electronic packages, to cope with the small size and tiny openings, yield gratings about 2 μm thickness.

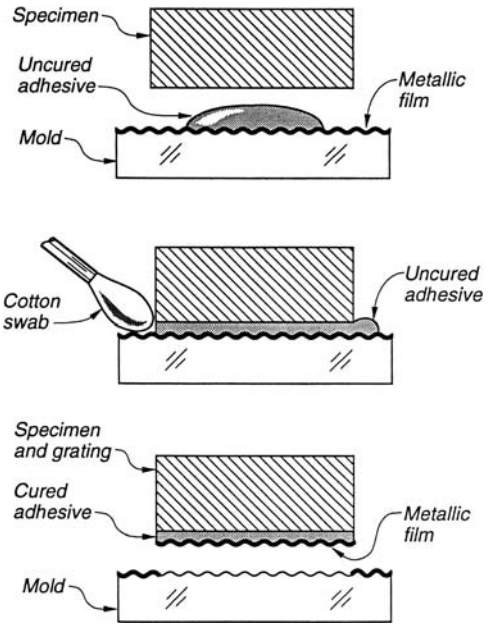


FIGURE 14.3. Steps in producing the specimen grating by a casting or replication process; the reflective metallic film is transferred to the specimen grating.

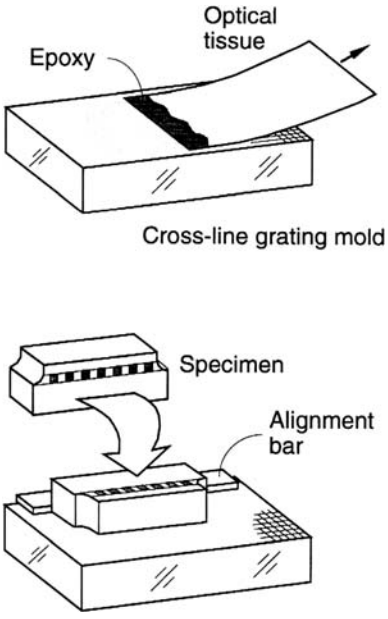


FIGURE 14.4. Procedure to replicate a specimen grating on a specimen with a complex geometry.

14.2.3.1. Grating Replication for Complex Geometry A special technique is required for replicating a specimen grating on the cross sections of microelectronics and photonics devices because they usually have such tiny and complex geometries that the excess adhesive produced by the grating replication procedure shown in Figure 14.3 cannot be swabbed away. The excess adhesive is critical since it could reinforce the specimen and change the local strain distribution.

An effective replication technique was developed to circumvent the problem [4,9,11,43]. First, a tiny amount of liquid adhesive, usually an epoxy is dropped onto the grating mold; the viscosity of the epoxy should be extremely low at the replication temperature. Then, a lintless optical tissue (a lens tissue) is dragged over the surface of the mold, as illustrated in Figure 14.4. The tissue spreads the liquid to produce a very thin layer of epoxy on the mold. The specimen is pressed gently into the epoxy, and it is pried off after the epoxy has polymerized. Before polymerization, the surface tension of the epoxy pulls the excess epoxy away from the edges of the specimen. The result is a specimen grating with a very clean edge. The specimen must be made very flat and smooth to be compatible with the thin film of epoxy.

14.2.4. Strain Analysis

Strains can be determined from the two displacement fields by the relationships for engineering strain

$$\varepsilon_x = \frac{\partial U}{\partial x} = \frac{1}{f}\left[\frac{\partial N_x}{\partial x}\right], \quad \varepsilon_y = \frac{\partial V}{\partial y} = \frac{1}{f}\left[\frac{\partial N_y}{\partial y}\right], \tag{14.3}$$

$$\gamma_{xy} = \frac{\partial U}{\partial y} + \frac{\partial V}{\partial x} = \frac{1}{f}\left[\frac{\partial N_x}{\partial y} + \frac{\partial N_y}{\partial x}\right], \tag{14.4}$$

where ε is the normal strain and γ is the shear strain at the surface of the specimen. Although it is not indicated here by the (x, y) suffix [shown in Equation (14.2)], these equations apply for every point in the field. Thus, it is the fringe gradients that determine the strains, both the normal strains and the shear strains.

14.2.5. Thermal Deformation Measured at Room Temperature

14.2.5.1. Bithermal Loading Thermal deformations can be analyzed by room temperature observations. In this technique, the specimen grating is applied at an elevated temperature, and it is allowed to cool to room temperature before it is observed in the moiré interferometer. Thus, the deformation incurred by the temperature increment is locked into the grating and recorded at room temperature [9,44]. The technique is called *bithermal loading*, implying two discrete temperatures [43].

A typical temperature increment is 80°C, whereby the grating is applied at about 100°C and observed at about 20°C. An adhesive that cures at elevated temperature is used, usually an epoxy. The specimen and mold are preheated to the application temperature, the adhesive is applied, and it is allowed to cure at the elevated temperature. The mold is a grating on a zero expansion substrate, so its frequency is the same at elevated and room temperatures. Otherwise, a correction is required for the thermal expansion of the mold.

These measurements can be achieved for cryogenic temperatures, too. In one test, the specimen grating was applied at −40°C using an adhesive that cured in ultraviolet light [1].

An example of bithermal loading is illustrated in Figure 14.5 [15]. The specimen is a flip-chip plastic ball grid array (FC-PBGA) package assembly. In the assembly, a silicon chip (6.8 mm × 6.1 mm × 1.2 mm) was first attached to an organic substrate through tiny solder bumps. The gap between the chip and the substrate was filled with an epoxy underfill

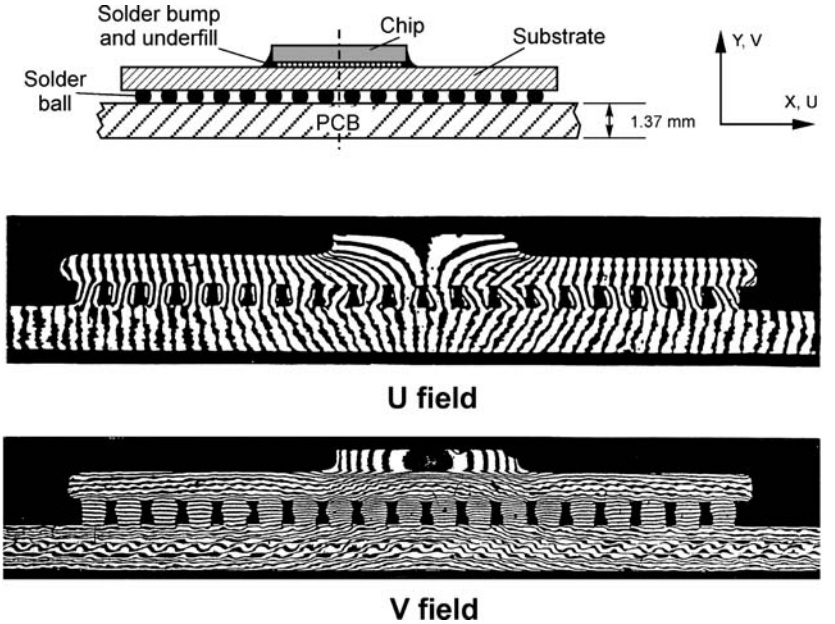


FIGURE 14.5. U and V displacement fields of a FC-PBGA package assembly, induced by thermal loading of $\Delta T = -60°C$.

to help reduce the thermal stresses induced in the solder bumps. This subassembly was then surface-mounted to a typical FR-4 printed circuit board (PCB) through larger solder ball arrays to form a final assembly. The assembly was cut and its cross-section was ground to produce a flat, smooth, cross-sectional surface. The drag method (Figure 14.4) was utilized to replicate a specimen grating at 82°C using a high temperature curing epoxy (Tra-230, Tracon), and the fringes were recorded at room temperature ($\Delta T = -60$°C). Very clean edges of the specimen grating are evident.

The V field fringe pattern reveals the detailed bending deformation of the substrate. The vertical displacements along the centerlines of the substrate and the PCB were determined from the fringe patterns and they are plotted in Figure 14.6(a). Two distinct curva-

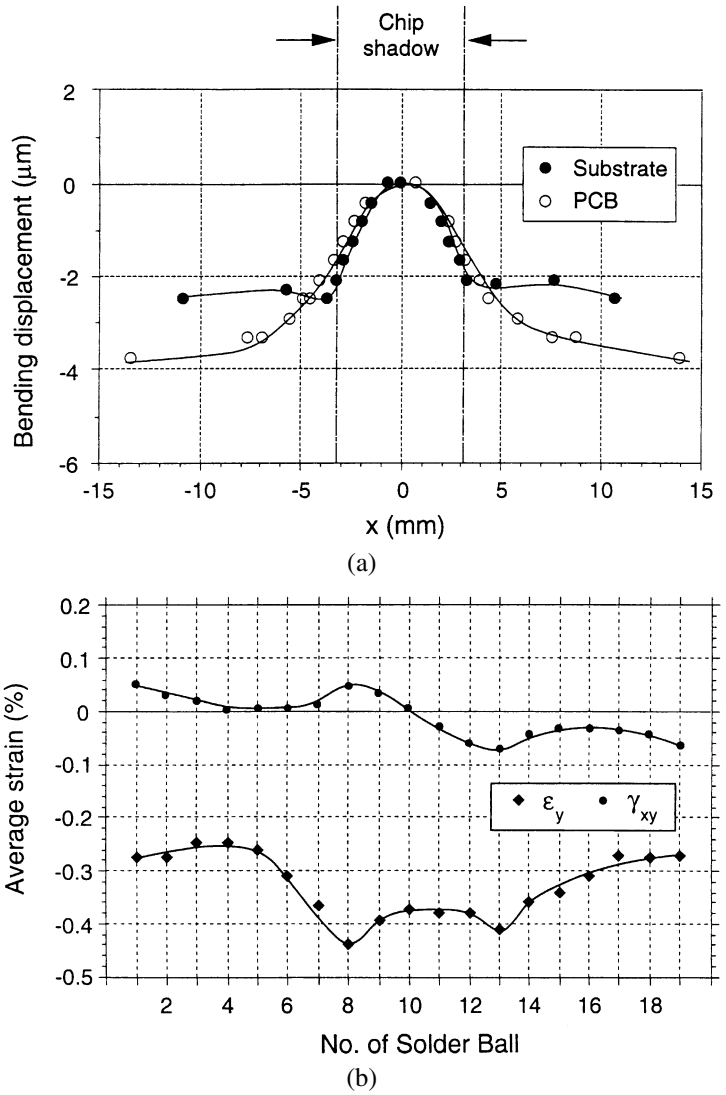


FIGURE 14.6. (a) Vertical displacements determined along the centerlines of the substrate and the PCB and (b) distribution of normal and shear strains (averaged along the vertical centerline) at each solder ball.

tures are observed, one in the area under the chip and the other in the rest of the substrate. The coefficient of thermal expansion (CTE) of the substrate was higher than that of the PCB. The substrate contracted more than the PCB during cooling, while the deformation of the substrate covered by the chip was constrained by the low CTE of the chip. This complicated loading condition produced an uneven curvature of the substrate, which resulted in an inflection point below the edge of the chip.

The substrate was connected to the PCB through the solder balls and the difference of curvature between the substrate and the PCB was accommodated by the deformation of the solder balls. The normal and shear strains (averaged along the vertical centerline) at each solder ball were calculated from the fringe patterns and they are plotted in Figure 14.6(b). The largest of these normal strains occurred in the solder ball located at the edge of the chip and its magnitude was nearly four times greater than the largest shear strain. Although symmetry about the central solder ball would be expected, the small deviations from precise symmetry are characteristic of real structures. The mirror symmetry of the shear curve is the results of shear forces acting in opposite directions on opposite sides of the centerline.

14.2.5.2. Standard Qualification Test of Optoelectronics Package Long-haul telecommunication systems are powered by high-performance semiconductor lasers. Light generated by a laser chip is coupled into an optical fiber using a system of lenses. These laser chips require an isolated environment for reliable performance over many years. In addition, their temperature must be controlled actively in order to produce a constant wavelength. Typically, the chips are assembled in a butterfly-type package [Figure 14.7(a)]. The package is manufactured by brazing metal and ceramic components together. It provides a hermetic environment once a lid is welded (resistance welding) in an inert atmosphere. The package assembly includes a thermo-electric cooler (TEC), which maintains a constant temperature of the laser chip. Qualification requirements for commercial use of these products are defined by the TELCORDIA standards. The completed assemblies are subjected to a variety of mechanical/thermal stress conditions and relevant performance parameters are monitored. The parameters must remain within a prescribed range to establish long-term reliability.

The laser assemblies produce high power light in order to transmit signals across long distances. High power is typically achieved by employing a lens system that maximizes the optical coupling between the laser chip and the fiber. A representative lens system is shown schematically in Figure 14.7(b). Initially, the laser chip is bonded to a chip carrier. The collimating lens is positioned relative to the laser chip with sub-micron accuracy while monitoring the output beam in real-time. The collimated beam is focused to a tiny spot by the second lens [focusing lens in Figure 14.7(b)]. The fiber is aligned with high precision to capture the light from the focused spot. The lenses and fiber are attached using a variety of processes like solder, epoxy and laser welding. Thus, the assembly is a complex structural system that consists of many materials joined by a hierarchy of attachment methods.

One of the main concerns for this high-coupling design is its sensitivity to relative displacements among the components. The spot-size of the light focused at the fiber is roughly equal to the diameter of the fiber-core, typically 8 μm. Tiny displacements of any optical component are sufficient to shift the focused light spot away from the fiber core and induce loss of coupling. During reliability testing, structural deformations can occur due to the mismatch in material properties and the inelastic behavior of attachment materials. Accordingly, characterization of the structural behavior is crucial to ensure a stable optical power output.

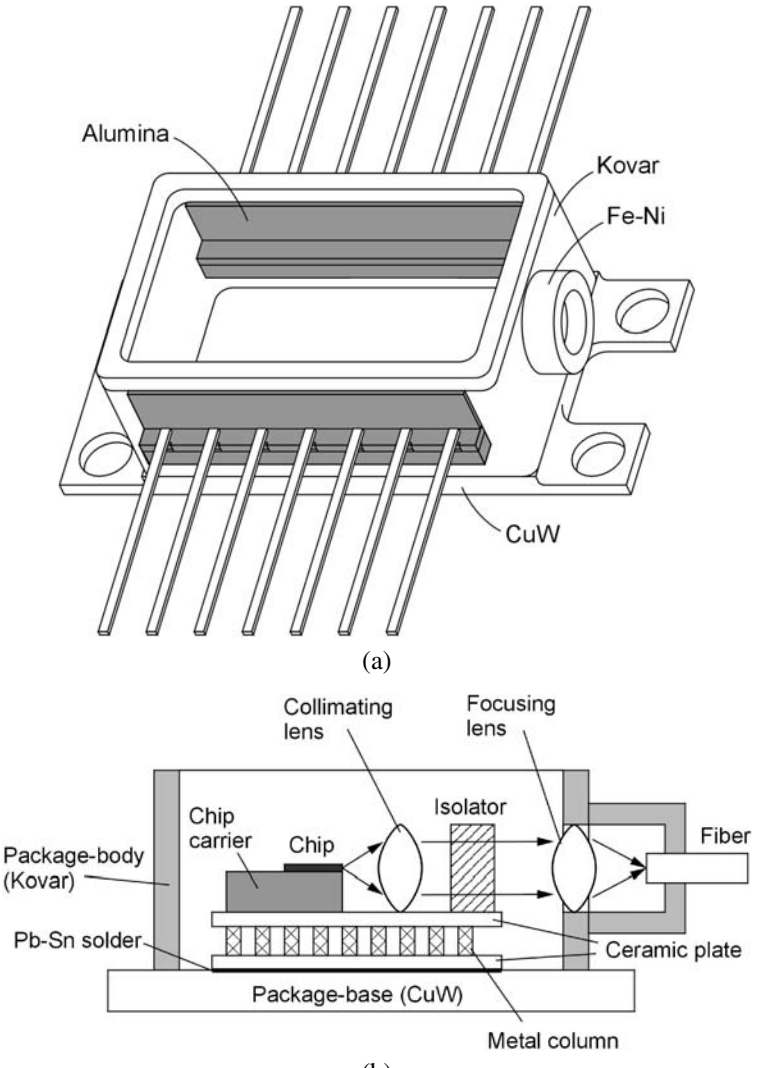


FIGURE 14.7. (a) Schematic illustration of a butterfly package, and (b) optical and electrical components in the package.

An experimental investigation of the global deformation of the package-TEC assembly was conducted using moire interferometry. In the package assembly, an array of metal columns (BiTe) is sandwiched between two ceramic plates (BeO) to form a TEC. The TEC is bonded to the package-base (CuW), with eutectic tin-lead solder. In a typical assembly process, a solder pre-form is placed on the package-base and the package is heated to melt the solder. The TEC is then placed on the molten solder and the whole assembly is cooled to room temperature. The solder solidifies to form a joint between the bottom plate of the TEC and the package base.

The package-TEC assembly was cut near the edge of the TEC to expose a row of BiTe columns, yet preserve the three-dimension integrity of the assembly. The exposed cross-section was polished and a specimen grating was replicated at room temperature.

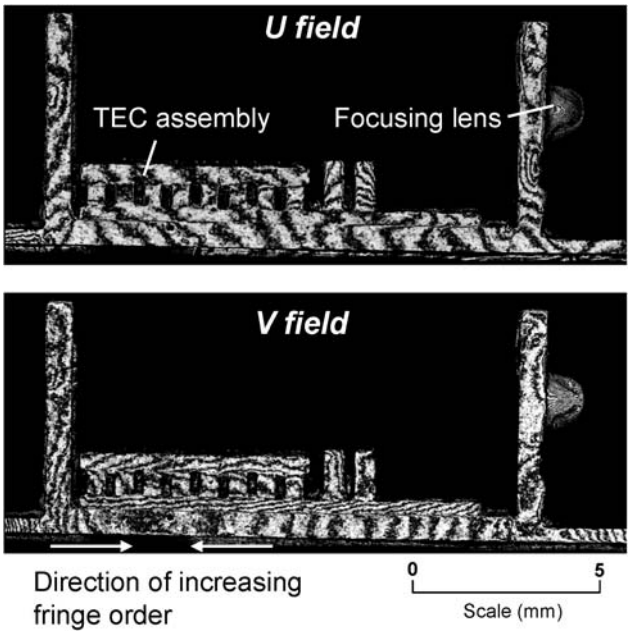


FIGURE 14.8. U and V field patterns caused by the high-temperature-storage test. The fringe numbers are multiplied by the O/DFM with $m = 4$ (the contour interval is 0.104 mm per fringe).

The interferometer was tuned with the specimen to check null fringe patterns in the U and V field to ensure no initial deformations. Then, the package assembly was subjected to the high-temperature-storage (HTS) test condition (85°C for 100 hours). Later, the specimen was cooled to room temperature and the U and V fields were recorded. The resultant fringe patterns after 100 hours of HTS are shown in Figure 14.8, where the fringes were multiplied by the O/DFM method with a fringe multiplication factor of $m = 4$ and the resultant contour interval was 0.104 μm per fringe. The presence of fringes clearly indicates that the deformation state of the assembly has changed due to the HTS.

The material of the package-base has a higher CTE than the TEC ceramic. Accordingly, the TEC had a concave deformation (center higher than the edges) at room temperature after it was soldered to the package-base. During the HTS, the stresses in the solder joint relaxed, resulting in decrease in concavity of the TEC. Since the fringe patterns measure the change in deformation, they show a convex deformation of the TEC.

The above observation provides important guidelines to the placement of the laser chip and collimating lens mounted on the TEC. It is clear that they should be located toward the front of the TEC or closer to the focusing lens. Otherwise, the change in deformation would lead to a change in angle of the beam and thus reduction in optical power coupled into the fiber.

14.2.6. Deformation as a Function of Temperature

14.2.6.1. Real-time Observation [22] The industry has been employing the *Accelerated Thermal Cycling* (ATC) test to reduce the time required for reliability assessment, where the devices are tested in much more severe environments. The results are used to predict the number of cycles to failure at the actual operating conditions by employing an acceleration

parameter. In the ATC test, the whole assembly is subjected to heating and cooling cycles in an environmental chamber.

When deformation measurements are required during the ATC, it is necessary to implement moiré interferometry with an environmental chamber that provides convection heating and cooling. The air inside the chamber must be circulated vigorously to achieve the heating/cooling rate required for a typical ATC condition. Consequently, the environmental chamber experiences vibrations, which are normally transmitted to the specimen. Moiré interferometry measures tiny displacements and those inadvertent vibrations can cause the moiré fringes to dance at the vibration frequency.

The real-time moiré setup developed to cope with the vibrations is illustrated in Figure 14.9 [22]. The major components in the setup include a portable moiré interferometer and a computer controlled environmental chamber. The specimen holder is not attached to the chamber. Instead, it is connected rigidly to the interferometer and it is essentially free from contact with the environmental chamber. Furthermore, the interferometer and the chamber are mounted on separate tables and thus the interferometer is mechanically isolated from the chamber. With this arrangement, the chamber vibrations are not transmitted to the specimen, and the moiré fringes can be documented while the chamber is being operated. Further details of the rod assembly and the temperature control can be found in Ref. [22].

Temperature dependent deformation of a wire-bond plastic ball grid array (WB-PBGA) package assembly was documented to illustrate real-time observation. The specimen is depicted schematically in Figure 14.10(a) with relevant dimensions. In the package, an active chip is first bonded to the substrate, which is a thin PCB, and the integrated circuits are connected electrically to the bond fingers on the substrate by thermo-sonic gold wire bonding. The device is then overmolded to form a PBGA package. For the final as-

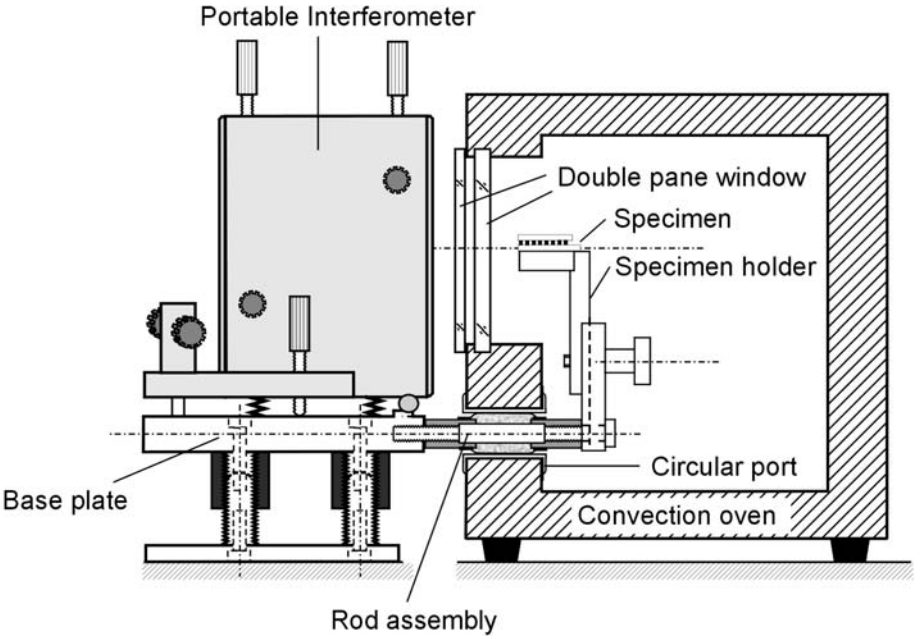


FIGURE 14.9. Schematic illustration of the moiré setup for real-time observation of thermal deformations.

sembly, the package is connected mechanically and electrically to a thicker PCB using a uniform array of solder balls.

A specimen grating was replicated on the cross section using a room temperature curing epoxy (Tra-100, Tracon). The initial null fields obtained at room temperature are shown in Figure 14.10(b). The specimen was subjected to a thermal cycle and the deformations were documented as a function of temperature. Figure 14.10(c) depicts the temperature

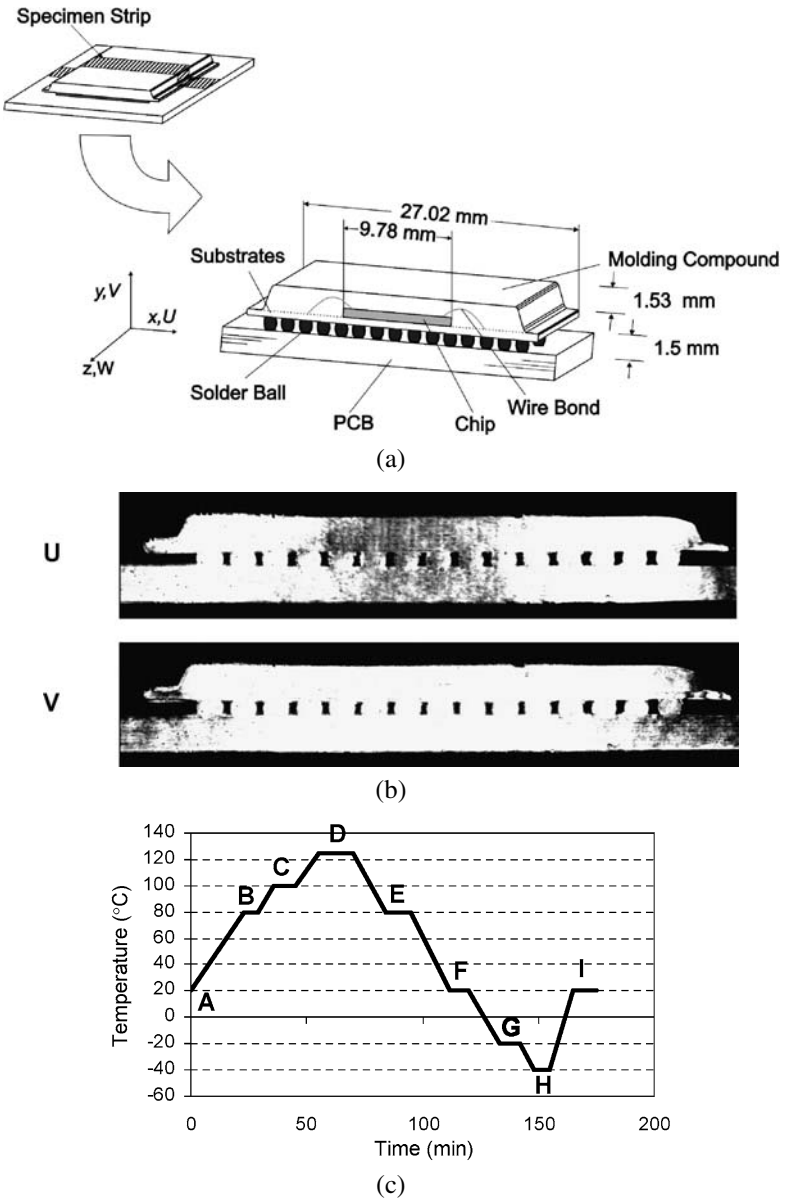
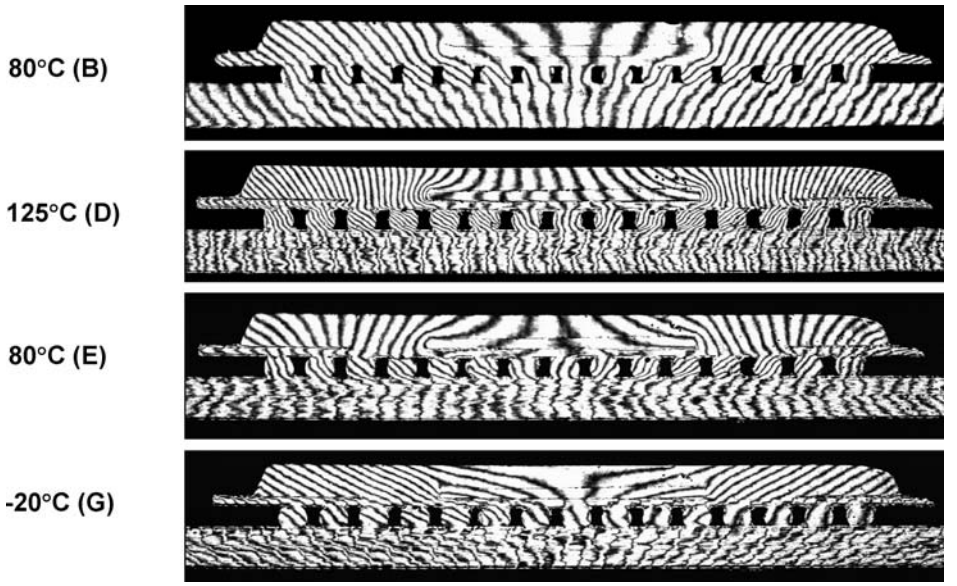


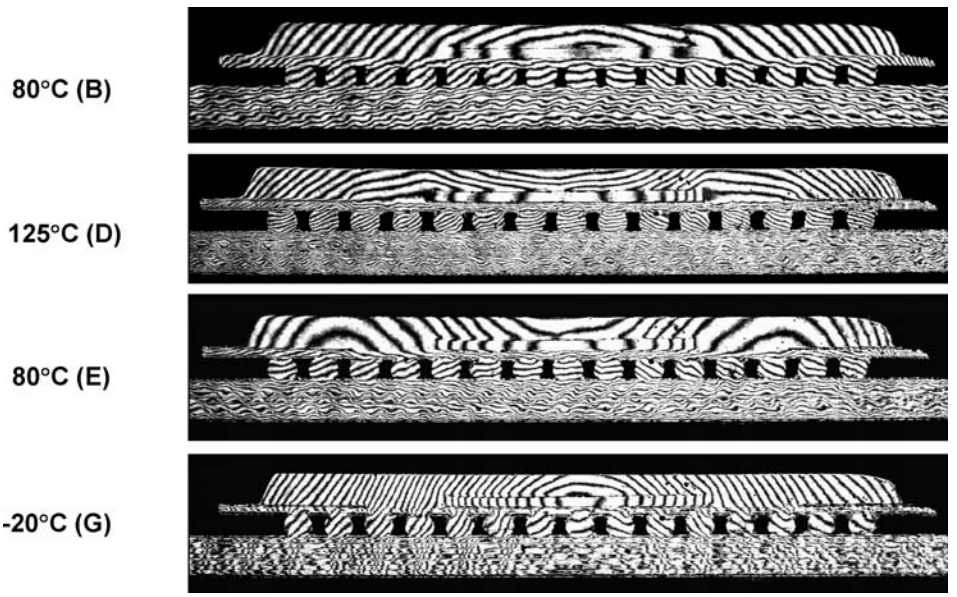
FIGURE 14.10. (a) Schematic diagram of specimen geometry with relevant dimensions, (b) initial null fields obtained at room temperature and (c) temperature profile used in the experiment.

profile used in the thermal cycle. The maximum and minimum temperature was 125°C and −40°C, respectively.

Fringe patterns were recorded at each of the lettered temperature levels. Representative fringe patterns obtained at 80°C (heating), 125°C, 80°C (cooling), and −20°C are shown in Figure 14.11. The thermal deformation in the assembly is very complicated.



(a)



(b)

FIGURE 14.11. Representative (a) U field and (b) V field fringe patterns.

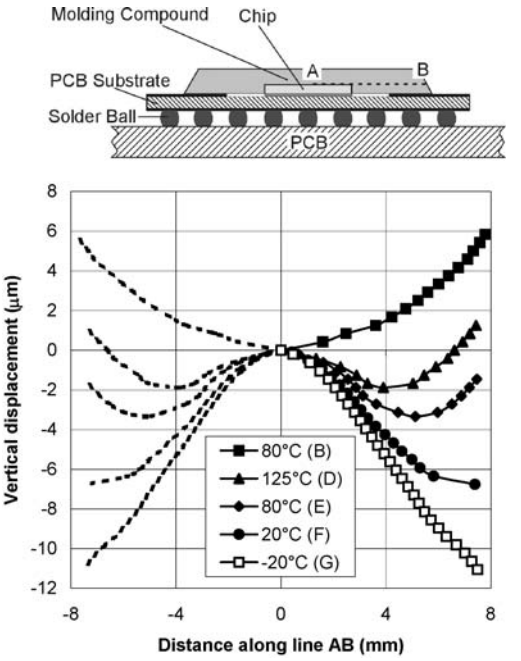


FIGURE 14.12. Bending displacements obtained along the line shown in the insert.

The magnitude and the direction of the deformation depend on the temperature dependent thermal/mechanical properties of materials used in the assembly. The vertical or bending displacements along the centerline of the package (along the line A–B in the insert of Figure 14.12) were determined from the fringe patterns and the results are plotted in Figure 14.12 to illustrate the complexity. Note that the deformations at 80°C are radically different for the heating and cooling phases, indicating non-linear creep behavior at higher temperatures. Such experimental data is critical when verification of a numerical model is undertaken for design optimization by a parametric study.

14.2.6.2. Non-Linear Deformation of Ceramic BGA Package Assembly [23] The objective of this application is to study the temperature and time dependent thermo-mechanical behavior of a ceramic ball grid array (CBGA) package assembly subjected to an ATC condition. The assembly used in the analysis was a 25 mm CBGA package assembly with 361 I/O's (19 × 19 solder interconnection array) assembled to an FR-4 PCB. A specimen with a strip array configuration was prepared from the assembly, containing five central rows of solder interconnection. The solder interconnection of the package assembly consists of a high melting point solder ball (90%Pb/10%Sn) and a eutectic solder fillet (63%Pb/37%Sn). The high melting point solder ball does not reflow during the assembly process, which provides a consistent and reproducible standoff between the ceramic package and the PCB.

The diameter of the solder ball was 0.89 mm. After the interconnection is formed, the actual separation from the package to the PCB was 0.97 mm. The pitch of the copper pads on the PCB was 1.27 mm.

The specimen was subjected to a thermal cycle and the deformations were documented as a function of temperature. Figure 14.13 depicts the temperature profile used in the thermal cycle. The heating/cooling rate was 5°C/min and the maximum and minimum

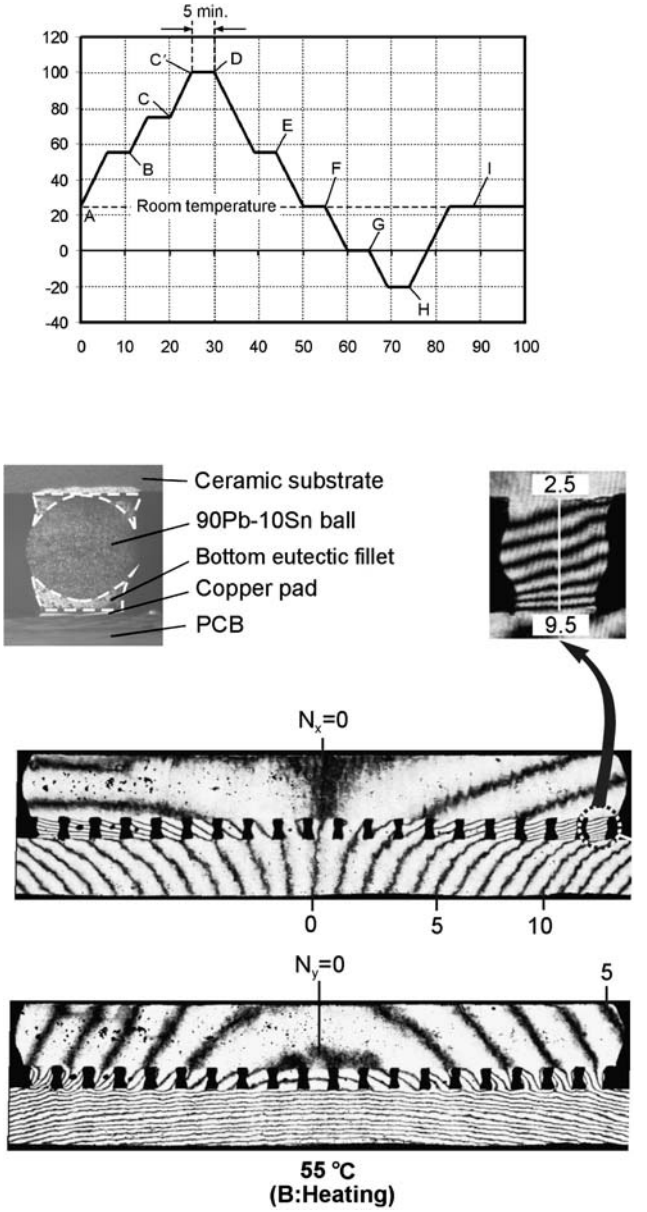


FIGURE 14.13. Temperature profile used in the thermal cycle and representative horizontal and vertical displacement fields of the assembly at 55°C during heating.

temperatures were 100°C and −20°C, respectively. To ensure a uniform temperature distribution, the specimen was kept at each target temperature for 5 minutes before measurement. The U and V moiré patterns were recorded within a fraction of a minute.

Representative horizontal and vertical displacement fields of the assembly at 55°C during heating are shown in Figure 14.13. The dominant mode of deformation of the solder interconnection is shear deformation, which is caused by the mismatch of the CTE of

the ceramic module and the PCB. Consequently, the shear strains at the interconnection increase as *the distance from the neutral point* (DNP) increases.

14.2.6.2.1. Accumulated Plastic Deformation in Solder Interconnection. The insert in Figure 14.13 shows a magnified view of the rightmost solder interconnection. This solder interconnection was analyzed to investigate the effect of the two solder materials in the interconnection. The development of inelastic strains in the solder joints during the thermal cycle is explained in Figure 14.14, where the horizontal displacements along the vertical centerlines are plotted for various stages in the thermal cycle.

It is important to note that the reflow process produced a thin eutectic solder bench between the solder ball and the copper pad while there was essentially no gap between the solder ball and the ceramic module. Consequently, the shear deformation of the top eutectic fillet was constrained by the solder ball but the bottom eutectic fillet was free to deform in shear. A detailed deformation history of the bottom fillet is described below.

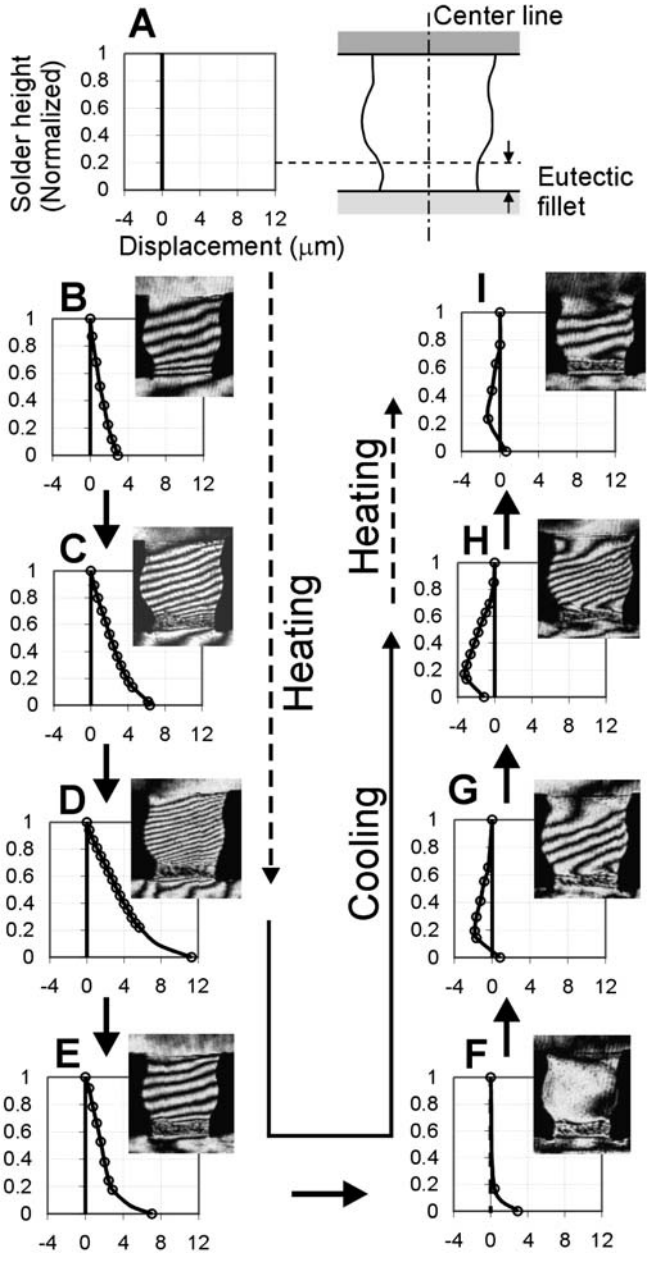
As the temperature increased, a relative horizontal displacement between the top and the bottom of the solder interconnection was caused by the CTE mismatch between the module and the PCB. The relative displacement was nearly linear over the height of the solder interconnection at the initial stage of heating (B), which indicated a nearly uniform shear strain in the interconnection. As the temperature increased, however, the slope at the high melting solder ball became distinctively different from that at the eutectic solder fillet. At the elevated temperatures, the shear strain of the eutectic fillet became much larger than that of the solder ball. The eutectic fillet has a much lower melting point compared to the solder ball, and thus it has a smaller modulus and a higher creep rate at elevated temperatures. As a result, the shear strain of the eutectic fillet increased at a much higher rate than that of the solder ball.

The most striking results were observed during cooling. When the assembly was cooled to 55°C (E) from the maximum temperature, the relative horizontal displacement (or average shear strain) of the high melting temperature solder ball was nearly the same as the deformation observed at the same temperature during heating (B). However, the shear strain in the eutectic fillet below the ball was much higher because the creep strain produced at the maximum temperature was not recovered during cooling. When the assembly was cooled to room temperature, this became more evident. The shear strain in the ball recovered completely but the shear strain in the fillet did not. As the assembly was cooled to cryogenic temperatures, the solder ball exhibited the opposite relative horizontal displacement as expected, but the direction of the relative horizontal displacement in the eutectic fillet remained unchanged. As a result, the sign of the shear strain of the solder ball became opposite to that of the eutectic fillet.

The shear strain in the eutectic fillet increased rapidly at high temperatures. Its maximum value was $\approx -2.4\%$ at 100°C, which was nearly six times as large as the shear strain in the solder ball. While the assembly was cooled to room temperature (F), the shear strain magnitude decreased at a much slower rate. The permanent shear strain in the eutectic fillet was $\approx -1.5\%$ after the heating cycle.

14.2.6.2.2. Effect of Substrate. It was observed that the shear strain in the bottom fillet remained nearly unchanged while cooling it from room to cryogenic temperatures. At high temperatures, the modulus of eutectic solder (E_{eu}) is much smaller than that of high melting solder (E_{hm}); at 100°C, $E_{eu} = 2.3$ GPa and $E_{hm} = 4.5$ GPa [45]. At low temperatures, however, E_{eu} becomes much larger than E_{hm}; at 0°C, $E_{eu} = 16.5$ GPa and $E_{hm} = 8.5$ GPa [45]. This temperature dependent stiffness of eutectic solder was attributed to the deformation behavior of the bottom fillet at low temperatures, i.e., the bottom eutectic fillet deformed more than the solder ball at high temperatures, but less at low temperatures.

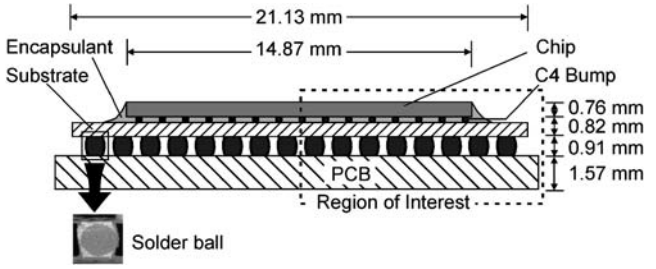
The inelastic deformation in the bottom fillet is governed only by the maximum temperature; i.e., the maximum temperature is more critical to the damage than the minimum temperature. Furthermore, due to complete stress relaxation at the maximum temperature, the maximum elastic stress occurs at the minimum temperature and its magnitude increases if the minimum temperature increases.



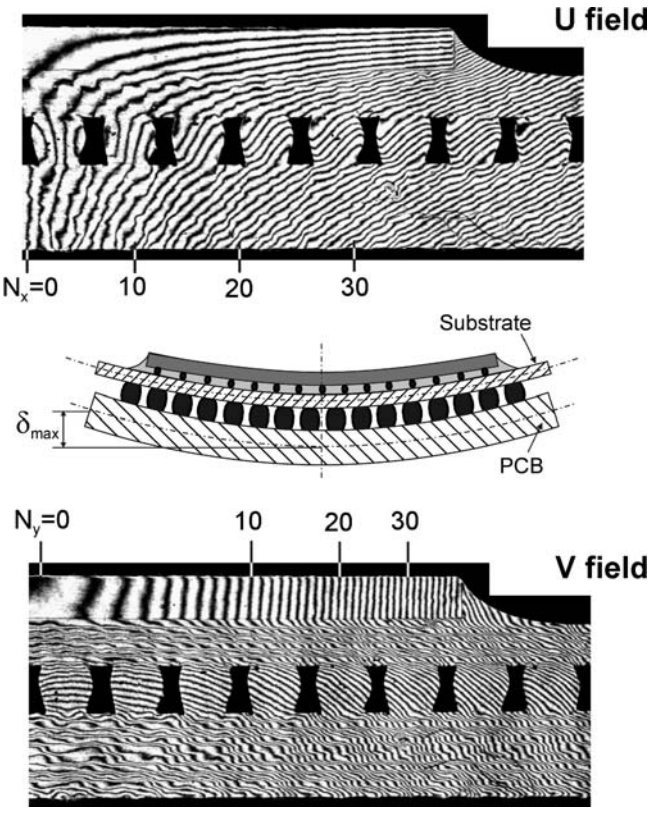
FIGUR E 14.14. *U* field fringe patterns of the rightmost solder interconnection of CBGA package assembly and the corresponding horizontal displacements determined along the vertical centerline.

The above argument is valid only when that the stresses produced by the CTE mismatch relax completely at the maximum temperature, i.e., a stress-free state exists at the maximum temperature. In the case of PBGA package assemblies, however, the stresses are much lower and complete relaxation of the stresses does not always occur during an ATC test.

An experiment indicating this alternate condition is illustrated in Figure 14.15, which shows a result obtained from a flip-chip PBGA package assembly. The cross section of the assembly is shown in Figure 14.15(a). It is important to note that the solder composition of



(a)



(b)

FIGURE 14.15. (a) Schematic diagram of FC-PBGA package assembly and (b) *U* and *V* field fringe patterns recorded at 125°C.

the interconnection is identical to that of the CBGA package assembly, i.e., high melting point solder balls with eutectic fillets. The fringe patterns in Figure 14.15(b) show U and V displacement fields of the right half of the assembly at 125°C. The patterns report very significant bending of the assembly, which was caused by the low modulus and high CTE of the organic substrate. The maximum bending displacement (δ_{max}) of the substrate and the PCB increased almost linearly as the temperature increased. The maximum displacement was about 25 μm at 125°C.

The deformations were monitored continuously while the assembly was kept at 125°C for 15 minutes. The bending deformation remained unchanged. Stress relaxation did not occur at the maximum temperature because of the low level of stress in the most of assembly, which otherwise would have been seen as a decrease in the bending displacements during the dwell period. For the same reason, the damage concentration at the eutectic fillets was not observed. The result was entirely different from that of the CBGA package assembly. Thus, to assume a stress-free condition at the elevated temperature would be improper for the analysis of typical flip-chip PBGA package assemblies.

14.2.7. Hygroscopic Deformation

A plastic encapsulated microcircuit (PEM) consists of a silicon chip, a metal support or leadframe, wires that electrically attach the chip's circuits to the leadframe, and a plastic epoxy encapsulating material, or mold compound, to protect the chip and the wire interconnects. The mold compound is a composite material made up of an epoxy matrix with silica fillers, stress relief agents, flame-retardants, and other additives.

In spite of many advantages over hermetic packages in terms of size, weight, performance, and cost, one important disadvantage of PEMs is that the polymeric mold compound absorbs moisture when exposed to a humid environment. Hygroscopic stresses arise when the mold compound and other polymeric materials swell upon absorbing moisture while the adjacent non-polymeric materials, such as the lead frame, die paddle, and silicon chip, do not experience swelling. The differential swelling that occurs between the mold compound and non-polymeric materials leads to hygroscopic mismatch stresses in the package.

In the applications discussed below, moire interferometry is used to characterize the hygroscopic swelling properties of a mold compound and subsequently to investigate the deformations of an actual package, caused by the mismatch in hygroscopic swelling. The hygroscopic deformation is compared with the thermal deformation and its implications are discussed.

14.2.7.1. Coefficient of Hygroscopic Swelling The overview of the experimental procedure is illustrated in Figure 14.16. (a) Two samples of a particular mold compound (∼2 mm thick) were first subjected to a 125°C bake for a minimum of 100 hours to remove any initial moisture that may have existed in the samples. The bake time was determined by periodically monitoring the weights of the samples until the measured weight of each sample remained unchanged for an extended period of time. (b) When the bake was completed, the samples were temporarily removed from the baking oven and a cross-line diffraction grating was replicated onto the samples at an elevated temperature of 85°C. (c) One of the two samples was left in the baking oven to ensure no extra moisture gain. This sample was referred to as the reference sample. (d) The second sample, referred to as the test sample, was subjected to an 85°C/85%RH environment and its weight was monitored periodically monitored until a virtual saturation state was reached. The virtual saturation state is defined

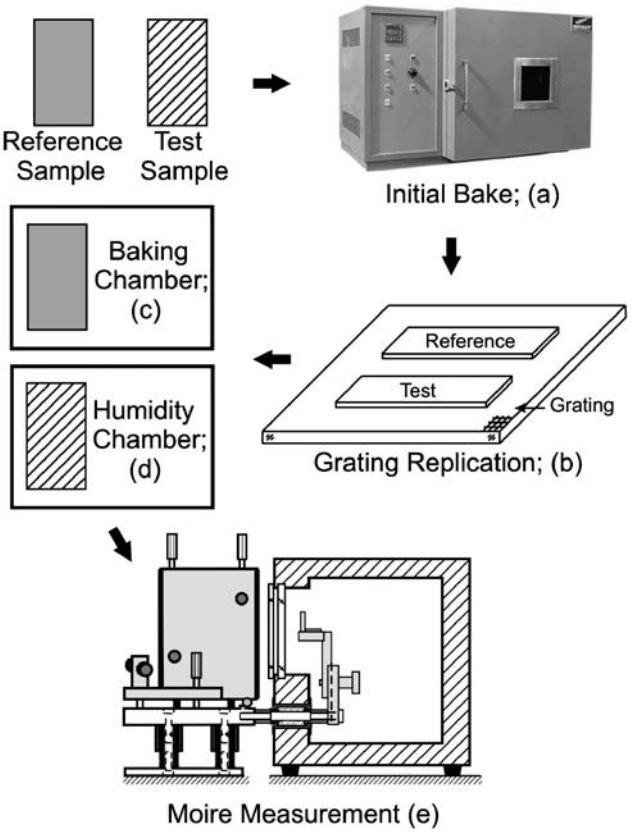


FIGURE 14.16. Overview of experimental procedure to measure hygroscopic deformations.

as the occurrence of no additional weight gain within the resolution of the balance for two to three days. (e) Once the virtual saturation state was achieved, the hygroscopic swelling measurement was performed by moiré interferometry during desorption process.

The above procedure was used to analyze the mold compound. The V field fringe patterns obtained during the desorption process are shown in Figure 14.17; desorption proceeded in the 85°C/0%RH environment. The null field pattern of the reference sample is shown in (a), and the fringe patterns of the test sample at time intervals of zero, sixteen, and four hundred hours are shown in (b), (c) and (d), respectively. The test specimen contracted as desorption progressed, as evidenced by a decrease in the number of fringes in the patterns. The fringe patterns at the zero hour [Figure 14.12(b)] represent the hygroscopic swelling at the virtual saturation point.

The hygroscopic strain, ε_h, can be determined directly from moiré fringe patterns by

$$\varepsilon_h = \frac{1}{2f_s}\frac{\Delta N}{\Delta L}, \tag{14.5}$$

where f_s is the initial frequency of the specimen grating (1200 lines/mm), ΔN is the change of fringe orders in the moiré pattern and ΔL is any gage length across which ΔN is determined.

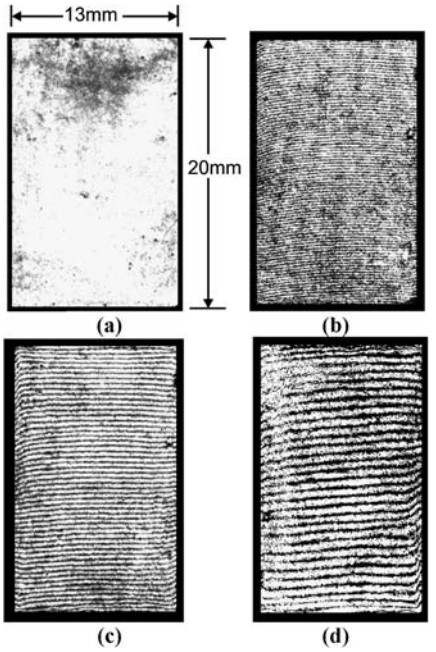


FIGURE 14.17. V field moiré patterns obtained from a mold compound. (a) Null field obtained from the reference sample; fringe patterns of the test sample at time intervals of (b) zero, (c) sixteen, and (d) four-hundred hours.

The V field hygroscopic strains are plotted against moisture content (%) in Figure 14.13. It is evident that a linear relationship exists between hygroscopic swelling and moisture content. The constant of proportionality, called the coefficient of hygroscopic swelling (CHS), is defined as

$$\beta = \frac{\%\varepsilon_h}{\%C}, \tag{14.6}$$

where β is the CHS and $\%C$ is the moisture content percentage calculated by

$$\%C = \frac{\text{Wet weight} - \text{Dry weight}}{\text{Dry weight}} \times 100.$$

"Wet weight" is defined as the weight of the sample including the weight of the absorbed moisture. The CHS is a material property of the mold compound and, if known, the hygroscopic swelling can be determined by measuring the moisture content in the mold compound. Although only V field fringes were shown in Figure 14.18, the corresponding U field patterns were documented and both fields were used to determine the CHS values. They were equal to within 5% in all cases. The average CHS value was 0.21 ($\%\varepsilon_h/\%C$) and the maximum moisture content was 0.2%.

14.2.7.2. Hygroscopic Stress in a Plastic Quad Flat Package Hygroscopic mismatch stresses in the package were documented. The package for the test was a square quad flat plastic package with 100 I/O's. The package contained a copper lead-frame and a chip with dimensions of $6.36 \times 6.36 \times 0.5$ mm. The package was prepared as shown in Figure 14.19 to investigate the interaction between the mold compound and the chip. The opposing sides

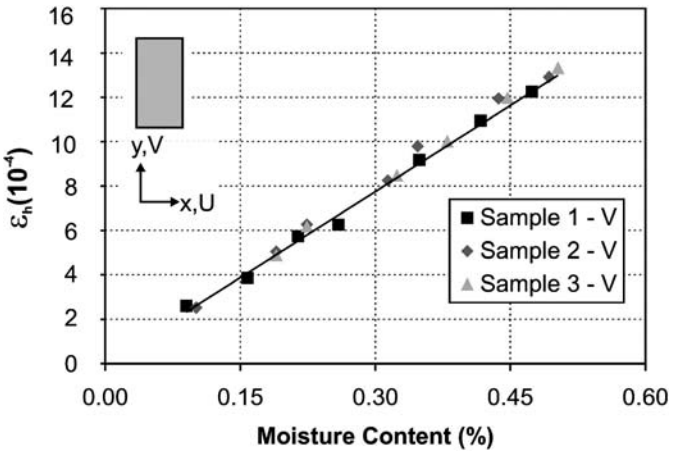


FIGURE 14.18. Hygroscopic strain vs. moisture content (%) obtained from the moiré fringes.

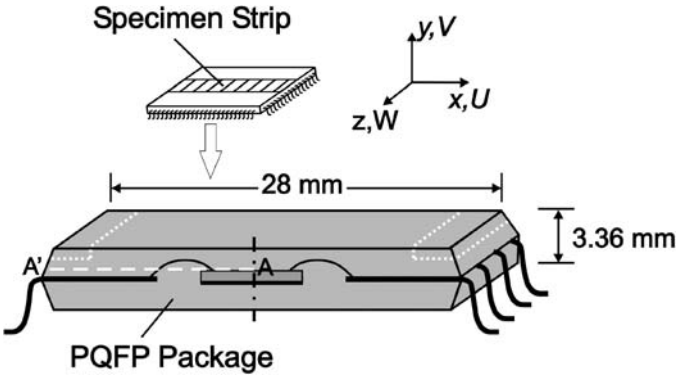


FIGURE 14.19. PQFP package for moiré experiments. The CTE and CHS of the mold compound of the package were determined from the regions marked by dashed boxes.

of the package were trimmed and ground using a precision grinding machine until the silicon chip was exposed on both sides. This specimen configuration preserved the symmetric boundary conditions. After the existing moisture was removed by baking at 125°C, the specimen grating was replicated onto the package surface at 85°C using an ultra-low expansion (ULE) grating mold.

The moiré system was tuned at the grating replication temperature and the null field patterns were taken as shown in Figure 14.20(a). The package was cooled to 25°C and the resulting thermal deformations were measured. These fringe patterns are shown in Figure 14.20(b), which represent in-plane displacement maps, induced by a bi-thermal loading of $\Delta T = -60°\text{C}$.

The package was then subjected to 85°C/85%RH until the saturation state was achieved. The package was installed in the real-time moiré system at 85°C and the deformations caused by hygroscopic swelling at the saturation state were measured. The resulting fringe patterns are shown in Figure 14.21(a). No reference specimen was used for

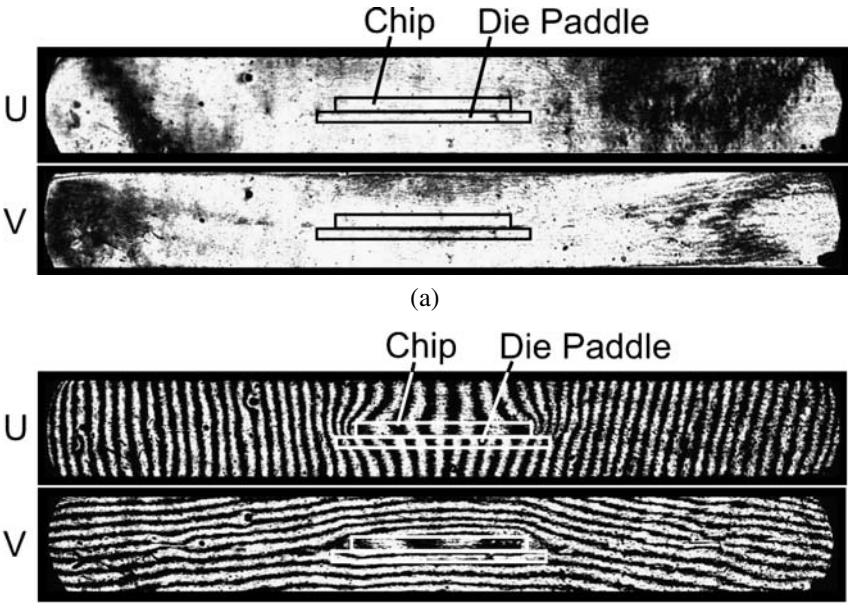


FIGURE 14.20. (a) Null field patterns documented at 85°C before moisture absorption, and (b) fringe patterns induced by cooling the package to 25°C ($\Delta T = -60$°C, zero moisture).

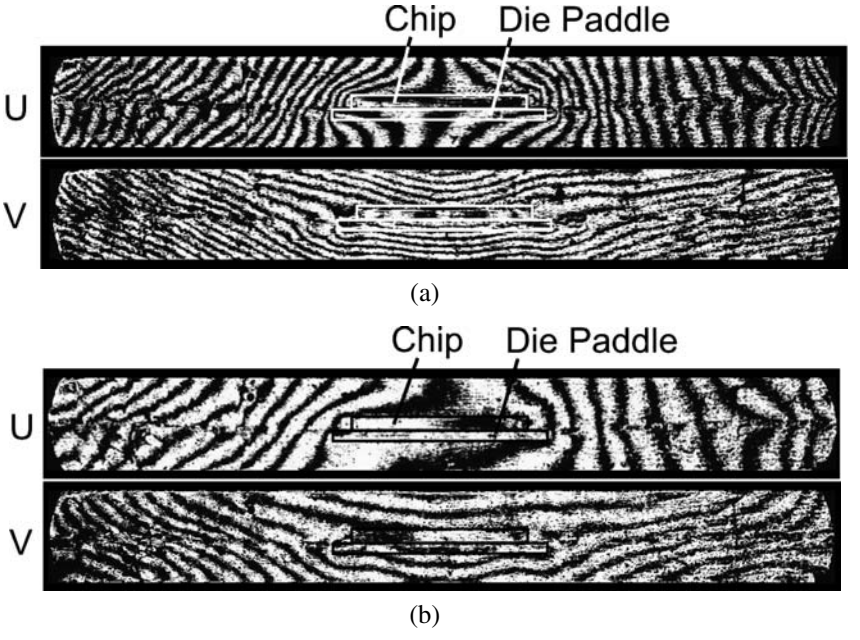


FIGURE 14.21. Fringe patterns obtained at 85°C during desorption process; (a) at its virtual equilibrium state and (b) after 8 hours.

this experiment since it was not practically possible to have two identical package specimens. Instead, the ULE grating mold was used as a reference to tune the interferometer to its initial condition.

The measurements were carried on while the desorption process continued. Representative fringe patterns of the package at a time interval of 8 hours are shown in Figure 14.21(b). A significant contraction of the package is evident; the number of fringes in the package decreased significantly after 8 hours of desorption, proceeding toward the condition of Figure 14.20(a) as the moisture content decreased.

The fringe patterns at the zero hour [Figure 14.21(a)] represent the hygroscopic mismatch deformation at the virtual saturation point. It is important to remember that the measurement was made at the grating replication temperature (85°C), and thus the fringe patterns shown in Figure 14.21 represent deformations induced only by hygroscopic swelling and do not contain any thermally induced deformations.

The displacement fields shown in Figures 14.15 and 14.16 represent the total deformation of the package, which include the free thermal (Figure 14.15) and the free hygroscopic (Figure 14.16) part of the deformation and the stress-induced part of the deformation. Mathematically, the total strain of the package is

$$\begin{aligned} &\textit{For thermal strain: } \varepsilon_\alpha^T = \varepsilon_\alpha^f + \varepsilon_\alpha^\sigma = \alpha \Delta T + \varepsilon_\alpha^\sigma, \\ &\textit{For hygroscopic strain: } \varepsilon_\beta^T = \varepsilon_\beta^f + \varepsilon_\beta^\sigma = \beta C + \varepsilon_\beta^\sigma, \end{aligned} \tag{14.7}$$

where ε^T is the total strain, ε^f is the free expansion/contraction part of strain, ε^σ is the stress-induced part of the strain; the subscript of α and β denotes the cases of thermal deformation and hygroscopic deformation, respectively.

The values of α and β of the mold compound were not known. They were determined from the regions sufficiently far away from the chip (regions marked by dashed boxes in Figure 14.14), where the deformations represent only ε^f of the mold compound. The value of α was determined from the fringe patterns in Figure 14.20(b) and it was 14.4 (ppm/°C). The value of β was determined using the procedure described in the previous section. Fringe patterns at various desorption times were analyzed, and the swelling in these regions was plotted versus moisture content. The maximum moisture content and the CHS value were determined as $C = 0.43\%$ and $\beta = 0.21$ ($\%\varepsilon_h/\%C$), respectively.

The total x direction strains along a line just above the top of the chip were obtained directly from the U field fringe patterns. The corresponding stress-induced strains were then calculated using Equation (14.7), with $\Delta T = -60$°C, and $C = 0.43\%$. The results along the dashed line (AA′ shown in Figure 14.19) are plotted in Figure 14.22(a), where a distance from the center of the package is normalized by the width of the chip. Along the chip/mold compound interface (distance < 0.5), $\varepsilon_\alpha^\sigma$ is tensile while ε_β^σ compressive. The magnitude of the stress-induced strain caused by CHS mismatch, ε_β^σ, is nearly twice as large as that produced by the CTE mismatch, $\varepsilon_\alpha^\sigma$, with $\Delta T = -60$°C. The strains change abruptly around the edge of the chip. These were caused by the material discontinuity. The signs of the strains were reversed but their magnitudes reduced to a uniform value at about half the chip width from the edge of the chip. Far from the chip, the strain was caused by the constraint produced by the copper leadframe.

Another interesting phenomenon was observed in the V displacement fields. The bending displacements along the line were determined from the V field fringe patterns, and they are plotted in Figure 14.22(b). Unlike the strains, the bending displacements have

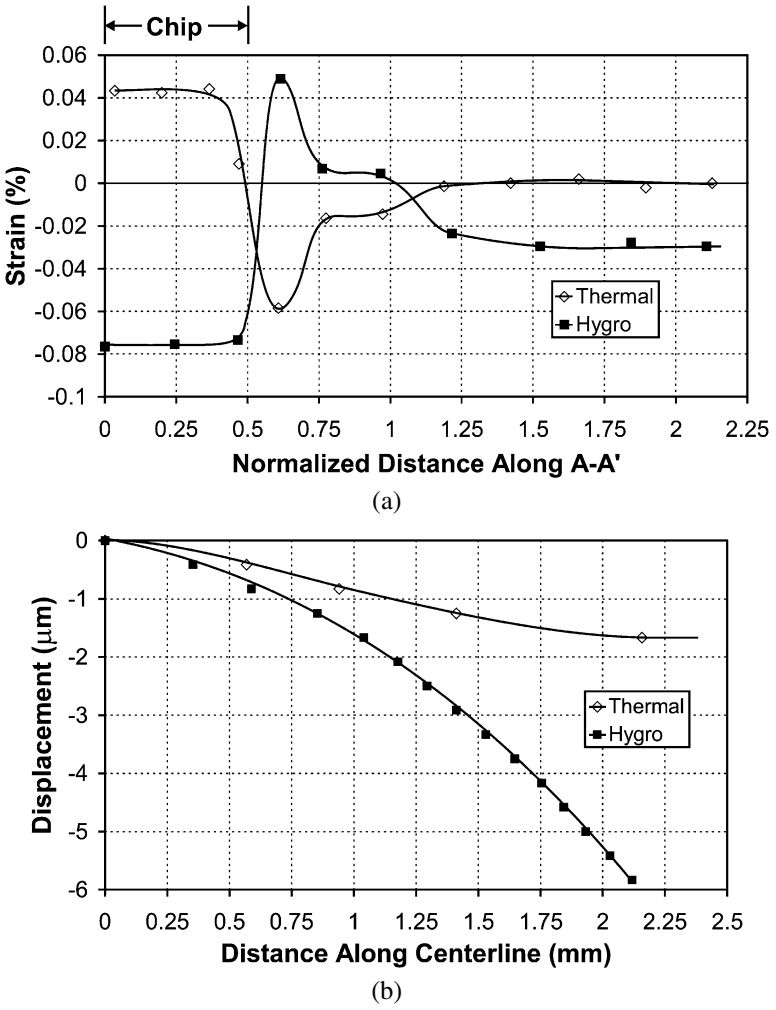


FIGURE 14.22. (a) Stress-induced strains and (b) bending displacements along the chip/mold compound interface.

the same sign. The total bending displacement induced by swelling is nearly three times as large as that caused by thermal deformation. The bending was caused by the fact that the lead-frame was not placed in the middle of the package. The CTE of the copper leadframe (17 ppm/°C) is reasonably close to that of the molding compound (14.4 ppm/°C) and the CTE mismatch was not highly significant. However, the swelling coefficient of the lead-frame is zero and the swelling mismatch caused a larger bending. The chip also contributed to the bending displacement.

It is well known that temperature changes and thermal expansion mismatches can cause stresses and deformations that lead to reliability problems in PEMs. The experimental evidence here indicates that hygroscopic stresses can also have a significant impact on PEM reliability. In fact, this study shows that the hygroscopic swelling induced deformations can be larger than thermally induced deformations in some packages. The analysis to assess reliability of microelectronics devices must include predictive capabilities of hygroscopic swelling if relative humidity is present in the field condition.

14.2.8. Micromechanics

14.2.8.1. Micro Via in Build-up Structure One of several purposes of a chip carrier is to provide conducting paths between the extremely compact circuits on the chip and the more widely spaced terminals on the PCB. Recent micro via technology enabled the industry to produce laminate substrates with high density, and fine pitch conductors as required for advanced assemblies. A cross-sectional view of a high-density organic substrate is illustrated in Figure 14.23. Photosensitive dielectric layers (insulators) are built up sequentially with patterned layers of copper plating (typically 25 μm thick).

Extensive research and development efforts have been and are being made to perfect the underfill process for these organic substrates, and to develop optimum underfill materials for the larger silicon devices. An important trend in newly developed underfill materials is its increased Young's modulus, which increases the coupling between the silicon chip and the substrate. This high degree of coupling transfers the CTE mismatch induced-loading to the build-up layers of the substrate.

Microscopic moiré interferometry was employed to quantify the effect of the underfill on the deformations of the microstructures within the build-up layers. Two specimen configurations were analyzed to study the deformations induced by the subsequent manufacturing process: a bare substrate and a flip chip package. The flip chip assembly is illustrated schematically in Figure 14.24(a) with its relevant dimensions. In the assembly, a silicon chip was attached to a high-density substrate by solder bumps and the gap between the solder bumps was filled with an underfill.

The specimens were cut and ground to expose the desired microstructures as illustrated schematically in Figure 14.24(b), where the insert depicts the detailed microstructures within the build-up layer. An epoxy specimen grating was applied at an elevated temperature of 92°C in a small region containing the microstructures. A tiny volume of the epoxy was applied around the region of interest with a sharp-pointed tool and an ULE grating mold was pressed against the epoxy to spread it into a thin film over the region. The fringes were recorded at a room temperature of 22°C, recording the thermal deformation for $\Delta T = -70°\text{C}$.

The displacement fields for a small region containing the microstructures were recorded by microscopic moiré interferometry. The region is marked by a dashed box in Figure 14.25(a); it is approximately 500 μm by 375 μm. The resultant fringe patterns are shown in Figure 14.25 for (b) the bare substrate and (c) the flip chip assembly. A fringe multiplication factor of $m = 4$ was used to produce a displacement contour interval of 52 nm/fringe.

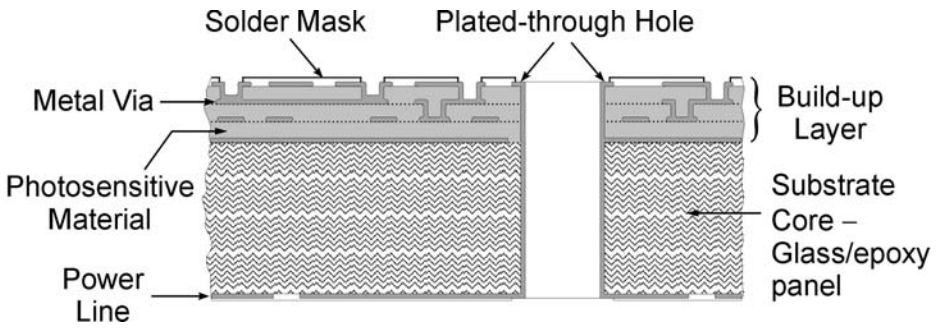


FIGURE 14.23. Schematic diagram of a high density organic substrate with build-up structures.

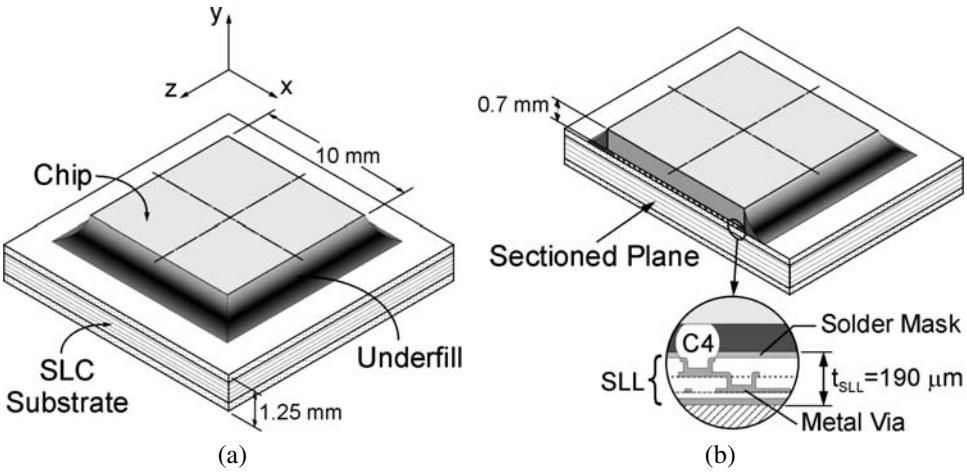


FIGURE 14.24. Schematic diagram of the flip chip assembly on a high-density substrate (a) before and (b) after specimen preparation.

The copper micro vias are embedded in the build-up layer. The CTE of the copper (17 ppm/°C) is much smaller than the CTE of the photosensitive material (>40 ppm/°C). Consequently, the photosensitive material contracts more than the vias during cooling. Since the photosensitive material is confined by the metal via and the adjacent layer, the different expansion rate causes deformations within the metal vias.

The deformation of a small segment of metal via, CC′ (see the insert of Figure 14.26), was analyzed to investigate the effect of the chip and underfill. The deformed shape of the portion CC′ in the flip chip assembly was evaluated from the fringe patterns in Figure 14.25(c) and the results are plotted in Figure 14.26(a).

The center portion of CC′ is connected rigidly to the solder bump/underfill layer while the left and right segments are extended into the photosensitive material. As can be seen from the deformed shape, these segments moved in the positive y direction (upwards) relative to the center portion. This movement produced a shear strain in the segments, which is plotted in Figure 14.26(b). The maximum shear strain occurred near the end of the right segment and its magnitude was 0.38%. The shear strain of the same segment in the bare substrate is also plotted in Figure 14.26(b). The effect of the extra constraint from the solder and underfill layer is evident.

14.2.8.2. Local CTE Variations in High Performance Substrate A high performance substrate accommodates various electrical components such as CPU, capacitors, resisters, and etc. The substrate is a complex composite system, which includes layers of epoxy, woven fiber glass, and copper planes. The substrate bridges high density I/O of CPU and next level structure with less dense I/O using embedded components such as micro via and paths as well as multiple build-up layers. With increasing number of transistors in the CPU, there is subsequent need for miniaturization of 1st lever interconnects, which connect the CPU and the substrate.

For reliability assessment of 1st level interconnects, the global CTE mismatch between the silicon chip (≈3 ppm/°C) and the substrate (≈17 ppm/°C), as well as the local CTE variations should be considered. Microscopic moiré interferometry was employed to determine the local CTE variations around two plated-through-holes (PTH). Figure 14.27

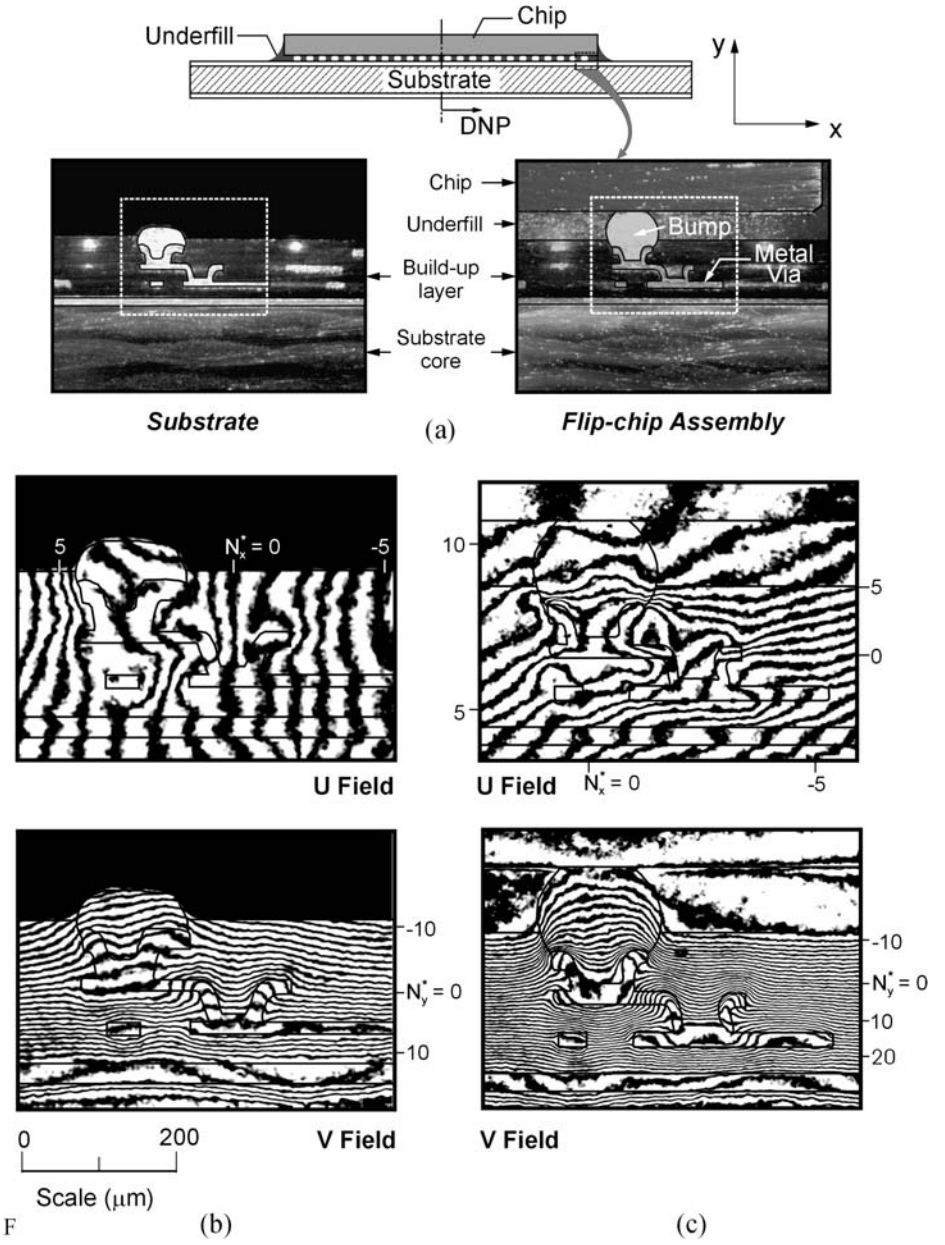


FIGURE 14.25. (a) Micrographs of the region of interest. Microscopic U and V displacement fields of (b) bare substrate and (c) flip chip assembly. The contour interval is 52 nm per fringe.

shows the microscopic U and V displacement fields of the PTH area, induced by ΔT of $-80°$C, where the contour interval is 104 nm/fringe.

The fringe patterns clearly show the homogeneous nature of the plug material inside the PTH, indicated by the uniformly spaced U and V fringes, and the heterogeneous nature of fiber/resin laminated areas located between the PTHs. For the regions marked as A

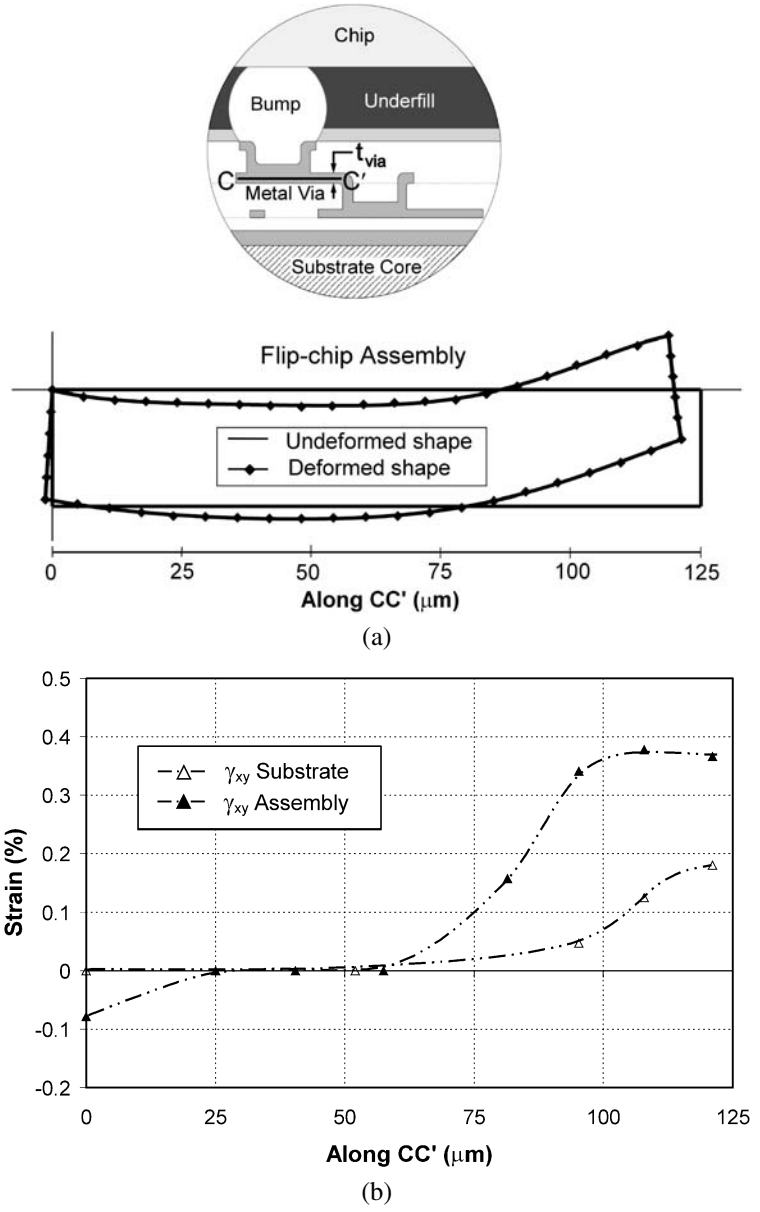


FIGURE 14.26. (a) Deformed shape of the micro via segment CC′ in the flip chip assembly and (b) shear strain distribution along CC′.

and B in Figure 14.27, the effective CTE differ by nearly 8 times in the U field (33.7 and 4.0 ppm/°C), whereas there is only 10% difference in the V field (63.0 and 69.0 ppm/°C). This resulted from variations of materials and features in the substrate. The local fluctuations of CTE around these features are the critical information needed to identify true failure risk areas within the substrate.

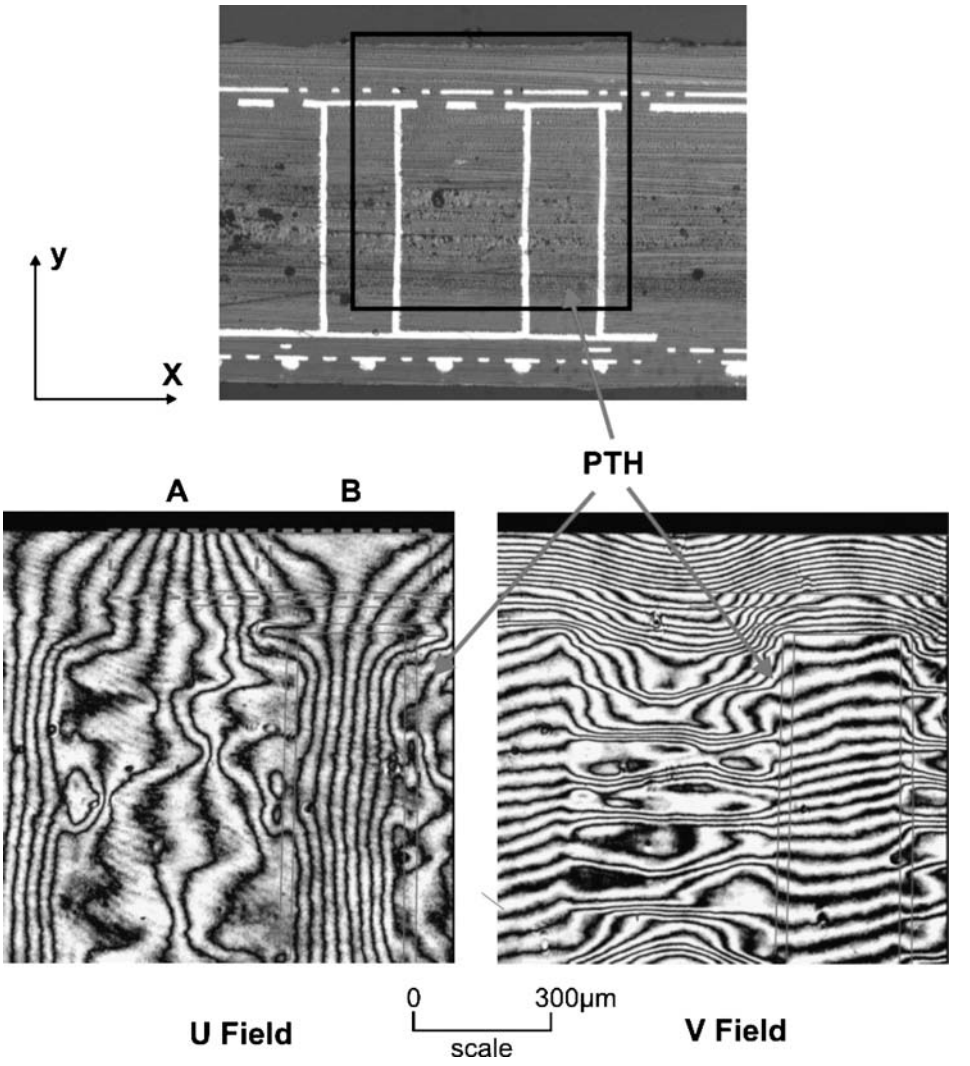


FIGURE 14.27. Local CTE variations around plated-through-hole, induced by ΔT of $-80°$C. The contour interval is 104 nm/fringe (Courtesy of S. Cho, Intel).

14.3. WARPAGE ANALYSIS

14.3.1. Twyman/Green Interferometry

14.3.1.1. Basic Principle Twyman/Green (T/G) interferometry is an optical method, which measures surface contours (out-of-plane displacements) with sub-micron sensitivity [1]. T/G interferometry is simple but it can be used very effectively to determine deflections of silicon wafers or chips. The silicon surface is polished during the manufacturing process and provides a specular (mirror-like) surface, which is a critical requirement for the method. More importantly, the method can be used in conjunction with moiré interferometry to provide a complete set of U, V, W displacement fields since the moiré specimen grating also provides the required specular surface on the specimen.

Figure 14.28 illustrates the optical setup schematically. In the method, an optically flat beam splitter directs half the light to the specimen inside an environmental chamber and the other half to a flat reference mirror. Similar to the real-time moire setup, the specimen holder is not attached to the chamber and it is connected directly to the optical table to avoid vibrations from the environmental chamber. After reflection from the specimen and reference surfaces, the beams meet again at the beam splitter and a portion of each propagates to be collected by an imaging system. The wave front from the specimen, which was originally flat, interferes with the wave front from the reference mirror to produce a contour map of the z coordinate of the specimen surface. The W displacement then can be determined by

$$W(x, y) = \frac{\lambda}{2} N_z(x, y), \tag{14.8}$$

where N_z is the fringe order at each point in the fringe pattern and λ is the wavelength of the laser light employed. When a He-Ne laser of $\lambda = 0.633$ μm is used, the contour interval of the fringe pattern is 0.316 μm per fringe order.

14.3.1.2. Application: Tape Automated Bonding Assembly [13] The package studied by T/G interferometry was a typical chip/heat sink assembly. The package used the tape automated bonding (TAB) technology, where a small active chip (5 mm square for this case) was joined to patterned metal on polymer tape by using thermo-compression bonding. Then, the top surface of the chip was protected by dispensing an encapsulant.

When the subassembly cooled to room temperature after the encapsulant was cured at an elevated temperature, the encapsulant bent the chip, as illustrated schematically in Figure 14.29. The bending was caused by thermal contraction of the encapsulant, which created a large tensile stress on the chip/heat sink adhesive layer. This bending had to be portrayed accurately, since excess tensile stresses caused by the chip bending could produce cracks in the adhesive layer.

The results from the subassemblies with different thicknesses of encapsulant are shown in Figure 14.29. The fringe patterns represent the bending deformation of the back

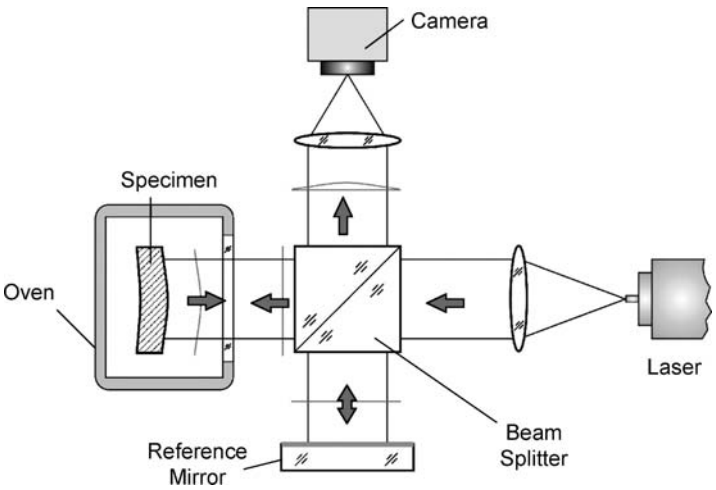


FIGURE 14.28. Schematic diagram of real-time T/G interferometry setup.

side of the chip at room temperature. The corresponding bending displacement was determined from Equation (14.8). The results indicated that the bending displacement was linearly proportional to the thickness of the encapsulant. With a given chip and heat sink material, the tensile stresses of the adhesive layer increased as the thickness of the encapsulant (and the subsequent bending) increased. The results were used to provide a manufacturing specification of the allowable maximum encapsulation thickness.

14.3.1.3. Application: Micro-Opto-Electro-Mechanical system [28] The packaging requirements of the Micro-Opto-Electro-Mechanical system (MOEMS) for video technology are significantly different from those of conventional electronic packages. Digital light switches (micro-mirrors) located inside the package reflect incident light and project a digital image. Therefore, the package must provide a path for incident and reflected light.

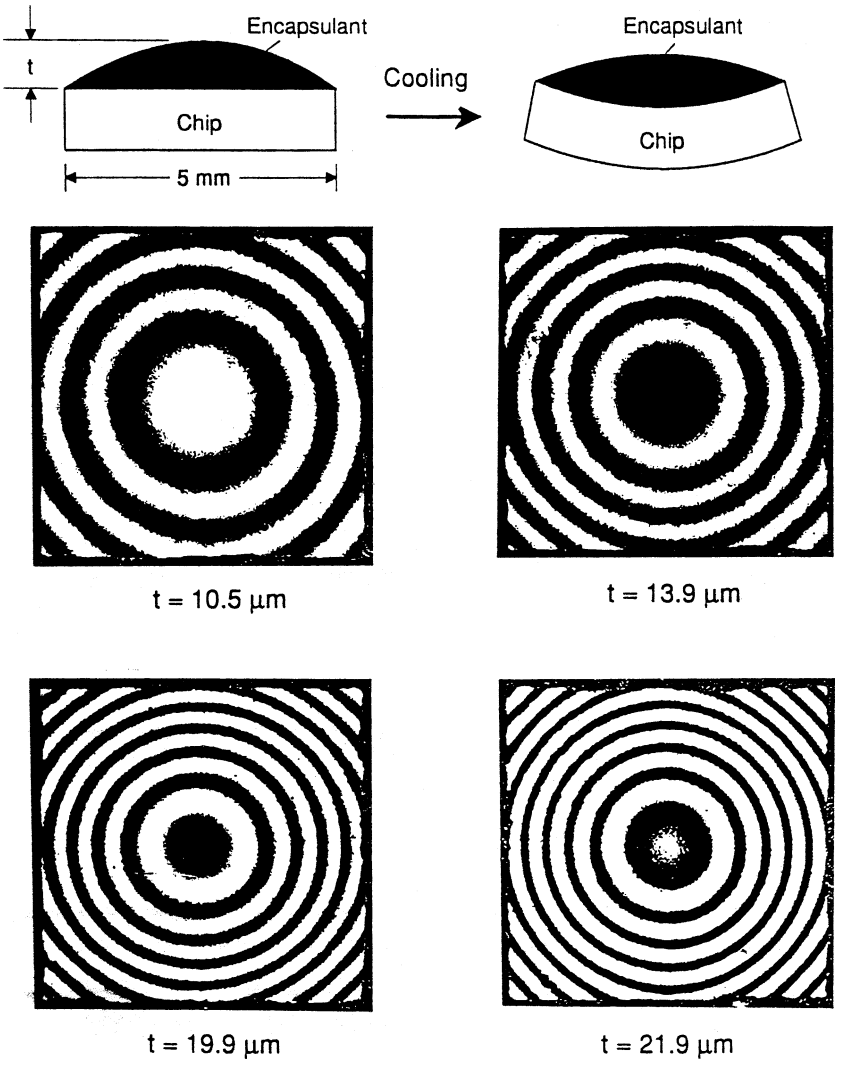


FIGURE 14.29. Surface contours of silicon chip caused by thermal contraction of the encapsulant.

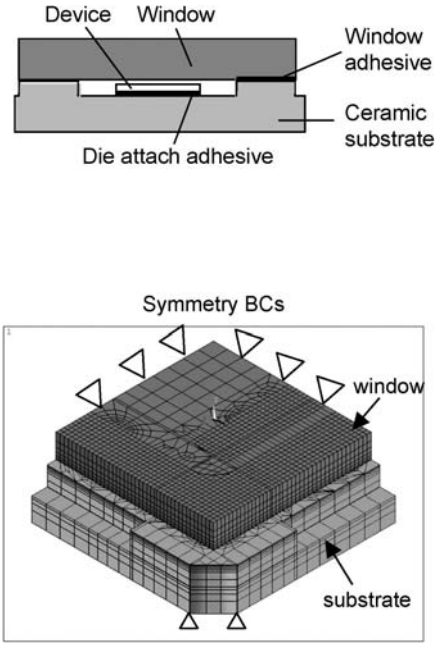


FIGURE 14.30. Cross sectional sketch of DMD™ device and corresponding 3D FEM model using quarter symmetry.

Alignment and focus, which are system optical requirements, require tight tolerances on package flatness and parallelism. Reliable device operation and extended life require a controlled operating environment inside the package. Therefore the package must be sealed from contaminants and moisture. To achieve all these, package structural integrity under a broad range of thermal and mechanical load conditions must be maintained.

Numerical models, such as finite element models (FEM), can be used to understand the integrity of the package during operations and tests. This application focuses on validation of such structural finite element models through Twyman/Green interferometry experiments.

The cross-sectional sketch of a MOEMS package is shown in Figure 14.30. The DMD™ device is attached to an alumina substrate with a heat-conducting adhesive. The die is wire-bonded to the substrate for electrical connectivity. A glass window is bonded to the substrate using an epoxy adhesive. The 3D FEM model is also shown in Figure 14.30.

The warpage of the window in the package was measured using T/G interferometry. The surface of the window was made reflective by depositing a thin aluminum layer. The specimen was then installed in the computer controlled environmental chamber, and the warpage information was documented while heating and cooling the specimen from $-40°C$ to $150°C$. The fringe patterns are shown in Figure 14.31, where only a quadrant of the window surface (defined by the vertical and horizontal centerlines) was viewed due to the symmetry of the package.

The FEM model was analyzed at various temperatures corresponding to the experimental conditions. Initial FEM models had assumed that the stress-free temperature was considered to be $25°C$ since the adhesive cured at room temperature. Models with this assumption had indicated that there would be relatively large window warpage at $150°C$. The experimental results showed an opposite trend; the warpage decreased as the temperature

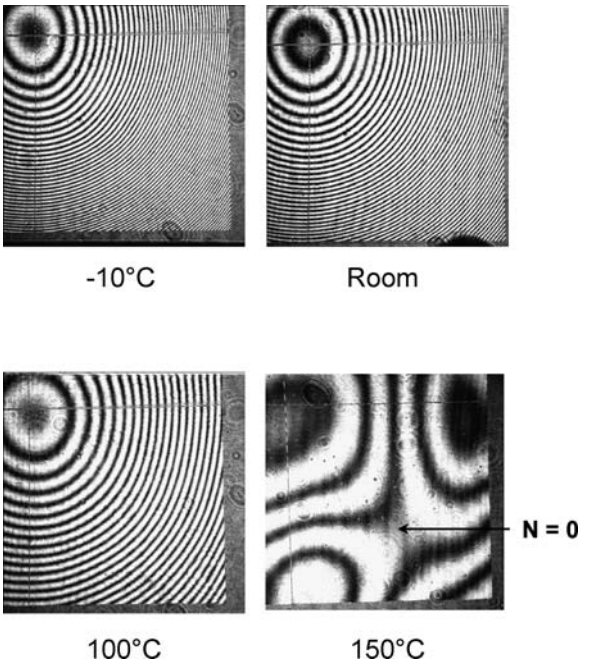


FIGURE 14.31. Surface contours of window as a function of temperature.

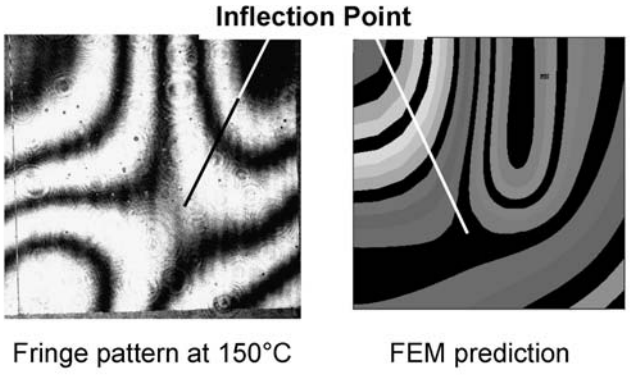
increased. At the maximum temperature, the window became nearly flat and it exhibited an inflection point in the center of the quadrant, as indicated by the zero fringe order in Figure 14.31. During the first thermal cycle, the warpage at room temperature was zero and it increased as the temperature increased. However, the stresses in the package completely relaxed at the glass transition temperature of the epoxy (135°C), and consequently, the stress-free temperature became 135°C after the first cycle.

The corrected stress-free temperature enhanced correlation between the modeling predictions and the experimental data. Yet, the correlation was poor at the temperature beyond 135°C and the inflection point observed from the experiment was not produced by the FEM modeling. This discrepancy was resolved by considering an internal pressure caused by the expansion and contraction of the trapped gases inside the DMD™ package. The internal pressure, which increased with temperature, offset the effect of the thermal expansion mismatch and caused the window to warp convexly when heated above 135°C.

Considering the internal pressure, the modeling predictions show remarkable correlation with the experimental data. As shown in Figure 14.32, the predicted window warpage at 150°C exhibits the inflection point, and the predicted warpages at different temperatures agree with the experimental data as well. The experimental data guided the modeling efforts and led the model into a predictive domain.

14.3.2. Shadow Moiré

14.3.2.1. Basic Principle Shadow moiré provides whole-field maps of out-of-plane displacements but with relatively coarse sensitivity (typically 25 to 50 μm per fringe order). The method does not require any special surface condition. The sensitivity range and relaxed surface requirement make the method ideally suited for the warpage measurement of PCB and chip/organic carrier packages.



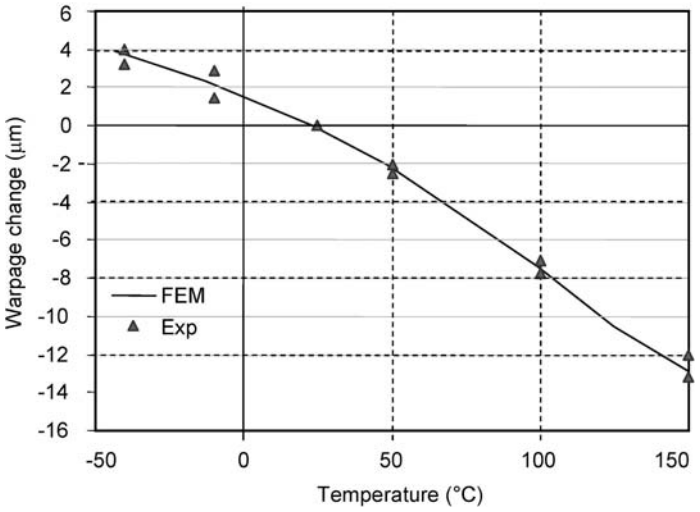


FIGURE 14.32. Comparison between experimental data and modeling prediction after experimental observations were used to guide the FEM analysis.

In the method, a real reference grating is located in front of a specimen and it creates moiré fringes by interacting with its shadow on the specimen. Figure 14.33 illustrates the basic concepts of real-time shadow moiré [17]. A linear reference grating of pitch g is fixed adjacent to the surface, which serves as a window of an environmental chamber. The grating is comprised of black bars and clear spaces on a flat glass plate. A light source illuminates the grating and specimen, and the observer (or camera) receives the light that is scattered in its direction by the matte surface of the specimen.

With the arrangement shown in Figure 14.33 where the source and camera lie at the same distance from the plane of the specimen, the relationship between W and fringe order N_z is

$$W(x, y) = \frac{g}{\tan\alpha + \tan\psi} N_z(x, y) = \frac{gL}{D} N_z(x, y). \tag{14.9}$$

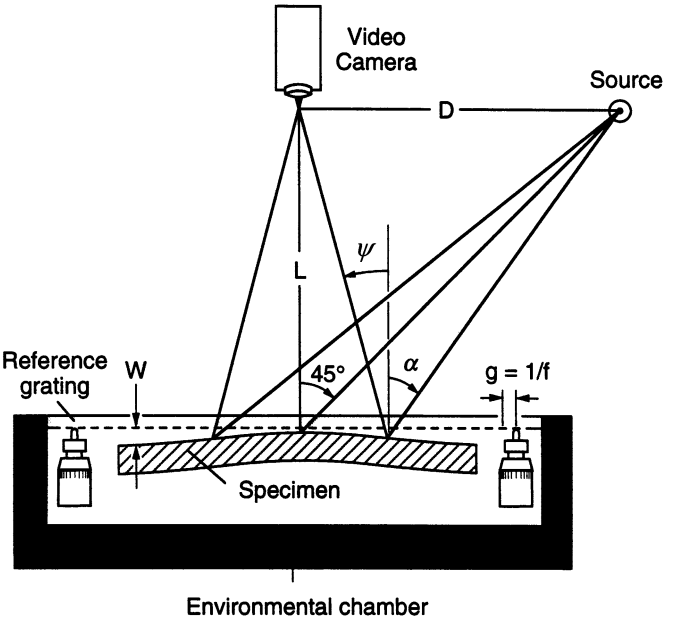


FIGURE 14.33. Schematic diagram of real-time shadow moiré setup.

Although incidence and viewing angles α and ψ, respectively, vary across the field, the sum of their tangents is a constant.

14.3.2.2. Popcorn Effect in Wire Bond Plastic Ball Grid Array Package [17] The WB-PBGA package is not a hermetically sealed package, and thus the plastic mold compound absorbs moisture while the package is in storage. The absorbed moisture is usually concentrated at the chip/substrate interface. When the package is heated rapidly in a reflow oven, the moisture vaporizes and produces a high pressure, which cracks the package; this phenomenon is known as the *popcorn effect*. Real-time shadow moiré was employed to document the effect.

The specimen was a 35 mm square WB-PBGA package. The results obtained from shadow moiré are shown in Figure 14.34, which represent the substrate (bottom side) with a contour interval of 50 μm (2 mil) per fringe. The out-of-plane displacements along a diagonal line were extracted from the fringe patterns and the results are plotted in the figure. At 200°C, a highly concentrated warpage occurred over the chip region. The total warpage in the region was 100 μm and it was ascribed to initial vapor expansion that delaminated the interface between the chip and chip pad. As temperature increased, the pressure developed by vapor increased. As a result, delamination eventually propagated through the mold-compound/substrate interface at 240°C, indicated by the uniformly distributed circular fringes in the pattern.

14.3.2.3. Substrate Topography The warpage of a PBGA package is attributed to a large mismatch of CTE between the chip and substrate. In the subsequent assembly process, electrical and mechanical connections are made by solder balls between the substrate and a PCB. If the bottom side of the substrate warps significantly at the solder reflow temperature as illustrated in Figure 14.35, it yields an uneven height of solder interconnections, which could cause premature failure of the assembly. Detailed knowledge of this out-of-plane

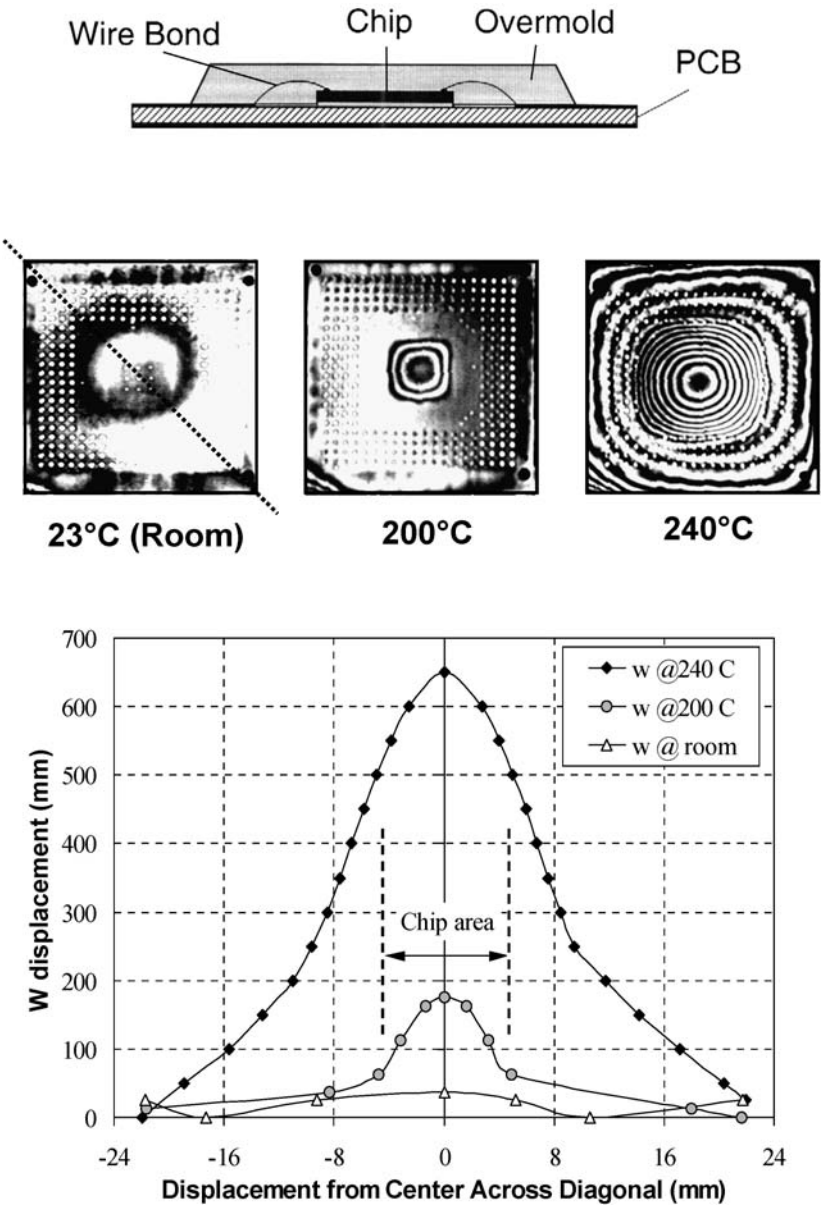
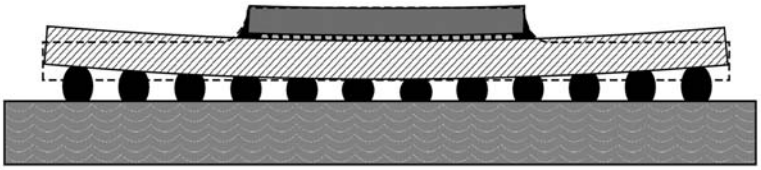


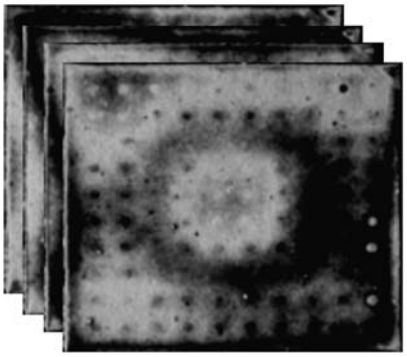
FIGURE 14.34. Warpage of WB-PBGA package caused by *popcorn effect*, where the contour interval is 50 μm per fringe order.

deformation is essential to optimize design and process parameters for reliable assemblies. The package warpage at the solder reflow temperature was documented by shadow moiré.

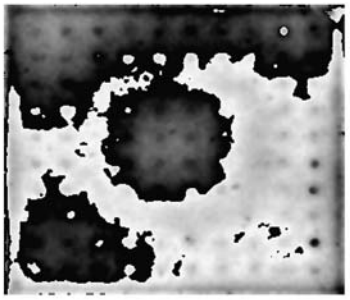
Figure 14.35 shows the shadow moire fringes on the bottom surface of the package. The fringes were processed by the *phase shifting technique*, where a series of fringe or phase shifted interferograms are utilized to compute the phases at every point in the field. Once the phase information is determined, the fringe order is determined at each



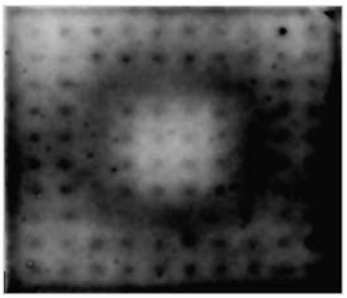
Undesired condition during assembly
(Warped substrate at solder reflow temperature)



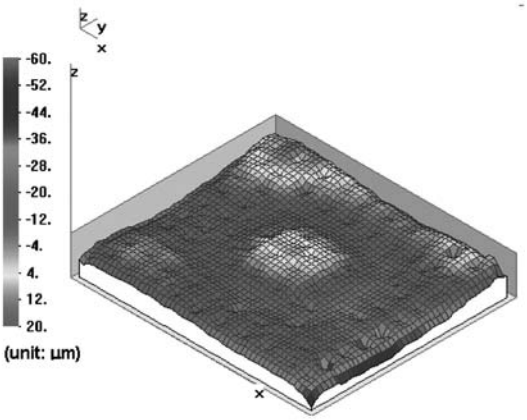
(a) Fringe patterns



(b) wrapped phase



(c) unwrapped phase map



(d) 3-D warpage

FIGURE 14.35. Warpage of the bottom surface of a plastic ball grid array package at the solder reflow temperature.

pixel to within a small fraction of a fringe order, and the displacement field is determined with enhanced accuracy. The displacement field can then be displayed as a contour map, pseudo-color map, or 3D graph. A series of shadow moiré patterns, each phase stepped by $1/4$ of 2π with respect to its neighbor, are shown in Figure 14.35(a). The patterns were processed to produce a wrapped phase map in Figure 14.35(b). Figure 14.35(c) shows the corresponding unwrapped fringe pattern, which represents contours of constant fringe order, or constant W displacement. A three-dimensional representation of the warpage is shown in Figure 14.35(d).

14.3.3. Far Infrared Fizeau Interferometry

14.3.3.1. Basic Principle Although simple, the application of T/G interferometry is limited since it requires a specular (mirror-like) surface. In addition, its measurement sensitivity is a fraction of micron, which is usually too high for the typical warpage observed in flip-chip packages. On the other hand, with shadow moiré, a small distance between the specimen and the reference grating is required for fringe formation; consequently, the chip and the substrate cannot be viewed simultaneously. In addition, shadow moire a diffusive surface for good fringe visibility; this is typically done by spraying a white matte paint and it is not desired for nondestructive testing.

For the FC-PBGA package applications, the measurement sensitivity in the range of microns is ideal. Large tolerance to specimen surface roughness is required to test the ground surface of the chip and the organic substrate without any specimen preparation. The warpage is to be measured as a function of temperature to simulate operating and accelerated testing conditions. These technical requirements were the immediate motivation of development of far infrared Fizeau interferometry (FIFI).

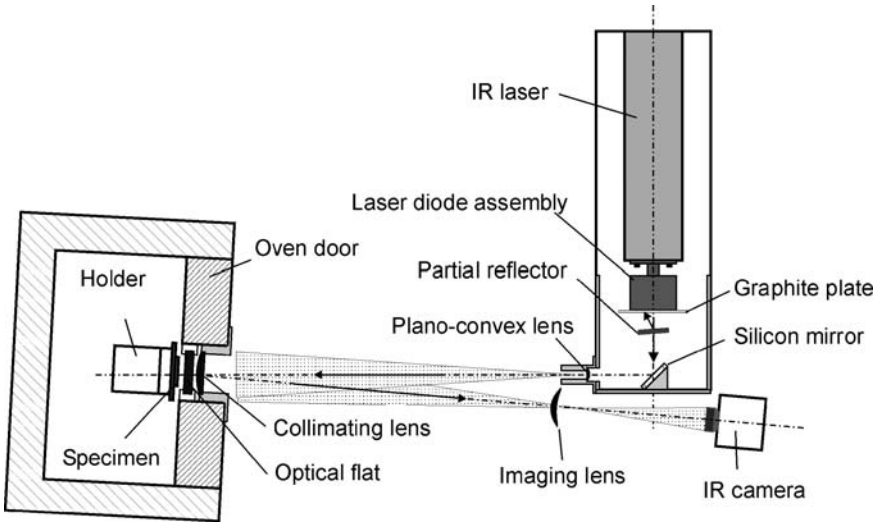
It is known that the specular component of reflected light (mirror-like refection) increases as the wavelength or the angle of incidence increases. Consequently, with a longer wavelength, a surface regarded as optically rough under visible light can be treated as a specular surface. The increase of specular reflection can be explained qualitatively by using the definition of effective roughness, ε_R, known as "Rayleigh criterion"

$$\varepsilon_R = \frac{4\pi h \cos\theta}{\lambda}, \tag{14.10}$$

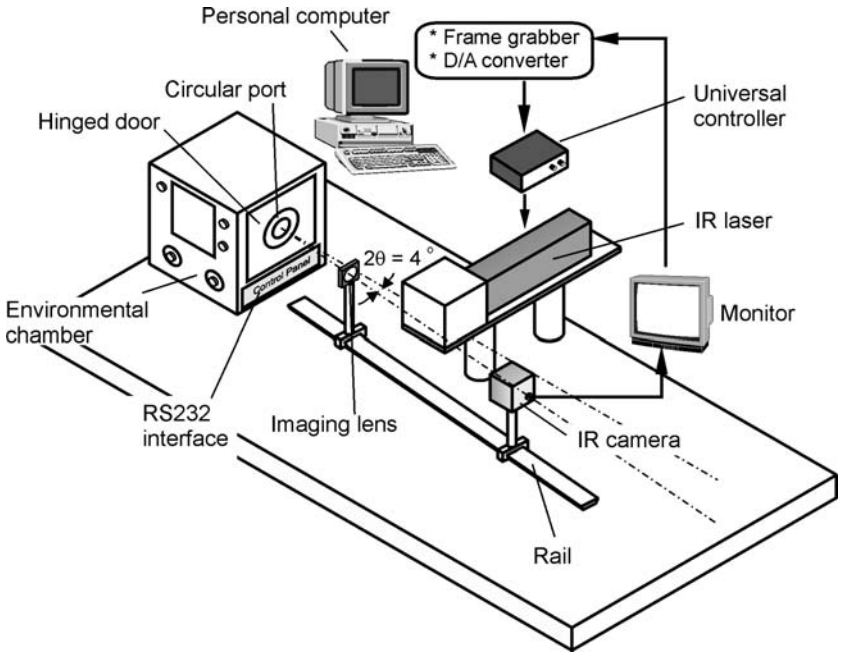
where h is the height of surface irregularities, λ is the wavelength and θ is the angle of incidence. Theoretically, a surface will appear perfectly smooth when h/λ approaches zero or θ approaches 90°.

Fizeau interferometry is a classical method interferometry using visible light, which measures surface topography of a slightly warped specular surface. Far infrared Fizeau interferometry extends the domain of its application by employing light with a long wavelength and thus decreasing the apparent roughness of the specimen surface [46]. Considering $\lambda = 10.6$ μm of CO_2 laser, the apparent roughness is reduced by a factor of 20 for a give angle of incidence, compared with a wavelength in the middle of visible spectrum. Consequently, optically rough surfaces such as the ground surface of silicon, the surfaces of organic substrates, etc., can be tested without any specimen preparation.

The optical configuration of FIFI is illustrated in Figure 14.36(a) [30]. The beam from the laser is first divided by a partial reflector to reduce the intensity of the laser. The transmitted light is expanded and it is subsequently collimated by a collimating lens. An optical flat is placed next to the collimating lens. The optical axis of the collimating lens



(a)



(b)

FIGURE 14.36. (a) Optical and (b) mechanical configuration of FIFI for real-time observation.

is perpendicular to the optical flat and the expanded beam illuminates them with a small angle of incidence. A portion of collimated beam is reflected from the optical flat while the transmitted beam is reflected from the specimen surface. The reflected beams are collected by an infrared CCD camera.

The plane wave front from the optical flat interferes with the wavefront from the specimen to produce a contour map of the z coordinate of the specimen surface. The W displacement can then be determined by

$$W(x, y) = \frac{\lambda}{2\cos\theta} N_z(x, y), \tag{14.11}$$

where θ is the angle of incidence and λ is the wavelength of the laser light employed. With the small angle used in the system ($\cos\theta \approx 1$), the measurement sensitivity is $\lambda/2$ per fringe order (5.3 μm per fringe order).

The mechanical configuration for real-time observation is illustrated in Figure 14.36(b) [30]. An air-cooled CO_2 laser is used as a coherent light source. An expanded beam illuminates the optical flat and the collimator mounted on a specially designed port of an environmental chamber. The imaging system is comprised of an imaging lens and an infrared CCD camera. The imaging lens is mounted on a translation stage that travels along a rail to provide the desired magnification factor for various sizes of specimens. The image is displayed on a monitor and the output from the monitor is digitized by a frame grabber for image processing.

14.3.3.2. Characterization of Mechanical Design of a FC-PBGA Package Assembly [31] It is common practice (but not universal practice) to mount the FC-PBGA package on a larger PCB, which provides electrical pathways to and from the device. As illustrated in Figure 14.4, connections to the PCB are made by an array of larger solder balls.

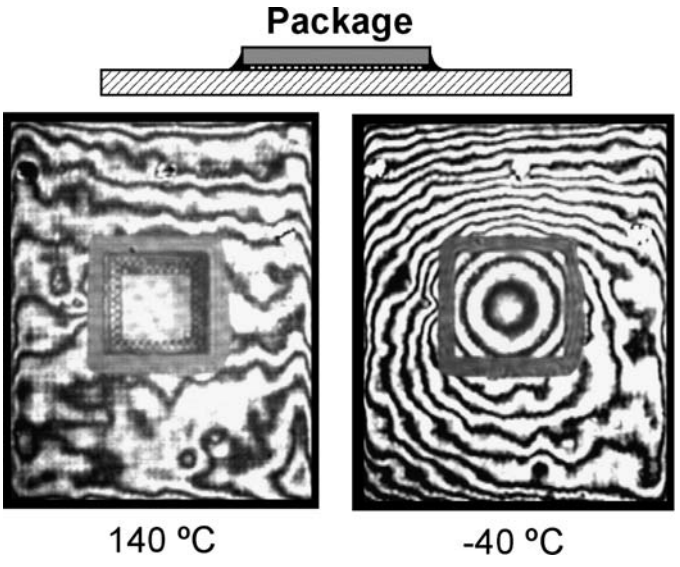
The FIFI system was employed to reveal the deformation mechanism of a FC-PBGA assembly. Deformation of the package was compared to the assembly for identical thermal loading. In the package, a silicon chip (6.9 mm × 6.1 mm × 0.76 mm) was attached to an organic substrate (1.02 mm thick) through eutectic solder bumps. The overall dimensions of the package body were 21 mm by 25 mm, which provided 304 I/O connections. The package was surface-mounted on an FR4 PCB using eutectic solder balls to form the assembly.

The thermal loading was applied to the two samples in separate experiments. In each case, they were initially heated until the stress-free (or zero chip bending) condition was achieved. Then, the deformations of the chip side were recorded while cooling the specimens to −40°C. The fringe patterns obtained at the stress-free condition and −40°C are shown in Figure 14.37(a) and (b) for the package and the assembly, respectively.

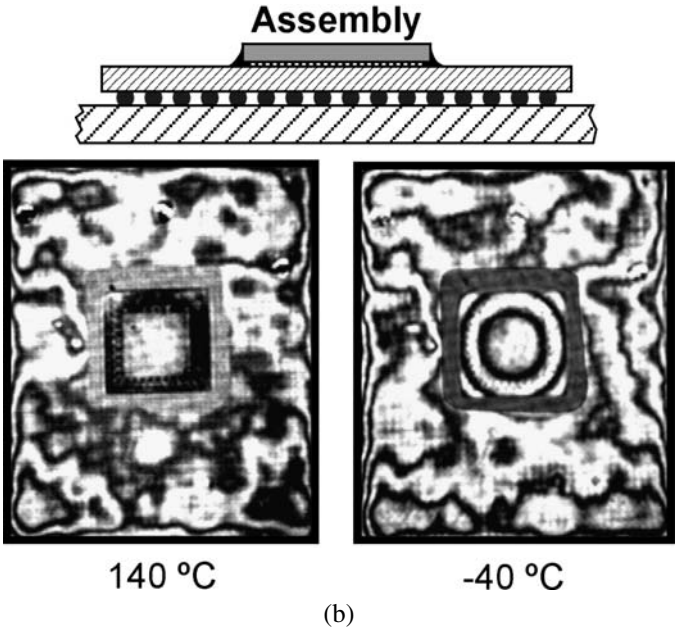
As expected, warpage of the chip increased significantly with cooling for both the package and the assembly. For the package, the warpage of the substrate also increased as the temperature decreased [Figure 14.37(a)]. For the assembly, however, the warpage of the substrate was nearly zero at the stress-free condition and it remained unchanged during the entire cooling process [Figure 14.37(b)].

The changes of warpage along the vertical centerline were extracted from the patterns for both specimens and the results are plotted in Figure 14.38. In the package specimen, the substrate under the chip (i.e., the *chip shadow portion*) was constrained through the underfill layer and it presumably experienced the same bending deformation as the chip. This conclusion is supported by the shape of the rest of the substrate (i.e., the *bare substrate portion*), which extended beyond the edges of the chip with essentially the same slope as the chip edges.

In the assembly specimen, the PCB constrained the bare substrate portion through the solder balls. The effect of the PCB constraint in the assembly is evident, as indicated by an inflection point at the end of the chip shadow portion (Figure 14.38).



(a)



(b)

FIGURE 14.37. Warpage contours of the chip side obtained by far infrared Fizeau interferometry; (a) flip chip package and (b) its assembly. The contour interval is 5.31 μm per fringe.

The substrate was connected to the PCB through the solder balls and the difference of curvature between the substrate and the PCB was accommodated in the chip shadow portion by *tensile* deformation of the solder balls. In the bare substrate portion, the differential contraction of the substrate and PCB was accommodated by *shear* deformation of the solder balls.

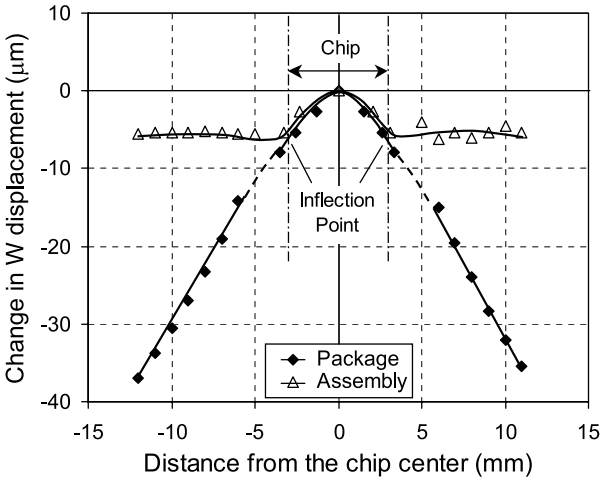


FIGURE 14.38. Changes in chip and substrate warpage along the vertical centerline, obtained from patterns in Figure 14.37.

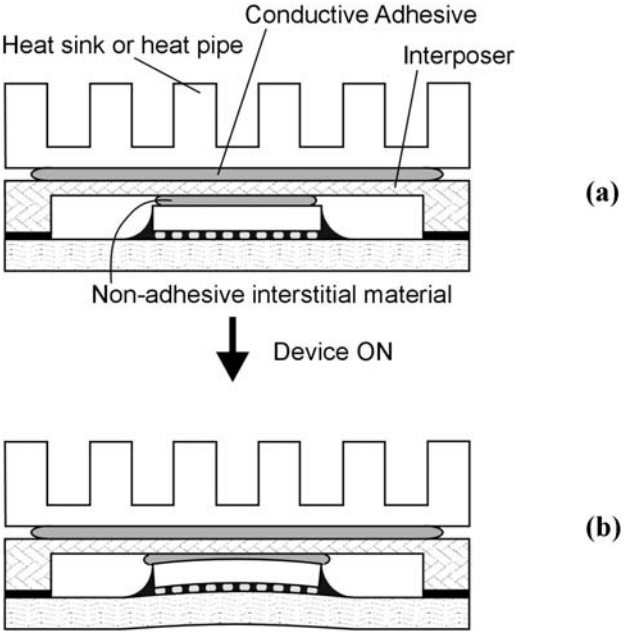


FIGURE 14.39. (a) Advanced thermal management solution uniquely available for flip chip packages and (b) illustration of the effect of warpage on the thermal interface.

These results guided the subsequent numerical parametric study, which quantified the effect of various substrate parameters on the solder ball strains [31]. The study indicated that the substrate CTE was one of the most critical design parameters for solder joint reliability. The experimental evidence provided by the FIFI system was essential to revealing this important design parameter.

14.3.3.3. Characterization of Thermal Design of FC-PBGA [47] Figure 14.39 illustrates an advanced thermal solution for a FC-PBGA package that has become important as power densities rise due to miniaturization of active elements within the chip. With the flip-chip package, the top surface of the chip is available for enhanced thermal management solutions. In the configuration shown in Figure 14.39(a), called *interposer* or *cap configuration*, a metal interposer is present between the chip and the heat sink that allows use of non-adhesive type interstitial materials with high thermal conductivity. These materials include thermal grease, thermal gel, elastomeric gasket, phase change material, etc.

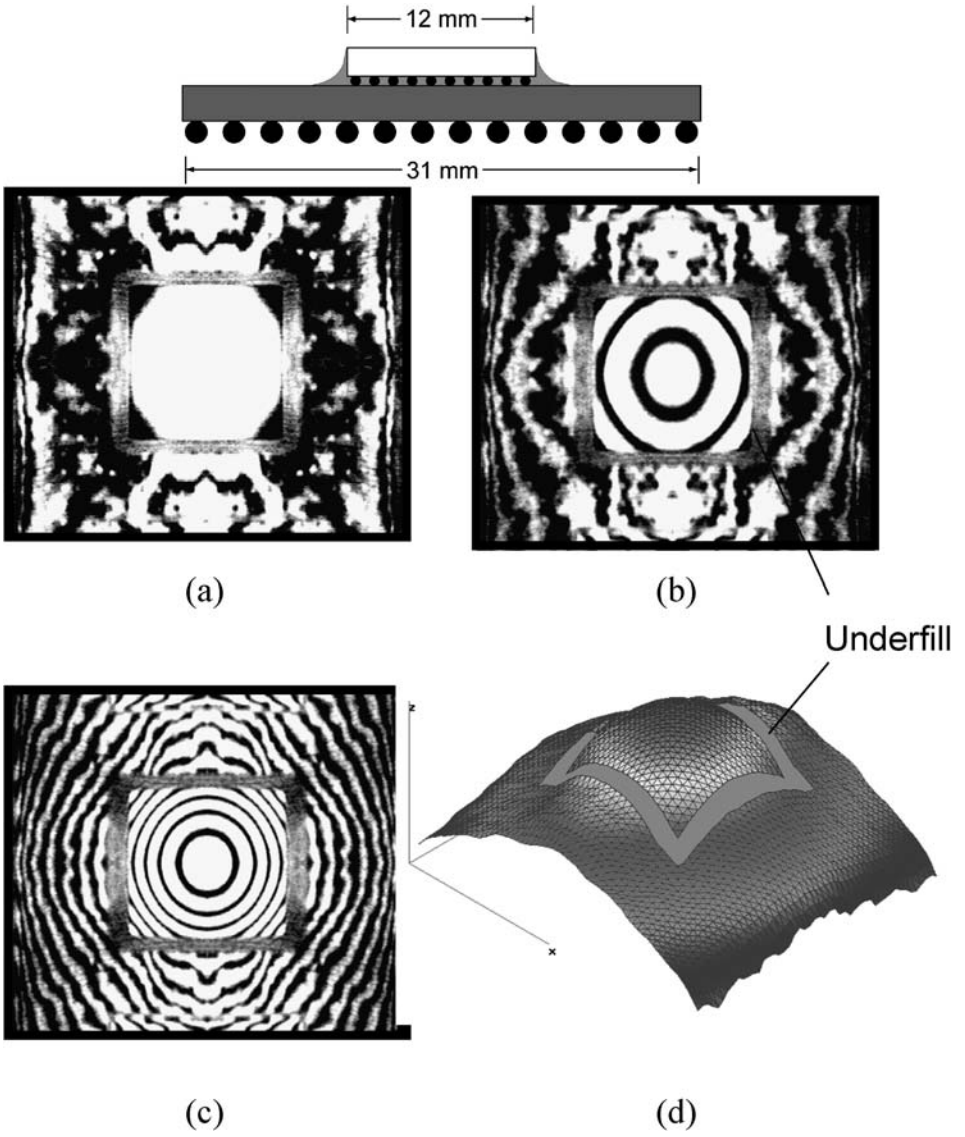


FIGURE 14.40. Warpage contours of FC-PBGA package documented at (a) 150°C, (b) 100°C and (c) room temperature, where the contour interval is 5.3 μm per fringe order. A 3D warpage map at room temperature obtained by a digital image processing is shown in (d).

Note that there is no stiff connection to a PCB in this case. In addition, the adhesive connection between the interposer and substrate has low stiffness. Accordingly, flexure of the substrate is essentially unconstrained. When the chip is powered and the device is subjected to a temperature excursion, the gap between the chip and the interposer changes due to warpage of the chip and the substrate. The gap changes again when the device is turned off and the package returns to its original shape. If a non-solid interstitial material such as thermal grease or gel is used, the material is gradually squeezed out during the repeated power cycling and eventually causes significant degradation of thermal performance; this phenomenon is known as *pump-out* and it is illustrated in Figure 14.39(b).

The *pump-out* phenomenon was evaluated by the FIFI system by testing the package illustrated in Figure 14.40. A square chip (12 mm $\times$ 12 mm) was mounted on an organic substrate (31 mm $\times$ 31 mm). Initially, the package was heated to the underfill curing temperature (150°C). Then, the deformation of the top surface of the package was recorded while the package was cooled to room temperature. The fringe patterns obtained at 150°C, 100°C and room temperature (22°C) are shown in Figure 14.40(a–c), where the contour interval is 5.31 μm per fringe order. The package was nearly flat (devoid of fringes) at the underfill curing temperature, but the large mismatch in CTE caused the package to warp as the temperature decreased.

Assuming the interposer is connected to the substrate, the change in the gap can be determined from the fringe patterns. It is 45 μm for a temperature change from 100°C to room temperature. The 3D shape of the upper surface of the chip and substrate can be determined from the fringe patterns, as illustrated in Figure 14.40(d), providing data for the change of volume between the chip and interposer. This change reduces the original volume of thermal grease or gel significantly and it is an essential consideration in the design.

If a solid interstitial material such as an elastomeric gasket or a phase change material is used at the thermal interface, warpage causes a nonuniform pressure distribution in the material, which causes non-uniform thermal conductance [48]. Thus, for both non-solid and solid materials, the warpage information produced by the FIFI system is critical to the mechanical design of the package.

ACKNOWLEDGMENT

The author wishes to thank all the co-authors of the cited references for their contributions.

REFERENCES

1. D. Post, B. Han, and P. Ifju, High Sensitivity Moiré: Experimental Analysis for Mechanics and Materials, Springer-Verlag, New York, 1994.
2. G. Cloud, Optical Methods of Engineering Analysis, Cambridge University Press, New York, 1995.
3. P. Rastogi, Ed., Photomechanics for Engineers, Springer-Verlag, New York, 2000.
4. B. Han, Thermal stresses in microelectronics subassemblies: quantitative characterization using photomechanics methods, Journal of Thermal Stresses, 26, pp. 583–613 (2003).
5. A.F. Bastawros and A.S. Voloshin, Transient thermal strain measurements in electronic packages, IEEE Transactions on Components, Hybrids and Manufacturing Technology, 13(4), pp. 961–966 (1990).
6. A.F. Bastawros and A.S. Voloshin, In situ calibration of stress chips, IEEE Transactions on Components, Hybrids and Manufacturing Technology, 13(4), pp. 888–892 (1990).

7. A.F. Bastawros and A.S. Voloshin, Thermal strain measurements in electronic packages through fractional fringe moiré interferometry, Journal of Electronic Packaging, Transaction of the ASME, 112(4), pp. 303–308 (1990).
8. Y. Guo, W.T. Chen, and C.K. Lim, Experimental determination of thermal strains in semiconductor packaging using moiré interferometry, Proceedings of 1992 Joint ASME/JSME Conference on Electronic Packaging, American Society of Mechanical Engineers, New York, 1992, pp. 779–784.
9. Y. Guo, W.T. Chen, C.K. Lim, and C.G. Woychik, Solder ball connect (SBC) assemblies under thermal loading: I. Deformation measurement via moiré interferometry, and its interpretation, IBM Journal of Research and Development, 37(5), pp. 635–648 (1993).
10. T.Y. Wu, Y. Guo, and W.T. Chen, Thermal-mechanical strain characterization for printed wiring boards, IBM Journal of Research and Development, 37(5), pp. 621–634 (1993).
11. B. Han and Y. Guo, Thermal deformation analysis of various electronic packaging products by moiré and microscopic moiré interferometry, Journal of Electronic Packaging, Transaction of the ASME, 117, pp. 185–191 (1995).
12. P.H. Tsao and A.S. Voloshin, Manufacturing stresses in the die due to die-attach process, IEEE Transactions on Components, Packaging and Manufacturing Technology, Part A, 18(1), pp. 201–205 (1995).
13. B. Han, Y. Guo, C.K. Lim, and D. Caletka, Verification of numerical models used in microelectronics packaging design by interferometric displacement measurement methods, Journal of Electronic Packaging, Transaction of the ASME, 118, pp. 157–163 (1996).
14. B. Han and Y. Guo, Determination of effective coefficient of thermal expansion of electronic packaging components: a whole-field approach, IEEE Transactions on Components, Packaging and Manufacturing Technology, Part A, 19(2), pp. 240–247 (1996).
15. B. Han, M. Chopra, S. Park, L. Li, and K. Verma, Effect of substrate CTE on solder ball reliability of flip-chip PBGA package assembly, Journal of Surface Mount Technology, 9, pp. 43–52 (1996).
16. B. Han, Deformation mechanism of two-phase solder column interconnections under highly accelerated thermal cycling condition: an experimental study, Journal of Electronic Packaging, Transaction of the ASME, 119, pp. 189–196 (1997).
17. B. Han and Y. Guo, Photomechanics tools as applied to electronic packaging product development, in B. Han, R. Mahajan and D. Barker, Eds., Experimental/Numerical Mechanics in Electronics Packaging, Vol. 1, Society for Experimental Mechanics, Bethel, CT, 1997, pp. 11–15.
18. J.-H. Zhao, X. Dai, and P.S. Ho, Analysis and modeling verification for thermal-mechanical deformation in flip-chip packages, Proceedings of 48th Electronic Components and Technology Conference, May 25–28, Seattle, WA, 1998, pp. 336–344.
19. A.S. Voloshin, P.H. Tsao, and R.A. Pearson, In situ evaluation of residual stresses in an organic die-attach adhesive, Journal of Electronic Packaging, Transaction of the ASME, 120(3), pp. 314–318 (1998).
20. B. Han, Recent advancement of moiré and microscopic moiré interferometry for thermal deformation analyses of microelectronics devices, Experimental Mechanics, 38(4), pp. 278–288 (1998).
21. B. Han, Z. Wu, and S. Cho, Measurement of thermal expansion coefficient of flexible substrate by moiré interferometry, Experimental Techniques, 25(3), pp. 22–25 (2001).
22. S.-M. Cho, S.Y. Cho, and B. Han, Observing real-time thermal deformations in electronic packaging, Experimental Techniques, 26(3), pp. 25–29 (2002).
23. S.-M. Cho, B. Han, and J. Joo, Temperature dependent deformation analysis of ball grid array package assembly under accelerated thermal cycling condition, Journal of Electronic Packaging, Transaction of the ASME, 126, pp. 41–47 (2004).
24. E. Stellrecht, B. Han, and M. Pecht, Measurement of the hygroscopic swelling coefficient in mold compounds using moiré interferometry, Experimental Techniques, 27(4), pp. 40–44 (2003).
25. E. Stellrecht, B. Han, and M. Pecht, Characterization of hygroscopic swelling behavior of mold compounds and plastic packages, IEEE Transactions on Components and Packaging Technologies, 27(3), pp. 499–506 (2004).
26. B. Han and P. Kunthong, Micro-mechanical deformation analysis of surface laminar circuit in organic flip-chip package: an experimental study, Journal of Electronic Packaging, Transaction of the ASME, 122(3), pp. 294–300 (2000).
27. S. Cho and B. Han, Effect of underfill on flip-chip solder bumps: an experimental study by microscopic moiré interferometry, The International Journal of Microcircuits and Electronic Packaging, 24(3), pp. 217–239 (2001).
28. M. Variyam and B. Han, DMD module calibration using interferometry, TI Technical Journal, pp. 1–8 (2001).
29. M.Y. Tsai, C.H. Jeter Hsu, and C.T. Otto Wang, Investigation of thermo-mechanical behaviors of flip chip BGA packages during manufacturing process and thermal cycling, IEEE Transactions on Components and Packaging Technologies, 27(3), pp. 568–576 (2004).

30. K. Verma and B. Han, Warpage measurement on dielectric rough surfaces of microelectronics devices by far infrared fizeau interferometry, Journal of Electronic Packaging, Transaction of the ASME, 122(3), pp. 227–232 (2000).
31. K. Verma, B. Han, S.-B. Park, and W. Ackerman, On the design parameters of flip-chip PBGA package assembly for optimum solder ball reliability, IEEE Transactions on Components and Packaging, 24(2), pp. 300–307 (2001).
32. C.-P. Yeh, K. Banerjee, T. Martin, C. Umeagukwu, and R. Fulton, Experimental and analytical investigation of thermally induced warpage for printed wiring boards, Proceedings of the 41th ECTC, Okland, CA, May 1991.
33. T. Martin, C.-P. Yeh, and C. Ume, Measurement of thermally induced warpage in printed wiring boards, Proceedings of the 1991 ASME Winter Annual Meeting, AMD-Vol. 131/EEP-Vol. 1, December 1991.
34. B. Han, Y. Guo, and H.-C. Choi, Out-of-plane displacement measurement of printed circuit board by shadow moiré with variable sensitivity, Proceedings of the 1993 ASME International Electronics Packaging Conference, Binghamton, NY, September 1993.
35. C.-P. Yeh, I.C. Ume, R.E. Fulton, K.W. Wyatt, and J.W. Stafford, Correlation of analytical and experimental approaches to determine thermally induced PWB, IEEE Transactions on Components, Hybrids and Manufacturing Technology, 16(8), pp. 986–995 (1993).
36. Y. Guo, Applications of shadow moiré method in determination of thermal deformations in electronic packaging, Proceedings of the 1995 SEM Spring Conference, Grand Rapids, MI, 1995.
37. M.R. Stiteler, I.C. Ume, and B. Leutz, In-process board warpage measurement in a lab scale wave soldering oven, IEEE Transactions on Components, Packaging, and Manufacturing Technology, Part A, 19(4), pp. 562–569 (1996).
38. D.B. Rao and M. Prakash, Effect of substrate warpage on the second level assembly of advanced plastic ball grid array (PBGA) packages, Proceedings of the 21st IEEE International Electronics Manufacturing Technology (IEMT) Symposium, Oct. 13–15, Austin, TX, 1997, pp. 439–446.
39. K.S. Chen, T.Y. Chen, C.-C. Chuang, and I.K. Lin, Full-field wafer level thin film stress measurement by phase-stepping shadow moire, IEEE Transactions on Components and Packaging Technologies, 27(3), pp. 594–601 (2004).
40. B. Han and D. Post, Immersion interferometer for microscopic moiré interferometry, Experimental Mechanics, 32(1), pp. 38–41 (1992).
41. B. Han, Higher sensitivity moiré interferometry for micromechanics studies, Optical Engineering, 31(7), pp. 1517–1526 (1992).
42. B. Han, Interferometric methods with enhanced sensitivity by optical/digital fringe multiplication, Applied Optics, 32(25), pp. 4713–4718 (1993).
43. B. Han, D. Post, and P. Ifju, Moiré interferometry for engineering mechanics: current practice and future development, Journal of Strain Analysis, 36(1), pp. 101–117 (2001).
44. D. Post and J. Wood, Determination of thermal strains by moiré interferometry, Experimental Mechanics, 29(3), pp. 318–322 (1989).
45. J.S. Corbin, Finite element analysis for solder ball connect (SBC) structural design optimization, IBM Journal of Research and Development, 37(5), pp. 585–596 (1993).
46. K. Verma and B. Han, Far infrared fizeau interferometry, Applied Optics, 40(28), pp. 4981–4987 (2001).
47. B. Han, Optical measurement of flip-chip package warpage and its effect on thermal interfaces, Electronics Cooling, 9(1), pp. 10–16 (2003).
48. E.E. Marotta and B. Han, Thermal control of interfaces with compliant interstitial materials for microelectronics packaging, Proceedings of 1998 MRS Annual Spring Meeting, San Francisco, CA, April 1998.

15

Analysis of Reliability of IC Packages Using the Fracture Mechanics Approach

Andrew A.O. Tay

Department of Mechanical Engineering, National University of Singapore, 9 Engineering Drive 1, Singapore 117576, Republic of Singapore

15.1. INTRODUCTION

A plastic IC package is a structure consisting of different materials. Figure 15.1 illustrates the structure of a typical plastic-encapsulated quad flat package. The package is typically fabricated by first attaching a semiconductor chip (also called die) onto the pad of a leadframe using an organic die-attach adhesive. Following curing of the die-attach adhesive and wirebonding, the assembly is encapsulated in a transfer molding process during which it is partially cured for a few minutes in the mold. This is followed by a longer period of post-mold curing to fully cure the encapsulant. Before the package is shipped to the customer, it has to be qualified for mechanical reliability, moisture sensitivity and pass other electrical functional tests. In the moisture qualification process the package is first subjected to moisture preconditioning at various levels of temperature and relative humidity e.g., 85°C/85%RH (Level 1), 85°C/60%RH (Level 2), 30°C/60%RH (Level 3), etc. before being subjected to solder reflow. After the solder reflow process, the package is then examined for delamination or cracking using a Scanning Acoustic Microscope (SAM).

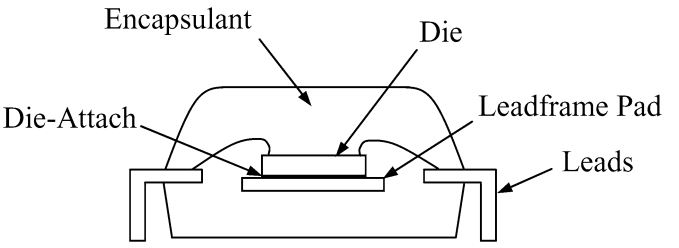


FIGURE 15.1. Typical structure of a plastic IC package.

TABLE 15.1.
Typical properties of packaging materials.

Material	Young's modulus, E (GPa)	Poisson's ratio, ν	CTE (ppm/K)
Encapsulant	12.25/1.23*	0.25/0.37*	17/68**
Die-attach	0.255	0.37	75
Silicon	165	0.25	2.6
Copper	115	0.35	17
Alloy 42	147	0.3	5

*at 20°C/at 200°C.
**below T_g/above T_g, $T_g = 160$°C.

One of the most commonly observed mechanical failure of plastic-encapsulated IC packages is that of "popcorn cracking." In 1985, Fukuzawa et al. [1] reported that many plastic packages which were preconditioned with moisture cracked with a "pop" sound when they were passed through a solder reflow oven. However, those packages which were dried did not crack. They postulated that the cause of this failure was due to the build-up of pressure at the pad/encapsulant interface caused by the evaporation of moisture in the package that had been absorbed in the encapsulant during the moisture preconditioning process.

Since the discovery of the popcorn failure mechanism, the finite element method has been applied to calculate the stress distribution within the plastic package in attempting to predict popcorning failure using a maximum stress criterion. When the plastic package is cured typically at 175°C, it may be assumed that the stress within the package is zero as the encapsulant is only fully hardened at this curing temperature. However, when the package is then cooled to room temperature or later raised to a high temperature during solder reflow, stresses will develop within the package owing to differences in the coefficient of thermal expansion (CTE) of the different constituent materials. Table 15.1 gives typical values of CTE together with the other mechanical properties of common packaging materials.

Hence at the maximum solder reflow temperature, say 215°C for eutectic solder and 260°C for leadfree solder, the mismatch in the CTE values of the constituent materials will give rise to thermal stresses. It is postulated that if the stress in the encapsulant exceeds a maximum value, called the rupture stress, package cracking will occur. The rupture stress may be determined by a simple tensile or 3-point bending test on a strip of encapsulant material. However, it was soon realized that owing to the existence of stress singularities at the corners within plastic packages, the maximum stress is theoretically infinite. Consequently, the maximum stress determined from finite element analysis is mesh dependent. When the mesh adjacent to the corner is refined, the value of the maximum stress calculated is increased *ad infinitum*. This being the case, the maximum stress criterion for the structural failure of a package becomes very difficult to implement as it is impossible to determine the true maximum stress in the encapsulant.

To overcome the above difficulty, Tay et al. [2,3] have proposed the use of the fracture mechanics approach to analyzing reliability of plastic packages. They calculated the stress intensity factors at the corners of the encapsulants formed with leadframe pads and showed that when these exceeded critical values measured using lead-pull tests, mechanical failure during solder reflow would occur. With this criterion, they have evolved a method of determining suitable mechanical properties of molding compounds to avoid popcorn failure.

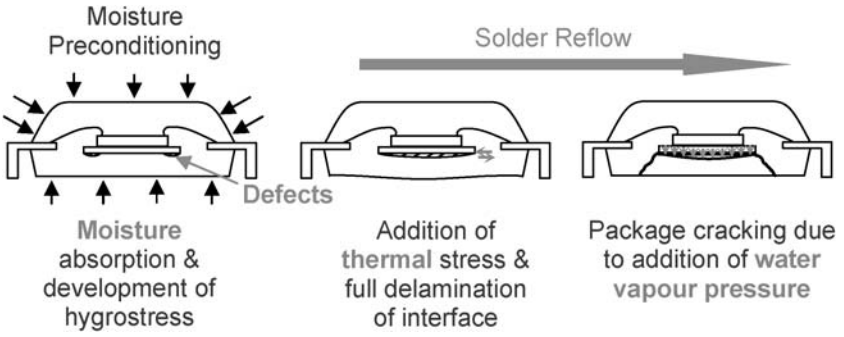


FIGURE 15.2. Mechanism of popcorning failure.

Later Tay and Lin [4] showed that delamination of the pad/encapsulant interface is necessary before popcorning failure can occur. Once the pad/encapsulant interface has delaminated, the structure becomes greatly weakened with the thin layers of encapsulants that are characteristic of packages nowadays. Hence they argued that the more important thing to predict the mechanical reliability of packages during solder reflow is to determine the delamination temperature. They also showed that during solder reflow, hygrostresses which arise due to moisture absorption by the plastic encapsulant add on to thermal stresses induced by CTE mismatch, exacerbating the problem. Tay and Lin then proposed the following mechanism for popcorn failure. With reference to Figure 15.2, moisture diffuses into the package during moisture preconditioning causing the encapsulant to swell but not the die or pad. This induces stress within the package which may be called hygrostress. During solder reflow, thermal stress is induced due to CTE mismatch. Due to manufacturing imperfections, defects in the form of very small delaminations (interfacial cracks) are present at interfaces within the package. Due to the high stress concentration in the vicinity of corners, very high stress intensity factors (SIFs) are induced at the tips of the interfacial cracks at the edge of the pad/encapsulant interface. If the combined hygrothermal SIF exceeds the fracture toughness of the interface, delamination will occur. Very often this delamination will propagate over the entire interface. Subsequently, the vapor pressure will press on the layer of encapsulant below the pad causing the SIF at the corner of the encapsulant to rise. If the toughness of the encapsulant is exceeded, a crack will propagate through the encapsulant to the external surface of the package.

15.2. HEAT TRANSFER AND MOISTURE DIFFUSION IN IC PACKAGES

Heat transfer during moisture preconditioning and solder reflow is governed by the following equation:

$$\frac{\partial T}{\partial \tau} = k'\left(\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2}\right), \tag{15.1}$$

where k' is the thermal diffusivity, τ the time and x, y the Cartesian coordinates. The governing equation for moisture diffusion is given by

$$\frac{\partial C}{\partial \tau} = D(T)\left(\frac{\partial^2 C}{\partial x^2} + \frac{\partial^2 C}{\partial y^2}\right), \tag{15.2}$$

where C is the moisture concentration and $D(T)$ is the coefficient of moisture diffusion given by

$$D(T) = D_0 \exp\left(-\frac{E_D}{RT}\right), \tag{15.3}$$

where D_0 is the permeability index; E_D is the activation energy of the diffusion process; $R = 8.31$ J/mol · K is the gas constant and T is the absolute temperature.

The solubility s can be described by the Arrhenius equation:

$$s = s_0 \exp(-E_s/RT). \tag{15.4}$$

Values of s_0 for a typical encapsulant are $s_0 = 4.96 \times 10^{-7}$ mg/mm^3 MPa and $E_s = -3.87 \times 10^4$ J/mol [5].

Details of the simulation of heat transfer and moisture diffusion are given in reference [6]. It has been found that the diffusion of heat is much faster than that of moisture. The temperature within the package becomes uniform within a minute while the moisture distribution may take hundreds of hours to become uniform within the package. Hence, the variation of D with temperature during the initial few minutes of moisture preconditioning can be ignored and a constant value of D corresponding to the environmental temperature can be assumed during the whole preconditioning process. A typical distribution of moisture distribution immediately after Level 1 moisture precondition is shown in Figure 15.3.

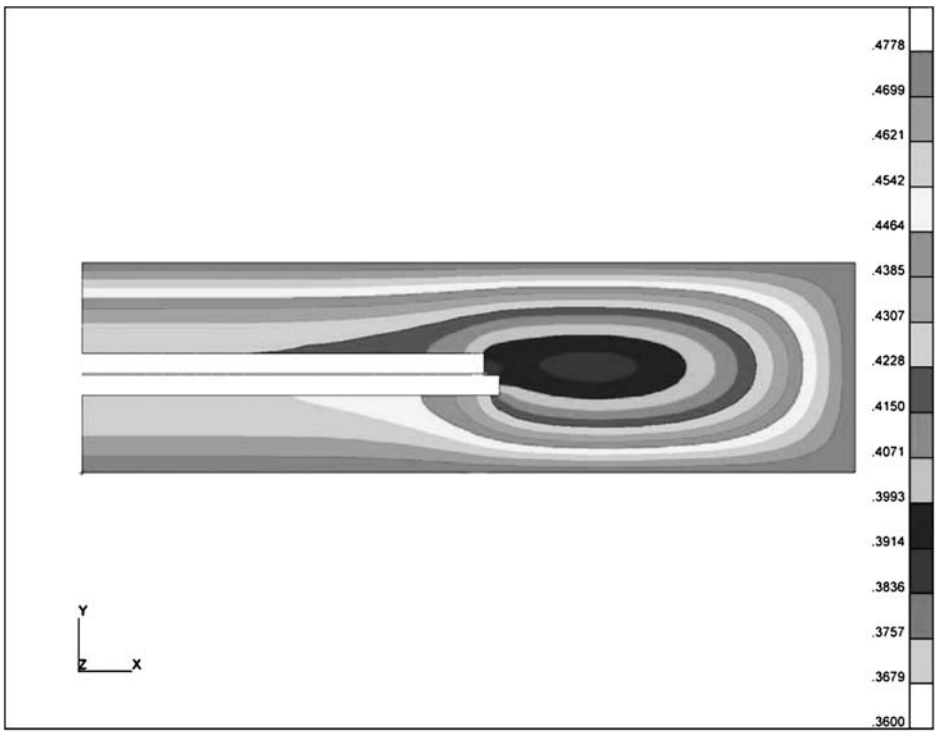


FIGURE 15.3. Distribution of moisture concentration (wt%) after 85°C/85%RH preconditioning for 168 hours.

15.3. FUNDAMENTALS OF INTERFACIAL FRACTURE MECHANICS

The interfacial crack problem has been investigated for many decades and details of solutions to the problem can be found in [8–13]. Only a brief summary will be given in the following. Consider a traction-free crack of length $2a$ at the interface between two half planes of homogeneous, isotropic and elastic materials, with material 1 above the interface and material 2 below as shown in Figure 15.4. It can be shown that the stress distribution in the vicinity of a crack tip is given by

$$\sigma_{yy} + i\tau_{xy} = \boldsymbol{K}(2\pi r)^{-1/2}(r/2a)^{i\varepsilon}, \tag{15.5}$$

where σ_{yy}, τ_{xy} are the normal and shear stresses in the vicinity of the crack tip, $\boldsymbol{K}$ is the complex stress intensity factor given by

$$\boldsymbol{K} = (K_{\mathrm{I}} + iK_{\mathrm{II}}), \tag{15.6}$$

where K_{I} and K_{II} are the opening and shearing mode components, and

$$\varepsilon = \frac{1}{2\pi}\ln\left[\left(\frac{\kappa_1}{\mu_1} + \frac{1}{\mu_2}\right)\Big/\left(\frac{\kappa_2}{\mu_2} + \frac{1}{\mu_1}\right)\right] \tag{15.7}$$

is the bielastic constant, where μ_j, E_j, and ν_j $(j = 1, 2)$ are the shear modulus, Young's modulus, and Poisson's ratio of the respective materials, and $\kappa_j = 3 - 4\nu_j$ for plane strain and $\kappa_j = (3 - \nu_j)/(1 + \nu_j)$ for plane stress.

The crack surface displacements behind the crack tip can be shown to be

$$\delta u = \delta u_y + i\delta u_x = \frac{8(K_{\mathrm{I}} + iK_{\mathrm{II}})}{(1 + 2i\varepsilon)\cosh(\pi\varepsilon)E^*}\left(\frac{r}{2\pi}\right)^{\frac{1}{2}}\left(\frac{r}{2a}\right)^{i\varepsilon}, \tag{15.8}$$

where r is the distance from crack tip and $\delta u_x, \delta u_y$ are the crack surface displacements which are respectively parallel and perpendicular to the interface (see Figure 15.5), and

$$\frac{1}{E^*} = \frac{1}{2}\left(\frac{1}{\overline{E}_1} + \frac{1}{\overline{E}_2}\right), \tag{15.9}$$

where $\overline{E}_j = E_j/(1 - \nu_j^2)$ for plane strain and $\overline{E}_j = E_j$ for plane stress.

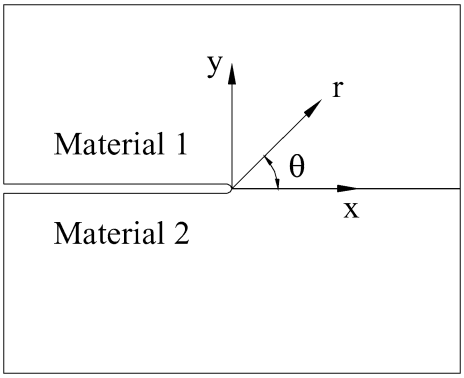


FIGURE 15.4. A crack along a bimaterial interface.

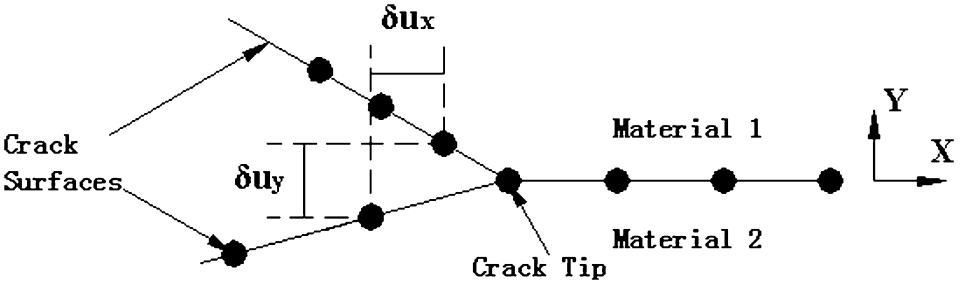


FIGURE 15.5. Crack surface displacements.

It can be shown from Equations (15.5) and (15.8) that for $\varepsilon \neq 0$, the stresses becomes oscillatory and the crack surfaces overlap for very small r, that is, at points very close to the crack tip. Although this is physically inadmissible, it has been pointed out that the oscillatory region can be ignored, since it is confined to an extremely small region around the crack tip. It is interesting to note that in the numerical simulations, it is sometimes found that the crack surfaces interpenetrate into each other. Hence, it is essential that contact surface elements be defined at the crack surfaces in order to ensure that interpenetration of crack surfaces does not occur.

From Equation (15.8) the stress intensity factors can be obtained as

$$K_{\mathrm{I}} = [A\cos(\varepsilon \ln r) + B\sin(\varepsilon \ln r)]/D, \tag{15.10}$$

$$K_{\mathrm{II}} = [B\cos(\varepsilon \ln r) - A\sin(\varepsilon \ln r)]/D, \tag{15.11}$$

where

$$A = \delta u_y - 2\varepsilon \delta u_x, \qquad B = \delta u_x + 2\varepsilon \delta u_y, \tag{15.12}$$

and

$$D = \frac{8}{\cosh(\pi\varepsilon)} \cdot \frac{1}{E^*} \cdot \left(\frac{r}{2\pi}\right)^{\frac{1}{2}}. \tag{15.13}$$

It can also be shown from Equation (15.8) that the magnitude K of the complex stress intensity factor $\boldsymbol{K}$ for the crack is given by

$$K = \sqrt{K_{\mathrm{I}}^2 + K_{\mathrm{II}}^2} = \lim_{r\to 0} \sqrt{1+4\varepsilon^2} \cdot \sqrt{\delta u_x^2 + \delta u_y^2}/D, \tag{15.14}$$

and the mode mixity by

$$\frac{K_{\mathrm{II}}}{K_{\mathrm{I}}} = \tan\psi, \tag{15.15}$$

where the mode mixity phase angle

$$\psi = \lim_{r\to 0}\left[\tan^{-1}\left(\frac{\delta u_x + 2\varepsilon\delta u_y}{\delta u_y - 2\varepsilon\delta u_x}\right) - \varepsilon \ln\left(\frac{r}{2a}\right)\right]. \tag{15.16}$$

15.4. CRITERION FOR CRACK PROPAGATION

It has been found that for a given bimaterial interface, the crack will propagate whenever the magnitude K of the complex stress intensity factor exceeds a critical value K_c called the interface fracture toughness which is a property of the interface. Now, instead of analyzing the crack problem in terms of stress and SIF as described in the previous section, an energy approach can also be used [14,15]. Using the energy approach, it can be shown [16] that the crack driving force is the strain energy release rate (SERR) G which is related to K by

$$G = \frac{(1-\beta_D^2)}{E^*} K^2, \tag{15.17}$$

where

$$\beta_D = \frac{\mu_1(\kappa_2 - 1) - \mu_2(\kappa_1 - 1)}{\mu_1(\kappa_2 + 1) + \mu_2(\kappa_1 + 1)}. \tag{15.18}$$

Since G and K are related by Equation (15.17), an equivalent criterion for crack propagation is when G for a crack exceeds a critical value G_c, also called the interface fracture toughness which is a property of the interface.

15.5. INTERFACE FRACTURE TOUGHNESS

Interface toughness is defined as the critical value of G (or K) required for crack propagation along the interface. This critical value G_c has to be measured. A typical variation of G_c with the mode mixity phase angle ψ is illustrated in Figure 15.6. It can be seen that G_c has a minimum at $\psi = 0°$ (pure mode I) and a maximum at $\psi = 90°$ (pure mode II). For interfaces between plastic encapsulants and leadframe materials, there is a further complication that the interface toughness is also a function of temperature and moisture content. In order to predict the onset of delamination at pad-encapsulant interfaces, data on the variation of interface toughness G_c with temperature T, moisture concentration C and mode mixity ψ, must be available. Unfortunately, such data are scarce in the literature, but are presently being accumulated [17–21]. Tay et al. [21] were the first to characterize the toughness of the interface between copper leadframe and a typical mold compound as a

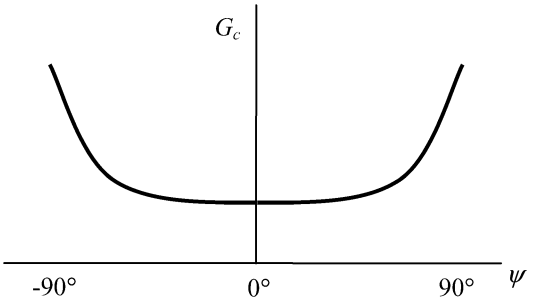


FIGURE 15.6. Typical variation of G_c with ψ.

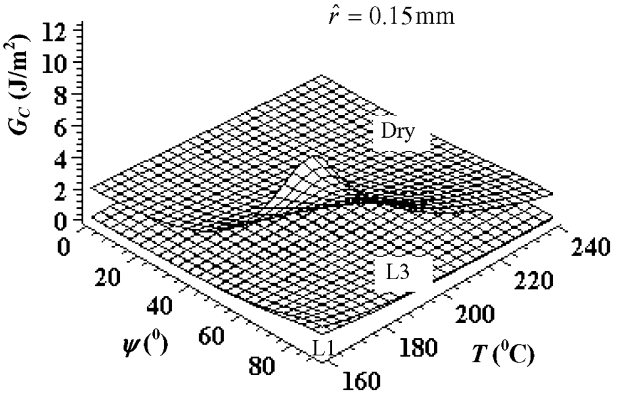


FIGURE 15.7. Variation of G_c with T, C and ψ.

comprehensive function of temperature, moisture and mode mixity. Their results are shown in Figure 15.7 and given in Equations (15.19)–(15.21):

(a) For dry samples ($C = 0$):

$$G_C(\psi, T) = \frac{4.826 - 0.01715T}{1 + (0.0032T - 1.354)\sin^2\psi}. \tag{15.19}$$

(b) For samples subjected to L3 (30°C/60%RH) preconditioning ($C = 0.0153$ mg/mm^3):

$$G_C(\psi, T) = \frac{0.4476 - 0.00147T}{1 + (0.00375T - 1.48)\sin^2\psi}. \tag{15.20}$$

(c) For samples subjected to L1 (85°C/85%RH) preconditioning ($C = 0.0344$ mg/mm^3):

$$G_C(\psi, T) = \frac{0.5112 - 0.0019T}{1 + (0.0028T - 1.286)\sin^2\psi}. \tag{15.21}$$

15.6. TOTAL STRESS INTENSITY FACTOR

It is reasonable to assume that when a plastic IC package is cured at the curing temperature T_{cure} it is in a state of zero stress. When the temperature T within the package is subsequently changed during solder reflow, thermal strains $\varepsilon_t = \alpha(T - T_{cure})$, and hence thermal stresses, will be induced in the package, where α is the coefficient of thermal expansion (CTE). For metals and silicon, the CTE is found to be constant. However, for plastic encapsulants the CTE is found to be a constant below the glass transition temperature T_g of the encapsulant and also constant above T_g but at a higher value. Due to the thermal stresses in the package a thermal SIF $\boldsymbol{K}_t$ will be developed at the crack tip where

$$\boldsymbol{K}_t = K_{\mathrm{I},t} + \boldsymbol{i}K_{\mathrm{II},t}. \tag{15.22}$$

When the plastic encapsulant absorbs moisture, it swells and hygrostrains $\varepsilon_h = \beta C$, and hence hygrostresses, will be induced in the package where β is the coefficient of hygro expansion (CHE), which is assumed to be constant, and C the moisture concentration. Due to the hygrostresses, a hygro SIF $\boldsymbol{K}_h$ will be induced at the crack tip where

$$\boldsymbol{K}_h = K_{\mathrm{I},h} + \boldsymbol{i} K_{\mathrm{II},h}. \tag{15.23}$$

Furthermore, if water vapor pressure develops in the delaminated (cracked) region, a SIF $\boldsymbol{K}_p$ due to the pressure will be developed at the crack tip given by

$$\boldsymbol{K}_p = K_{\mathrm{I},p} + \boldsymbol{i} K_{\mathrm{II},p}. \tag{15.24}$$

Now in a package undergoing solder reflow, it is possible that all three types of stresses, thermal, hygro and pressure-induced stresses, may exist simultaneously. In this situation, one must combine the effects of all three stresses into a total composite SIF. Since we are assuming linear elastic fracture mechanics, the principle of superposition may be employed to obtain the total SIF arising from the combined loading as

$$\boldsymbol{K}_{tot} = (K_{\mathrm{I},t} + K_{\mathrm{I},h} + K_{\mathrm{I},p}) + \boldsymbol{i}(K_{\mathrm{II},t} + K_{\mathrm{II},h} + K_{\mathrm{II},p}), \tag{15.25}$$

$$K_{tot} = \sqrt{(K_{\mathrm{I},t} + K_{\mathrm{I},h} + K_{\mathrm{I},p})^2 + (K_{\mathrm{II},t} + K_{\mathrm{II},h} + K_{\mathrm{II},p})^2}. \tag{15.26}$$

15.7. CALCULATION OF SERR AND MODE MIXITY

There are several numerical methods available for calculating the strain energy release rate (SERR) G and the phase angle ψ of the mode mixity. Several of these will be described in the following.

15.7.1. Crack Surface Displacement Extrapolation Method

A Crack Surface Displacement Extrapolation Method (CSDEM) was proposed by Yuuki and Cho [12]. Using Equations (15.13) and (15.16), values of K and $K_{\mathrm{II}}/K_{\mathrm{I}}$ are calculated and plotted against r. It will usually be found that the points can be well-fitted by a straight line, except for a few points nearest to the crack tip which should be disregarded. Thus K can be obtained by extrapolating the best straight line to $r = 0$. It will be found that when $K_{\mathrm{II}}/K_{\mathrm{I}}$ or ψ is plotted against r, a virtually constant value is obtained.

However, it has been shown that, in some cases, this extrapolation procedure may not yield accurate results under certain conditions, even for homogeneous cracks [22]. This is because, in some cases, when K_{I} and K_{II} calculated using the crack surface displacements are plotted against r, neither a linear variation with r nor a constant is observed. In this situation, extrapolating the curve to obtain K_{I} and K_{II} at $r = 0$ is bound to incur some uncertainty. A more certain procedure for the determination of the SIF K of the bimaterial interface crack has recently been proposed by Hu and Tay [23]. Based on a new second-order analysis of the interface crack, a parameter A is used to engender a K calculated from the crack surface displacements which remained constant for a considerable region around the crack tip. A modified crack surface displacement extrapolation method (MCSDEM) is then developed for obtaining K and the mode mixity $K_{\mathrm{II}}/K_{\mathrm{I}}$ of the crack. However, based on the examples Hu and Tay studied, it was found that the CSDEM did give an accurate value for bimaterial cracks.

15.7.2. Modified J-integral Method

Rice [24] first proposed the expression of an integral, called the J-integral, along an arbitrary contour path Γ around the crack tip (Figure 15.8) in a homogeneous material in the absence of body forces such as thermal stress:

$$J_\Gamma = \int_\Gamma \left[W n_1 - \sigma_{ij} \frac{\mathrm{d}u_i}{\mathrm{d}x_1} n_j \right] \mathrm{d}\Gamma, \tag{15.27}$$

where W is the total strain-energy density defined as

$$W = \int_0^{\varepsilon_{ij}} \sigma_{ij} \mathrm{d}\varepsilon_{ij}, \tag{15.28}$$

where σ_{ij} is the stress tensor using standard indicial notations, ε_{ij} the total strain tensor and n_j is the component in the jth direction of the unit vector $\boldsymbol{n}$ normal to Γ. Rice has shown that the J-integral was path-independent and corresponded to the SERR G. The J-integral is very useful for determining the SERR especially for ductile materials.

For the case where a thermal loading is applied, Wilson and Yu [25] have shown that the SERR G_t due to the thermal loading is given by the modified J-integral (MJI) J_t as follows:

$$J_t = J_{\Gamma,t} - \int_{\Gamma^+ + \Gamma^-} t_i \frac{\partial u_i}{\partial x_1} \mathrm{d}s - \frac{E\alpha}{1-2\nu} \int_{A_0} \left[\frac{1}{2} \frac{\partial}{\partial x_1} (\theta \varepsilon_{ii}) - \varepsilon_{ii} \frac{\partial \theta}{\partial x_1} \right] \mathrm{d}A, \tag{15.29}$$

where $J_{\Gamma,t}$ is the J-line integral computed along any arbitrary contour Γ around the crack-tip subjected to thermal loads, α is the coefficient of thermal expansion (CTE), θ the temperature relative to the reference temperature where thermal stress is zero, t_i the surface tractions along the crack surfaces Γ^+ and Γ^-, u_i the displacement component, x_1 the Cartesian coordinate along the interface, and A_0 the domain around the crack tip bounded by Γ, Γ^+ and Γ^-.

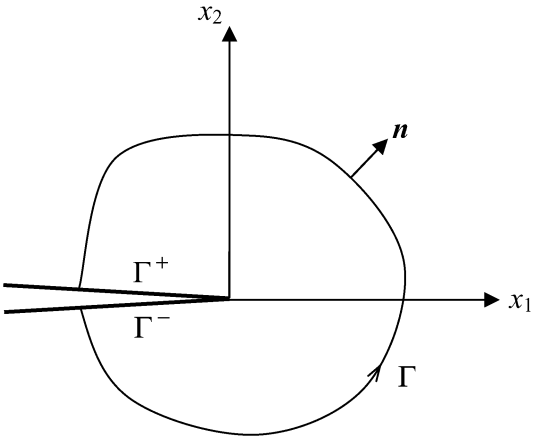


FIGURE 15.8. Arbitrary contour Γ around crack tip.

The corresponding modified J-integral J_h for hygrostress was obtained by Lin and Tay [26,27] as

$$J_h = J_{\Gamma,h} - \int_{\Gamma^+ + \Gamma^-} t_i \frac{\partial u_i}{\partial x_1} \mathrm{d}s - \frac{E\beta}{1-2\nu} \int_{A_0} \left[\frac{1}{2} \frac{\partial}{\partial x_1} (C\varepsilon_{ii}) - \varepsilon_{ii} \frac{\partial C}{\partial x_1} \right] \mathrm{d}A, \tag{15.30}$$

where $J_{\Gamma,h}$ is the J-line integral computed along any arbitrary contour Γ around the crack-tip subjected to hygro loads.

For a bimaterial interface crack, Malyshev and Salganik [16], showed that as long as the crack plane is flat, the modified J-integral will give the SERR G. They further showed that K, G and J are related by

$$G = J = \frac{(1-\beta_D^2)}{E^*} K^2. \tag{15.31}$$

The value of J is calculated automatically by some finite element codes such as ABAQUS and is hence easily obtained by performing a finite element analysis under thermal or hygro loading. Hence under hygrothermal loading, K_t and K_h can be obtained from J_t and J_h and Equation (15.31). However, a difficulty arises here as it is necessary to obtain the separate components K_{I} and K_{II} of both $\boldsymbol{K}_t$ and $\boldsymbol{K}_h$ in order to obtain the composite hygrothermal $\boldsymbol{K}_{tot}$ according to Equation (15.25). This difficulty can be resolved by obtaining the ratio $K_{\mathrm{II}}/K_{\mathrm{I}}$ using the extrapolation method suggested by Yuuki and Cho [13] which is described in the previous section. Once the ratio $K_{\mathrm{II}}/K_{\mathrm{I}}$ and K are determined, the K_{I} and K_{II} components of both $\boldsymbol{K}_t$ and $\boldsymbol{K}_h$ can be calculated and then used to obtain the composite $\boldsymbol{K}_{tot}$ according to Equation (15.25). G_{tot} is then calculated from Equations (15.26) and (15.17).

15.7.3. Modified Virtual Crack Closure Method

The strain energy release rate G is a measure of the energy available for an increment of crack extension. Irwin's virtual crack-closure method [15] is based on the principle that the work necessary to extend the crack from a to $a + \Delta$ is the same as that necessary to close the crack tip from $a + \Delta$ to a. This procedure will provide a set of formulas for G that depends on the crack opening displacement and the nodal forces at and ahead of the crack tip. Following Irvin's virtual crack-closure method, Rybicki and Kanninen [28] developed the modified virtual crack closure method (MVCCM) of calculating stress intensity factors for cracks in homogeneous materials. Raju [29] continued to develop the procedure for calculating G for higher order elements and singular elements. With reference to Figure 15.9, Raju showed that the mode I and mode II components of G are given by

$$G_{\mathrm{I}} = \lim_{\Delta \to 0} \left\{ -\frac{1}{2\Delta} [F_{yi}(v_k - v_{k'})] \right\}, \tag{15.32}$$

$$G_{\mathrm{II}} = \lim_{\Delta \to 0} \left\{ -\frac{1}{2\Delta} [F_{xi}(u_k - u_{k'})] \right\}, \tag{15.33}$$

where F_{xi} and F_{yi} are the x- and y-components of the nodal force at node i, and, u_k and v_k are the x- and y-components of the displacement at node k, respectively.

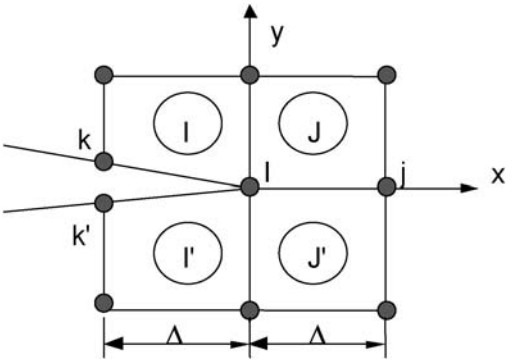


FIGURE 15.9. Mesh of 4-noded elements around crack tip.

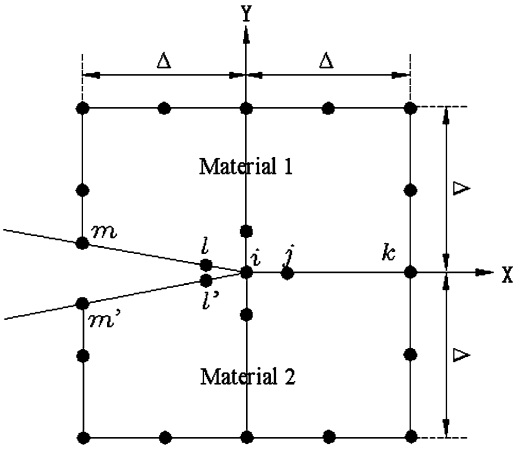


FIGURE 15.10. Mesh of 8-noded singular elements around crack tip.

For an 8-noded singular element (Figure 15.10),

$$G_{\mathrm{I}} = -\frac{1}{2\Delta}\big\{F_{yi}[t_{11}(v_m - v'_m) + t_{12}(v_l - v'_l)] + F_{yj}[t_{21}(v_m - v'_m) + t_{22}(v_l - v'_l)]\big\}, \tag{15.34}$$

$$G_{\mathrm{II}} = -\frac{1}{2\Delta}\big\{F_{xi}[t_{11}(u_m - u'_m) + t_{12}(u_l - u'_l)] + F_{xj}[t_{21}(u_m - u'_m) + t_{22}(u_l - u'_l)]\big\}, \tag{15.35}$$

where $t_{11} = 6 - (3/2)\pi$, $t_{12} = 6\pi - 20$, $t_{21} = 1/2$ and $t_{22} = 1$.

The SERR G is then given by

$$G = G_{\mathrm{I}} + G_{\mathrm{II}}. \tag{15.36}$$

It will be found that for cracks in homogeneous media ($\beta_D = 0$), the components of G and K are related directly by

$$G_{\mathrm{I}} = K_{\mathrm{I}}^2/E^* \quad \text{and} \quad G_{\mathrm{II}} = K_{\mathrm{II}}^2/E^*. \tag{15.37}$$

Hence the mode mixity can be given by

$$\tan\psi \equiv \frac{K_{\mathrm{II}}}{K_{\mathrm{I}}} = \sqrt{\frac{G_{\mathrm{II}}}{G_{\mathrm{I}}}}. \tag{15.38}$$

However, for bimaterial interface cracks ($\beta_D \neq 0$), it is found that [30], the MVCCM can still be used to calculate the magnitude of the SERR but the mode mixity phase angle ψ should be obtained from

$$\frac{K_{\mathrm{II}}}{K_{\mathrm{I}}} = \frac{-2A_{12}(1+R_0) + \sqrt{4A_{12}^2(1+R_0)^2 - 4(A_2 - R_0A_1)(A_1 - R_0A_2)}}{2(R_0A_2 - A_1)}$$
$$\text{if } u_k - u_{k'} < 0, \tag{15.39}$$

or

$$\frac{K_{\mathrm{II}}}{K_{\mathrm{I}}} = \frac{-2A_{12}(1+R_0) - \sqrt{4A_{12}^2(1+R_0)^2 - 4(A_2 - R_0A_1)(A_1 - R_0A_2)}}{2(R_0A_2 - A_1)}$$
$$\text{if } u_k - u_{k'} > 0, \tag{15.40}$$

where $R_0 = G_{\mathrm{II}}/G_{\mathrm{I}}$, m and m' are the corresponding nodes at the upper and lower crack surfaces (Figure 15.9), and

$$A_1 = \frac{(c_1+c_2)}{8\cosh^2(\pi\varepsilon)(1+4\varepsilon^2)}\left[\cos\left(2\varepsilon\ln\frac{\Delta}{2a}\right)\left(\frac{1}{4} - \frac{67}{42}\varepsilon^2\right)\right.$$
$$\left. + \sin\left(2\varepsilon\ln\frac{\Delta}{2a}\right)\left(\frac{17}{14}\varepsilon - \frac{1}{3}\varepsilon^3\right) + \left(\frac{1}{4} + \varepsilon^2\right)\right], \tag{15.41}$$

$$A_2 = \frac{(c_1+c_2)}{8\cosh^2(\pi\varepsilon)(1+4\varepsilon^2)}\left[\cos\left(2\varepsilon\ln\frac{\Delta}{2a}\right)\left(-\frac{1}{4} + \frac{67}{42}\varepsilon^2\right)\right.$$
$$\left. + \sin\left(2\varepsilon\ln\frac{\Delta}{2a}\right)\left(-\frac{17}{14}\varepsilon + \frac{1}{3}\varepsilon^3\right) + \left(\frac{1}{4} + \varepsilon^2\right)\right], \tag{15.42}$$

$$A_{12} = \frac{(c_1+c_2)}{8\cosh^2(\pi\varepsilon)(1+4\varepsilon^2)}\left[\cos\left(2\varepsilon\ln\frac{\Delta}{2a}\right)\left(\frac{17}{14}\varepsilon - \frac{1}{3}\varepsilon^3\right)\right.$$
$$\left. + \sin\left(2\varepsilon\ln\frac{\Delta}{2a}\right)\left(-\frac{1}{4} + \frac{67}{42}\varepsilon^2\right)\right], \tag{15.43}$$

where $c_1 = (\kappa_1 + 1)/\mu_1$, $c_2 = (\kappa_2 + 1)/\mu_2$.

15.7.4. Variable Order Boundary Element Method

Classical linear elastic fracture mechanics shows the presence of stresses of singularity $r^{1/2}$ at the crack tip, and the standard isoparametric quadratic element is unable to give a good representation of the surrounding stress field, leading to inaccurate results unless fine meshes are used. From Equation (15.5) it can be seen that interface cracks have a singularity of $r^{1/2-i\varepsilon}$.

In order to address the problem of the correct singularity in the stress field, a variable-order singular field boundary element has been developed [31,32] for use at such singularities in place of the standard isoparametric element. In this new method, called the Variable Order Boundary Element Method (VOBEM), the SIFs, which may be complex, replace the tractions as unknowns to be solved for at the crack tip node. In the proposed element, the tractions T_i and displacements u_i take the following forms:

$$T_i = \sum_{h=1}^{s} a_i^h \beta_i^h(\theta)\xi^{\lambda_h - 1} + b_i + c_i\xi, \tag{15.44}$$

$$u_i = a_i^h \gamma_i^h(\theta)\xi^{\lambda_1} + d_i + e_i\xi. \tag{15.45}$$

Here, s is the number of included singularities, $\lambda_h - 1$ is the order of the hth stress singularity, $\beta_i^h(\theta)$ and $\gamma_i^h(\theta)$ are the eigen functions [33] of the hth singularity, and ξ is a coordinate measured along the element from the singular node. During the numerical implementation, the unknowns $a_i^h, b_i, c_i, d_i, e_i$ are expressed in terms of the physically meaningful SIFs, nodal tractions and nodal displacements. Using this singular field boundary element, problems involving stress singularities can be solved accurately and efficiently without much mesh refinement near the singular nodes. The SIFs are also evaluated directly as nodal unknowns in the numerical scheme, and this reduces the post-processing required to obtain the values of the SIFs.

15.7.5. Interaction Integral Method

The strain energy release rate G can be evaluated numerically using the interaction integral method proposed by Shih and Asaro [34]. With reference to Figure 15.11, taking into account possible material nonlinearity, thermal strain and crack face tractions, the following general expression for G and the J-integral in two dimensions using standard indicial notations can be obtained as:

$$G = J = \int_A \left[\left(\sigma_{ij} \frac{\partial u_j}{\partial x_1} - W\delta_{1i} \right) \frac{\partial q}{\partial x_i} + \left(\alpha\sigma_{ii} \frac{\partial \theta}{\partial x_1} \right) q \right] \mathrm{d}A - \int_{C^+ + C^-} t_i \frac{\partial u_i}{\partial x_1} q \mathrm{d}C, \tag{15.46}$$

where C^+ and C^- are the upper and lower crack faces, respectively, t_i is the traction on the crack faces, W is the strain energy density, α is the coefficient of thermal expansion, θ the temperature (relative to the reference temperature where thermal stress is zero), u_i are the displacement components, σ_{ij} are the stress components and δ_{ij} is the Kronecker's delta. q can be interpreted as a normalized virtual displacement, although the above derivation does not require such an interpretation. The q function is merely a mathematical device that

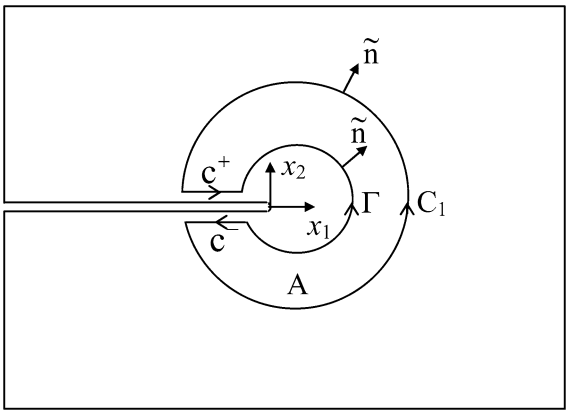


FIGURE 15.11. Conventions for domain integral.

enables the generation of an area integral, which is better suited to numerical calculations. Here q is an arbitrary but sufficiently smooth function that is equal to unity on Γ and zero on C_1.

For uncoupled thermoelasticity, W is given by

$$W(\varepsilon_{ij},\theta)=\int_0^{\varepsilon_{ij}^m}\sigma_{ij}\mathrm{d}\varepsilon_{ij}^m=\frac{1}{2}(\sigma_{jk}\varepsilon_{jk}-\alpha\theta\varepsilon_{kk}), \tag{15.47}$$

where ε_{ij}^m is the mechanical strain. Assuming open cracks, the traction on the crack faces t_i will be zero and Equation (15.46) reduces to

$$G=J=\int_A\left[\left(\sigma_{ij}\frac{\partial u_j}{\partial x_1}+\frac{1}{2}\alpha\theta\varepsilon_{kk}\delta_{1i}-\frac{1}{2}\sigma_{jk}\varepsilon_{jk}\delta_{1i}\right)\frac{\partial q}{\partial x_i}+\alpha\sigma_{ii}\frac{\partial\theta}{\partial x_1}q\right]\mathrm{d}A. \tag{15.48}$$

From Equations (15.6) and (15.17), G and K are related by

$$G=\frac{1-\beta_D^2}{E^*}K^2=\frac{1-\beta_D^2}{E^*}\left(K_{\mathrm{I}}^2+K_{\mathrm{II}}^2\right). \tag{15.49}$$

To extract K_{I} and K_{II} from $\boldsymbol{K}$, an auxiliary field of known intensity k_1 is superimposed onto the actual field, resulting in an auxiliary field with energy release rate

$$G_{\mathrm{aux},1}=\frac{1-\beta_D^2}{E^*}k_1^2, \tag{15.50}$$

and the resultant combined energy release rate is

$$G_{\mathrm{com},1}=\frac{1-\beta_D^2}{E^*}\left[(K_{\mathrm{I}}+k_1)^2+K_{\mathrm{II}}^2\right]. \tag{15.51}$$

The interaction energy release rate is defined as

$$G_{\mathrm{int},1}=G_{\mathrm{com},1}-(G+G_{\mathrm{aux},1}). \tag{15.52}$$

From Equations (15.49) to (15.52) we obtain

$$G_{\text{int},1} = 2\left(\frac{1-\beta_D^2}{E^*}\right)K_{\text{I}}k_1. \tag{15.53}$$

Thus for an auxiliary field of known intensity factor, say $k_1 = 1$, we obtain from Equation (15.53)

$$K_{\text{I}} = \frac{1}{2}\left(\frac{E^*}{1-\beta_D^2}\right)G_{\text{int},1}. \tag{15.54}$$

Similarly K_{II} can be determined from

$$K_{\text{II}} = \frac{1}{2}\left(\frac{E^*}{1-\beta_D^2}\right)G_{\text{int},2}, \tag{15.55}$$

where $G_{\text{int},2}$ is the interaction energy associated with an auxiliary field of intensity, $k_2 = 1$.

Using the reciprocity theorem

$$\frac{1}{2}\sigma_{ij}\varepsilon_{ij}^{\text{aux}} + \frac{1}{2}\sigma_{ij}^{\text{aux}}\varepsilon_{ij} = \sigma_{ij}\varepsilon_{ij}^{\text{aux}}, \tag{15.56}$$

and Equations (15.48) and (15.52), we obtain

$$G_{\text{int}} = \int_A \left[\left(\sigma_{ij}\frac{\partial u_j^{\text{aux}}}{\partial x_1} + \sigma_{ij}^{\text{aux}}\frac{\partial u_j}{\partial x_1} - \sigma_{jk}\varepsilon_{jk}^{\text{aux}}\delta_{1i}\right)\frac{\mathrm{d}q}{\mathrm{d}x_i} + \alpha\sigma_{ii}^{\text{aux}}\frac{\mathrm{d}\theta}{\mathrm{d}x_1}q\right]\mathrm{d}A,$$
$$(i, j, k = 1, 2). \tag{15.46}$$

The auxiliary fields σ_{ij}^{aux}, $\varepsilon_{ij}^{\text{aux}}$, $\partial u_j^{\text{aux}}/\partial x_1$ are known and the actual fields are obtained from a full field finite element analysis. Thus by evaluating the above integral, K_{I} and K_{II} can be obtained from Equations (15.54) and (15.55).

15.8. EXPERIMENTAL VERIFICATION

To verify the above-described methodology for predicting the onset of propagation of an initial delamination at the edge of the pad-encapsulant interface under hygrothermal loading, experiments were conducted as described briefly below. Further details are given in [35]. The lead frame employed in the plastic IC packages used for the verification tests had a very thin layer of silver of width 1 mm plated all around the edge of the copper lead frame pad. This was chosen to introduce a precisely-controlled initial delamination region of 1 mm width at the periphery of the pad since the adhesion between silver and mold compound is known to be poor. In the normal situation, a die would be attached to the pad on the same side as the silver border. This would have interfered with the propagation of the initial delamination. Hence, the packages used in the verification tests were fabricated without the die. The packages were transfer molded at 175°C and post-mold cured at the same temperature for 5 hours. Several specimens were then examined using a scanning

acoustic microscope. It was found that the samples thus made did in fact have a delaminated border of width 1 mm around the pad.

The fabricated packages were then divided into three groups. One group was fully dried by baking in an oven until all moisture was driven off while the other two were preconditioned in an environmental chamber at 30°C/60%RH (L3) and 85°C/85%RH (L1), respectively, for 504 hours. One package from each of the three groups was taken and placed in an oven where the temperature was set at 160°C. After about 15 minutes to allow for thermal equilibrium in the package to be established, the packages were retrieved. The oven temperature was then increased by a certain amount and another package from each of the three groups was taken and placed in the oven. The procedure was repeated until an oven temperature of 240°C was reached. In this manner, the packages in each of the three groups would have been subjected to various oven temperatures between 160°C and 240°C. Afterwards all the packages tested were examined with a scanning acoustic microscope to determine the temperature at which delamination started to grow.

Figure 15.12 shows the variation of G_t, the thermal strain energy release rate, for a dry package, with temperature. The SERRs were calculated using the interaction integral method described earlier. Since the package is dry there is no hygrostress, and $G_{tot} = G_t$. The phase angle ψ was found to be 70.1°. From Equation (15.19), the interface toughness G_c for $C = 0$, mode mixity $\psi = 70.1°$ and varying temperatures was calculated and also plotted in Figure 15.12. The temperature at which the G_{tot} curve intersects the G_c curve is the temperature at which the initial delamination is predicted to propagate along the pad-encapsulant interface. From Figure 15.12, the delamination temperature for the dry IC package is predicted to be about 225°C.

Figure 15.13 shows the variation of G_c, G_t and G_{tot} with temperature for an L3-preconditioned package. Here G_h is not zero and G_{tot} is everywhere greater than G_t. As

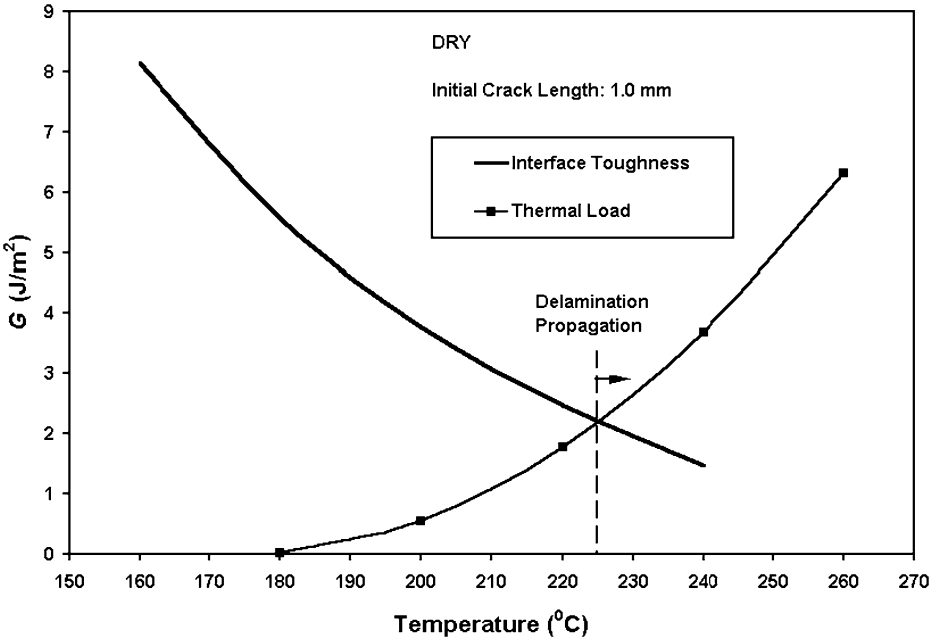


FIGURE 15.12. Variation of G_c and G_{tot} with temperature for dry package.

can be seen from Figure 15.13, the delamination is expected to propagate at about 187°C. Similarly from Figure 15.14 one can observe that the delamination propagation temperature for the package subjected to L1 preconditioning is about 180°C. It can be seen from Figures 15.12–15.14 that as the moisture content of the packages increases, the solder reflow temperature at which delamination propagation occurs will decrease. This is because the

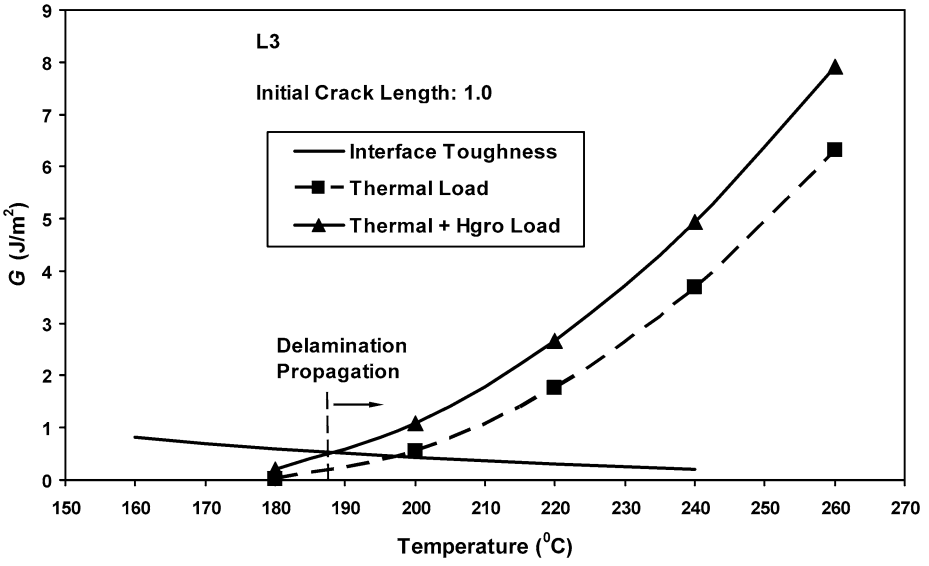


FIGURE 15.13. Variation of G_c, G_t and G_{tot} with temperature for L3-preconditioned package.

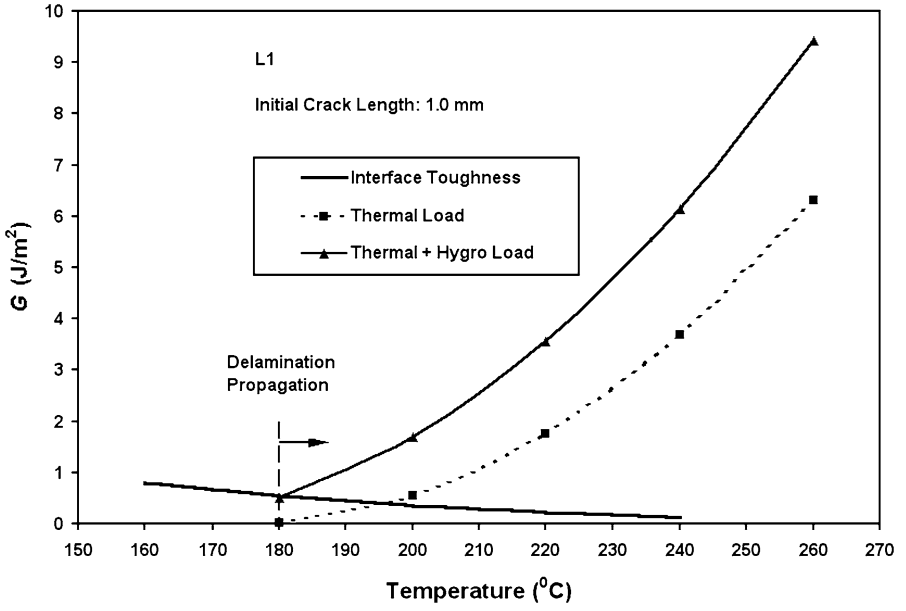


FIGURE 15.14. Variation of G_c, G_t and G_{tot} with temperature for L1-preconditioned package.

moisture absorbed by the package both increases the crack driving force G_{tot} and decreases the interface toughness G_c. This trend was also observed in the verification experiments.

Images obtained from the scanning acoustic microscopic analysis of the packages subjected to varying levels of moisture preconditioning and varying oven temperatures are shown in Figures 15.15–15.17. From the figures, it is obvious that the initial delaminated region is exactly the area coated with the thin layer of silver, a border 1 mm wide around the pad. It can be seen that the delamination temperature for a dry package is about 226°C while those for L3 and L1 moisture preconditioning are about 186°C and 178°C, respectively. The predicted and measured values of delamination temperature are tabulated in

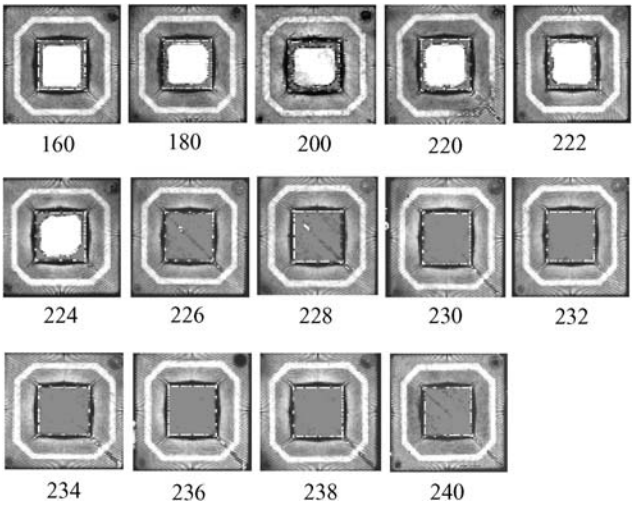


FIGURE 15.15. SAM images of delamination propagation along the pad-encapsulant interface in fully dried packages subjected to increasing temperature.

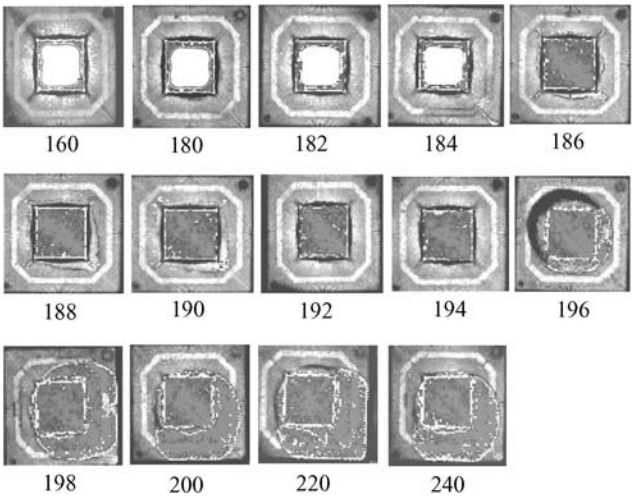


FIGURE 15.16. SAM images of delamination propagation along the pad-encapsulant interface in L3 preconditioned packages subjected to increasing temperature.

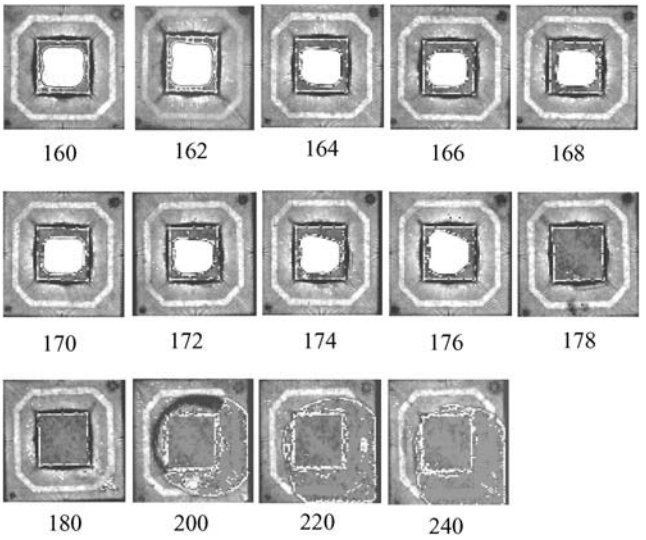


FIGURE 15.17. SAM images of delamination propagation along the pad-encapsulant interface in L1 preconditioned packages subjected to increasing temperature.

TABLE 15.2.
Comparison between predicted and measured delamination temperature.

Level of moisture preconditioning	Predicted delamination temperature	Measured delamination temperature
Dry	225°C	226°C
30°C/60%RH (L3)	187°C	186°C
85°C/85%RH (L1)	180°C	178°C

Table 15.2 for comparison. As can be seen, very good agreement within 2°C has been obtained which verifies the accuracy of the predictive methodology presented.

15.9. CASE STUDIES

15.9.1. Delamination Along Pad-Encapsulant Interface

The delamination of the pad-encapsulant interface of an 80-pin plastic quad flat package (PQFP) of outer dimensions 20 mm × 14 mm × 2.7 mm and a die-pad of dimensions 8 mm × 8 mm × 0.15 mm was studied by Tay and Lin [36]. An initial crack of 1 mm was assumed at the edge of the pad-encapsulant interface. The modified J-integral method was employed to calculate the SIF. Figure 15.18 illustrates the extrapolation process in obtaining the mode mixity $K_{\mathrm{II}}/K_{\mathrm{I}}$ using Yuuki and Cho's method [12]. It also shows that the mode mixity does not vary much with temperature. Figure 15.19 shows that the delamination temperature for packages preconditioned for 168 hours at 85°C/85%RH is 187°C while that for packages preconditioned at 85°C/60%RH is 195°C.

Verification experiments similar to that described above [36] were conducted which verified the predictions. Before encapsulation of the package specimens, a very thin layer

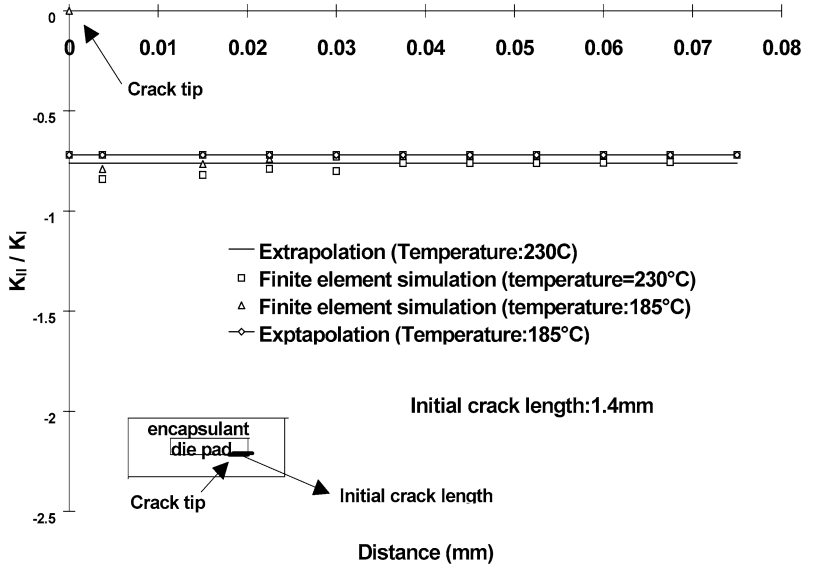


FIGURE 15.18. Determination of tan ψ by extrapolation.

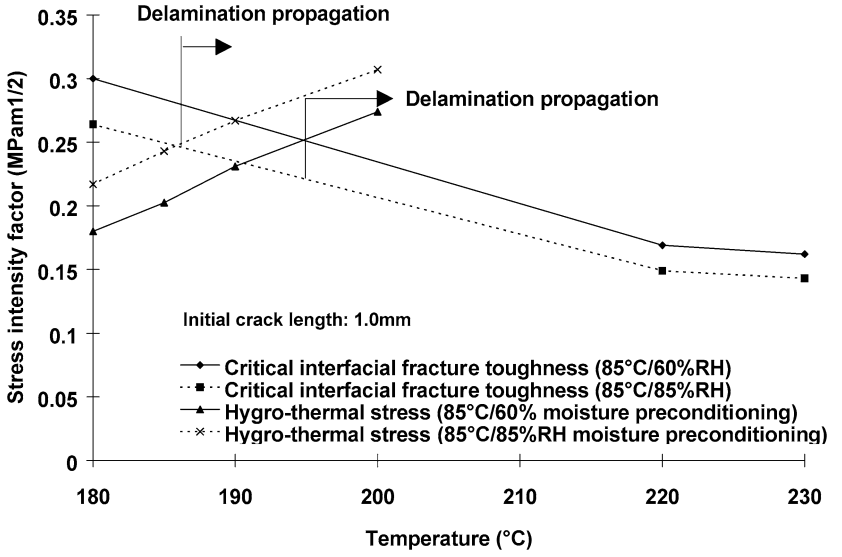


FIGURE 15.19. Prediction of delamination temperature.

of a mold release agent was painted along an area of the pad of width 1 mm parallel to the edge of the pad. The purpose of this was to induce an initial delaminated region 1 mm wide adjacent to one edge of the pad. The results for the specimens preconditioned at 85°C/60%RH are shown in Figure 15.20. It can be seen that the predicted delamination temperature of 195°C agreed well with the observed delamination temperature 190–195°C.

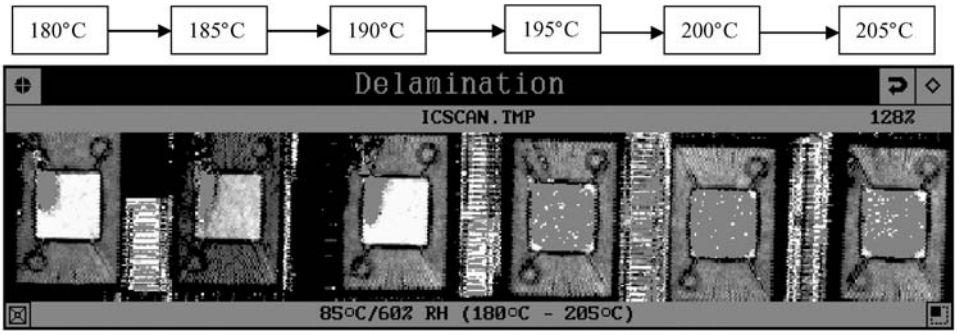


FIGURE 15.20. Delamination propagation of 85°C/60%RH preconditioned packages during solder reflow.

15.9.2. Delamination Along Die-Attach/Pad Interface

The package used in this study by Tay and Goh [37] was an 80-pin quad flat package (QFP) with outer dimension of 20 mm × 14 mm × 2.7 mm and a die-pad size of 10.8 mm × 8.4 mm × 0.15 mm (Figure 15.21). It was assumed that a small crack existed initially at the interface between the die-attach and the die-pad. G and ψ were calculated using the MVCCM. The effect of the location of this initial crack was studied. Three locations #1, #2 and #3 (Figure 15.21) with crack tips at 0 mm, 2.73 mm and 5.16 mm from the reference edge were selected for study.

The variation of K_{tot} with distance from the edge of the die-attach layer is plotted in Figure 15.22. It can be seen that K_{tot} is maximum at the edge crack and decreases almost exponentially with increasing distance from the edge of the die-attach layer. K_{tot} for the edge crack (#1) is 123–300 times the value for the central crack (#3). This strongly suggests that delamination will most likely initiate at the edge than at the central region of the pad/die-attach interface, even though the mode mixity ψ_{tot} for the edge crack is higher than that for the central crack which implies that the interface fracture toughness is higher at the edge. It has been reported by Hutchinson and Suo [13] that the interface toughness for $\psi \cong 90°$ can be about 6–10 times the value for $\psi \cong 0°$.

K_t was calculated for an edge crack with crack length ranging from 0.1 mm to 2 mm. From Figure 15.23 it can be seen that the stress intensity factor K_t, due to thermal load only, increases with increasing crack length. The rate of increase is rapid up to a crack length of about 0.2 mm and then becomes almost linear beyond 0.2 mm up to 2 mm. From Figure 15.24, it can be seen that ψ remains approximately constant up to 0.2 mm before decreasing rapidly with increasing crack length. The variation of K_{tot} and ψ with crack length shown in Figures 15.23 and 15.24 suggests that as an edge crack extends along the die-attach/pad interface, the magnitude of K_t increases while the magnitude of K_c decreases. This suggests that once the initial delamination starts to propagate, it will continue to extend until the entire interface is fully delaminated.

The effect of water vapor pressure acting on the crack surfaces of the interface crack was also studied. There have been some ambiguity about the magnitude of the water vapor pressure at the interfacial crack [4]. Some researchers use the saturated vapor pressure P_{sat} corresponding to the solder reflow temperature while others use a much lower pressure P_{comp} computed considering the diffusion of moisture into the interfacial crack [38]. In this study, both $P_{sat} = 2.114$ MPa and the computed value $P_{comp} = 0.2074$ MPa were employed. The results are also shown in Figures 15.23 and 15.24. It can be seen that the

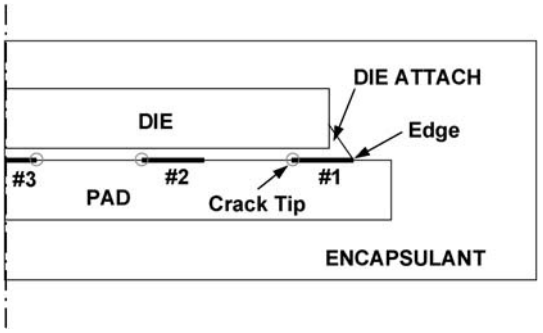


FIGURE 15.21. Schematic diagram of the finite element model.

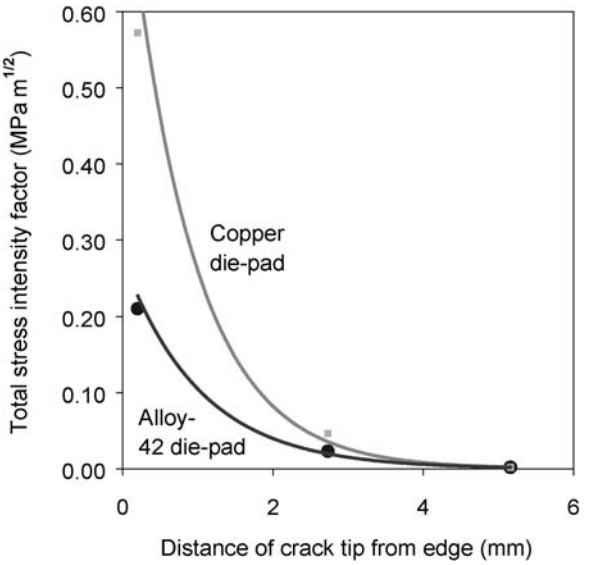


FIGURE 15.22. Variation of K_{tot} along die-attach layer.

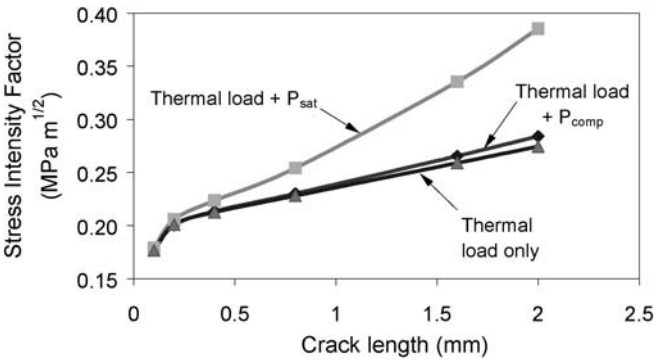


FIGURE 15.23. Variation of SIF with crack length.

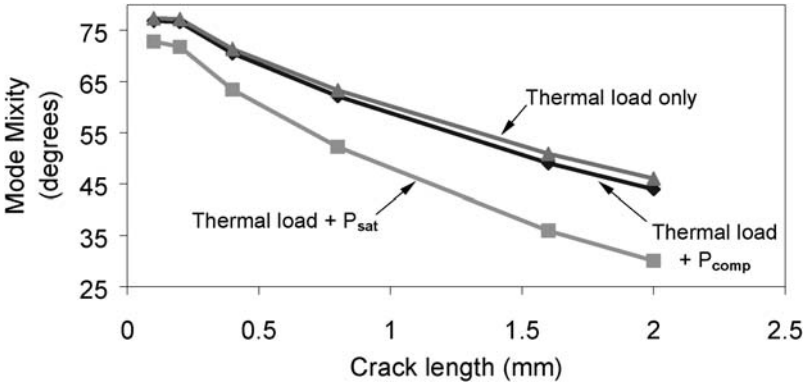


FIGURE 15.24. Variation of mode mixity with crack length.

additional contribution from interfacial water vapor pressure to the total SIF is small for small crack lengths. However, as the crack extends, the contribution from interfacial water vapor pressure increases exponentially. The effect is much larger when P_{sat} is used. The results here suggest that the effect of interfacial water vapor pressure on crack propagation is negligible for small cracks but becomes exponential for larger cracks. It can also be seen from Figure 15.24 that ψ is decreased by the interfacial water vapor pressure. This is reasonable since the vapor pressure acts normal to the surface of the crack, enhancing the opening mode (mode I) component and decreasing ψ. Thus, for a large crack with built-up vapor pressure, the cracking mode is predominantly mode I.

15.9.3. Analysis Using Variable Order Boundary Element Method

The effect of crack location and crack size on G and ψ along the pad-encapsulant interface has been studied by Tay et al. [31] using the variable order boundary element method (VOBEM). The package considered is shown in Figure 15.25. In the boundary element model, the boundaries of each domain were discretized with isoparametric quadratic elements, except at the corners and crack-tips where variable-order singular field elements were used to model the stress singularities. The number of elements used in the final mesh is indicated in parenthesis in Figure 15.25. A thermal stress analysis is carried out on this model to calculate the SIFs induced in the crack of various lengths at specific locations along the pad-encapsulant interface when the package temperature is increased from the stress-free temperature of 175°C to 215°C.

To determine the possibility of delamination propagation in the package, G and ψ values of the tips of the interfacial cracks were analyzed. In the present analysis, a study on the effect of the size of interfacial defects and their positions relative to the corner of the leadframe pad was carried out. This was done by considering interfacial defects or delaminations of varying sizes at different locations A, B and C at varying distances from the pad corner, as shown in Figure 15.25. The magnitudes of G and ψ for both left and right tips of the interfacial crack are plotted against their distance r from the corner in Figures 15.26 to 15.29. It can be seen that both the energy release rate G and the magnitude of the phase angle ψ increases with the size of the delamination as well as with proximity to the pad corner. It can also be seen that for any given crack, G at the right crack tip is always higher than that at the left crack-tip.

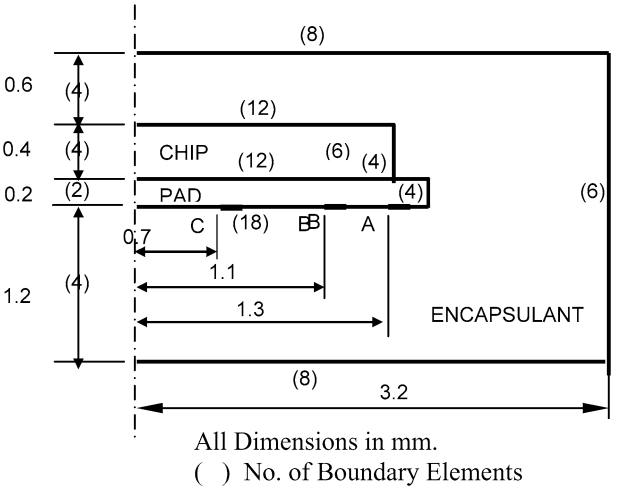


FIGURE 15.25. Boundary element model of IC package.

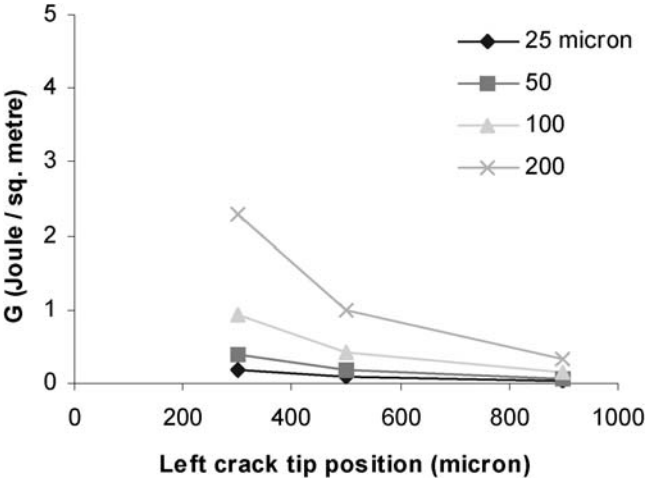


FIGURE 15.26. G (left crack tip) for different crack lengths and positions.

Consider the case of a small 25-μm defect or initial delamination at location C of the pad-encapsulant interface. From Figures 15.27 and 15.29, the magnitude of $|\psi|$ at the right tip and the left tip are about the same, so that the fracture toughness G_c is about the same at both tips. From Figures 15.26 and 15.28, it can be seen that when the package temperature (thermal loading) is gradually increased during solder reflow, the value of G at the right crack tip is always greater than that at the left tip. Thus the value of G at the right tip will exceed G_c first causing the crack to propagate towards the corner. As the crack grows, the size of the delamination increases and the distance from the corner decreases. Both these effects make the value of G greater (Figure 15.28). At the same time they both make $|\psi|$ smaller (Figure 15.29). As can be seen from Figure 15.6, G_c decreases monotonically with decreasing $|\psi|$. These trends suggest that the right crack tip should continue to propagate

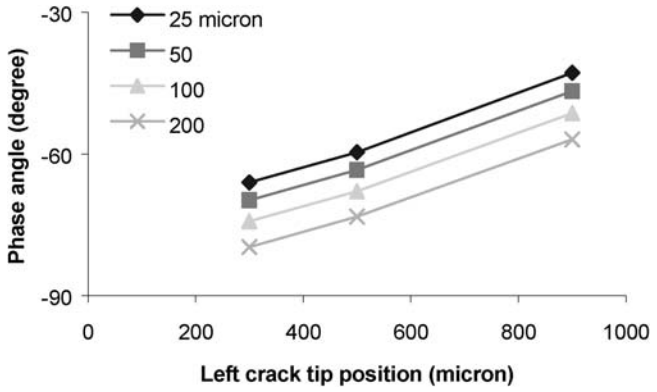


FIGURE 15.27. Phase angle (left crack tip) for different crack lengths and positions.

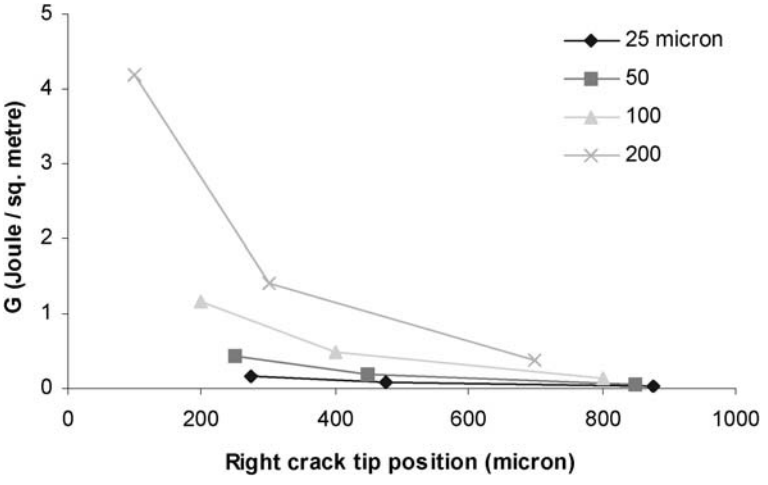


FIGURE 15.28. *G* (right crack tip) for different crack lengths and positions.

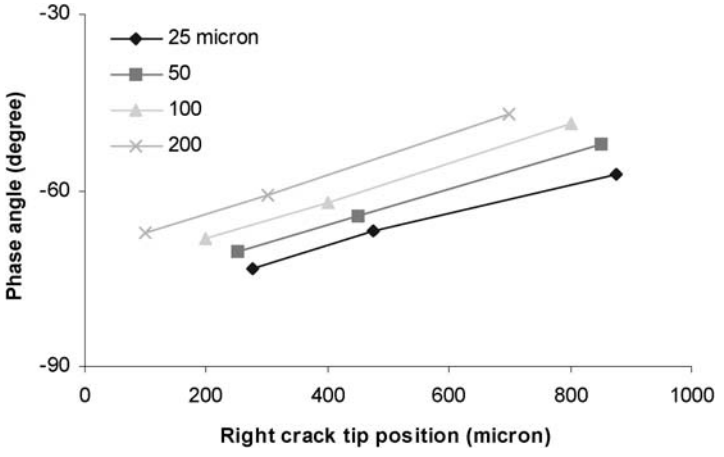


FIGURE 15.29. Phase angle (right crack tip) for different crack lengths and positions.

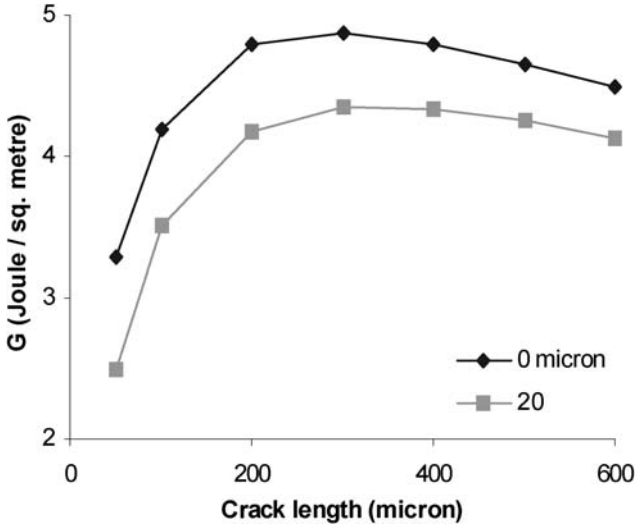


FIGURE 15.30. G (left crack tip) with right crack tip at 0 and 20 μm from corner.

until it reaches the pad corner. Beyond this point, as the temperature is increased further, one of the following 3 possible events can take place:

(a) the right crack tip can propagate up the vertical face of the pad
(b) the right crack tip can propagate into the encapsulant
(c) the right crack tip can be arrested at the corner and the left tip propagates.

Usually interfaces are weaker than the bulk material, hence event (b) is unlikely to occur before (a) or (c). Between events (a) and (c), the latter is more likely to occur as the toughness at the vertical face of the pad is likely to be much higher due to its higher roughness. Hence, as the temperature is further increased, the left tip is expected to propagate.

However, as the left crack tip propagates, it moves away from the corner, and the two effects of size and distance from the corner oppose each other (Figure 15.26). To investigate which factor dominates, an analysis is performed, keeping the right crack tip fixed at the pad corner and varying the length of the crack. The variation of G and ψ of the left crack tip with distance from the corner is given in Figures 15.30 and 15.31, respectively. As the crack grows and the crack tip moves away from the corner, it can be seen that the value of G still increases, showing that the effect of the size of the crack is more significant than the effect of distance from the corner. However, a maximum value is reached at a crack length of about 300 μm. It is also noted that as the crack grows, $|\psi|$ also increases, which implies that G_c increases. The trends of increasing G and increasing G_c with increasing crack length suggests that once the left crack tip propagates, whether the crack will continue to extend depends on the nature of variation of G and G_c with increasing crack length.

15.10. DISCUSSION OF THE VARIOUS NUMERICAL METHODS FOR CALCULATING G AND ψ

The MJI method requires the calculation of the J integral which may appear quite difficult to evaluate accurately. However, it can be shown that the J line integral can be re-

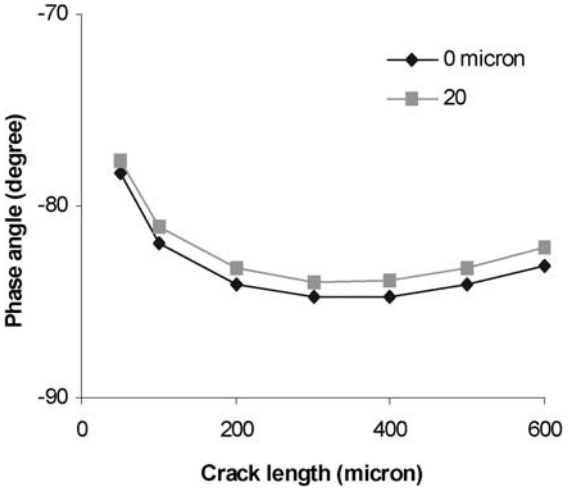


FIGURE 15.31. Phase angle (left crack tip) with right crack tip at 0 and 20 μm from corner.

cast as an area integral which makes it easier and more suited to the finite element method, resulting in accurate values being obtained. Some softwares have incorporated such calculations so that users only need to issue a simple command to have the value of the J integral around a crack tip determined. However, as mentioned earlier, for bimaterial cracks, just knowing the magnitude of J (and hence G) is not sufficient for predicting crack propagation. The mode mixity is also required and this can be done using the CSDEM as described earlier. However, this is not ideal as the analyst has to do some graph plotting.

For those finite element softwares which have not incorporated the calculating of the J integral, the CSDEM or MVCCM may be ideal. The CSDEM only requires the displacements of nodes along the crack surfaces to be computed using finite element analysis but some work has to be done by the analyst to plot and extrapolate the values of K and the ratio $K_{\mathrm{II}}/K_{\mathrm{I}}$, from which G and ψ may be calculated. Some uncertainty is inherent in the extrapolation process. The MVCCM only requires the displacements and nodal forces to be computed at a small number of nodes near the crack tip. Also, the formulas required to calculate the magnitude of G and ψ are relatively simple. However, the mesh around the crack tip needs to be sufficiently fine in order to get good accuracy. The use of singular elements at the crack tip can improve the accuracy by about 10%.

While the VOBEM should be more accurate than the MVCCM, since it models the actual order of singularity of the crack tip and also formulates the problem such that the SIFs become the primary unknowns which will come out in the solution directly without further postprocessing. Figure 15.32 shows the difference in G calculated using the VOBEM and the MVCCM for the package shown in Figure 15.25 for a crack with the right tip at the pad corner. As can be seen the difference is about 10%. However, in practice the formulation of the VOBEM is much more complex and no general purpose user-friendly boundary element software is widely available to date.

With the interaction integral method, it is possible to avoid any further work by the analyst. All the integrals required can be done automatically by the software in a manner transparent to the analyst, and the values of K_{I} and K_{II} can be outputted by the software when required. If a finite element software incorporates the calculation of K_{I} and K_{II} using

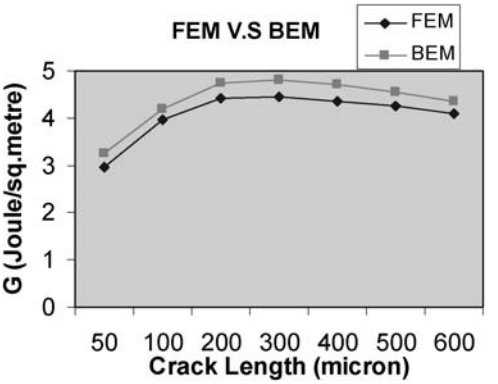


FIGURE 15.32. Values of G calculated using MVCCM (FEM) and VOBEM (BEM).

the interaction integral method, then this would appear to the method which requires the least effort from the analyst.

15.11. CONCLUSION

The mechanical failure of plastic IC packages undergoing solder reflow has been described. It was established that interfacial delamination in a plastic package is a precursor to popcorn failure of the package. The fracture mechanics approach to analyzing the reliability of plastic IC packages has been described. A method of determining the total stress intensity factor combining the effects of temperature, moisture and vapor pressure in the delamination has been developed. Some experimental data on the toughness of the copper/encapsulant interface as a comprehensive function of temperature, moisture concentration and mode mixity has been presented. A methodology for predicting delamination temperature in a plastic package undergoing solder reflow has been developed and verified by experiments involving actual packages. Several numerical methods that can be used to determine the strain energy release rate and the mode mixity have been described together with some discussion on their merits and demerits. Some case studies have been presented which provided a very good insight to the mechanics of delamination at various interfaces in plastic IC packages.

REFERENCES

1. I. Fukuzawa, S. Ishiguro, and S. Nanbu, Moisture resistance degradation of plastic LST's by reflow soldering, Proceedings, 23rd International Reliability Physics Symposium, 1985, pp. 192–197.
2. A.A.O. Tay, G.L. Tan, and T.B. Lim, A criterion for predicting delamination in plastic IC packages, Proceedings, International Reliability Physics Symposium, 1993, pp. 236–243.
3. A.A.O. Tay, G.L. Tan, and T.B. Lim, Predicting delamination in plastic IC packages and determining suitable mold compound properties, IEEE Trans. On CPMT, Part B: Advanced Packaging, 17(2), pp. 201–208 (1994).
4. A.A.O. Tay and T.Y. Lin, Effects of moisture and delamination on cracking of plastic IC packages during solder reflow, Proc. 46th Electronic Components and Technology Conference, 1996, pp. 777–782.
5. M. Kitano, et al., Analysis of package cracking during reflow soldering process, Proc. IRPS, 1988, pp. 90–95.
6. A.A.O. Tay and T.Y. Lin, Moisture diffusion and heat transfer in plastic IC packages, IEEE Transactions on Components, Packaging and Manufacturing Technology, Part A, 19(2), pp. 186–193 (1996).

7. M.L. Williams, The stress around a fault or crack in dissimilar media, Bull. Seism. Soc. Am., 49, pp. 199–204 (1959).
8. F. Erdogan, Stress distribution in a nonhomogeneous elastic plane with cracks, J. Appl. Mech., 30, pp. 232–236 (1963).
9. F. Erdogan, Stress distribution in bonded dissimilar materials with cracks, J. Appl. Mech., 87, pp. 403–410 (1965).
10. G.C. Sih and J.R. Rice, The bending of plates of dissimilar materials with cracks, J. Appl. Mech., 31, pp. 477–482 (1964).
11. J.R. Rice and G.C. Sih, Plane problems of cracks in dissimilar media, J. Appl. Mech., 32, pp. 418–423 (1965).
12. R. Yuuki and S.B. Cho, Efficient boundary element analysis of stress intensity factors for interface cracks in dissimilar materials, Engineering Fracture Mechanics, 34, pp. 179–188 (1989).
13. J.W. Hutchinson and Z. Suo, Mixed mode cracking in layered materials, Advances in Applied Mechanics, 29, pp. 63–191 (1991).
14. A.A. Griffith, The phenomenon of rupture and flow in solids, Phil. Trans. Roy. Soc. (London), 22, pp. 163–198 (1921).
15. G.R. Irwin, Fracture mechanics, in Structural Mechanics, Pergamon Press, New York, NY, 1960.
16. B.M. Malyshev and R.L. Salganik, The strength of adhesive joints using the theory of cracks, International Journal of Fracture Mechanics, 1, pp. 114–128 (1965).
17. N. Tanaka and A. Nishimura, Measurement of IC molding compound adhesion strength and prediction of interface delamination within package, ASME EEP-Vol. 10-2, Advances in Electronic Packaging, 1995, pp. 765–773.
18. K.M. Liechti and Y.S. Chai, Biaxial loading experiments for determining interfacial toughness, ASME Journal of Applied Mechanics, 58, pp. 680–687 (1991).
19. S. Liu, Y. Mei, and T.Y. Wu, Bimaterial interfacial crack growth as a function of mode-mixity, IEEE Transactions on Components, Packaging, and Manufacturing Technology, Part A, 18(3), pp. 618–626 (1995).
20. A.A.O. Tay and T.Y. Lin, Influence of temperature, humidity and defect location on delamination in plastics packages, IEEE Transactions on Components, Packaging and Manufacturing Technology, Part A, 22(4), pp. 512–518 (1999).
21. A.A.O. Tay, Y. Ma, S.H. Ong, and T. Nakamura, Measurement of interface toughness as a function of temperature, moisture concentration and mode mixity, Advances in Electronic Packaging, EEP-Vol. 26-2, pp. 1129–1136 (1999).
22. W.C. Carpenter, Extrapolation techniques for determining stress intensity factors, Engineering Fracture Mechanics, 18, pp. 325–332 (1983).
23. G.J. Hu and A.A.O. Tay, A new second-order analysis of bimaterial interfacial cracks, submitted for publication.
24. J.R. Rice, A path independent integral and approximate analysis of strain concentration by notch and cracks, Journal of Applied Mechanics, 35, pp. 379–386 (1968).
25. W.K. Wilson and I.W. Yu, The use of the J-integral in thermal stress cack problems, International Journal of Fracture, 15(4) (1979).
26. T.Y. Lin and A.A.O. Tay, A J-integral criterion for delamination of bi-material interfaces incorporating hygrothermal stresses, ASME EEP-Vol. 19-1, Advances in Electronic Packaging, 1997, pp. 1421–1428.
27. T.Y. Lin and A.A.O. Tay, Dynamics of moisture diffusion, hygrothermal stresses and delamination in plastic IC packages, ASME EEP-Vol. 19-1, Advances in Electronic Packaging, 1997, pp. 1429–1436.
28. E.F. Rybicki and M.F. Kanninen, A finite element calculation of stress intensity factors by a modified crack closure integral, Engineering Fracture Mechanics, 9, pp. 931–938 (1977).
29. I.S. Raju, Calculation of strain-energy release rates with higher order and singular finite elements. Engineering Fracture Mechanics, A8(3), pp. 251–274 (1987).
30. G.J. Hu and A.A.O. Tay, A modified virtual crack closure method for bimaterial interfacial crack analysis, submitted for publication.
31. A.A.O. Tay, K.H. Lee, and K.M. Lim, Numerical simulation of delamination in IC packages using a new variable-order singular boundary element, ASME Journal of Electronic Packaging, 125, pp. 569–575 (2003).
32. K.M. Lim, K.H. Lee, A.A.O. Tay, and W. Zhou, A new variable-order singular boundary element for two-dimensional stress analysis, Int. J. Numer. Methods Eng., 55, pp. 293–316 (2002).
33. P.S. Theocaris, The order of singularity at a multi-wedge corner of a composite plate, Int. J. Engng. Sci., 12, pp. 107–120 (1974).
34. C.F. Shih and R.J. Asaro, Elastic-plastic analysis of cracks on bimaterial interfaces: Part I. Small scale yielding, Journal of Applied Mechanics, 55, pp. 299–316 (1988).

35. A.A.O. Tay, Y. Ma, T. Nakamura, and S.H. Ong, A numerical and experimental study of delamination of polymer-metal interfaces in plastic packages at solder reflow temperatures, Proceedings of The Nineth Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems, ITHERM 2004, 1–4 June 2004, pp. 245–252.
36. A.A.O. Tay and T.Y. Lin, Moisture-induced interfacial delamination growth in plastic IC packages during solder reflow, Proceedings of 48th Electronic Components and Technology Conference, 1998, pp. 777–782.
37. A.A.O. Tay and K.Y. Goh, A study of delamination growth in the die attach layer of plastic IC packages under hygrothermal loading during solder reflow, IEEE Trans. on Device and Materials Reliability, 3, pp. 144–151 (2004).
38. K. Sawada, T. Nakazawa, N. Kawamura, and T. Sudo, Package deformation and cracking mechanism due to reflow soldering, Proceedings of Japan International Electronics Manufacturing Technology Symposium, 1993, pp. 295–298.

16

Dynamic Response of Micro- and Opto-Electronic Systems to Shocks and Vibrations: Review and Extension

E. Suhir

University of California, Santa Cruz, CA, University of Maryland, College Park, MD, and ERS/Siloptix Co., USA

16.1. INTRODUCTION

Electronic and photonic assemblies, components, devices and systems often experience dynamic loading. The ability to predict and, if necessary, minimize the adverse consequences of such a loading is of obvious practical importance [1,2].

In commercial electronics, dynamic loading can take place during handling or transportation of the equipment. In military, avionic, space, automotive, and marine electronics, dynamic loading occurs even during normal operation of the system. On the other hand, random vibrations are often applied deliberately (in addition to, or even instead of, thermal cycling or mechanical testing) as an effective and fast means to detect and weed out infant mortalities and to develop the most feasible and robust product. Shock loading is part of MIL-specs and other qualification requirements (see, for instance, [3,4]).

In the recent years, the necessity to protect portable electronics from impact loading (e.g., because of an accidental drop) triggered the development of both theoretical methods and experimental techniques for the prediction of the response of the system to, and minimizing the adverse consequences of, accidental shocks. The number of studies devoted to the dynamic response of portable electronic products and, particularly, solder joint interconnections is rapidly growing.

This chapter contains a brief review of the recent literature on the dynamic response of micro-electronic systems to shocks and vibrations. The extension part of the chapter has to do with the application of wires or wire-grid-arrays as a possible shock protection means in portable devices.

16.2. REVIEW

Goyal et al. [5] and Seah et al. [6] addressed stresses arising in printed circuit boards (PCB's), including those used in portable electronic products. The authors discussed also ways to protect these products from shocks.

It is the maximum acceleration (deceleration) that is usually viewed as an adequate criterion of the dynamic strength of a microelectronic component. It has been demonstrated, however [7], that, in some cases, such an approach could be misleading: a structural element that experiences high accelerations (decelerations) does not necessarily experience high dynamic stresses. For instance, a light and a short structural element that could be idealized as a single span beam simply supported at the ends can experience very high frequencies, when its supports are subjected to an impact loading or to continuous vibrations, and, hence, high accelerations (which are proportional to the frequency of vibrations squared), although the induced dynamic stresses could be very low. On the other hand, a structural element that could be idealized as a long cantilever beam with a heavy lump mass (electronic component) at its free end will experience relatively low frequencies of vibrations, when its support gets subjected to a shock load or to continuous vibrations. As a result of that, such a cantilever beam experiences low accelerations, although the dynamic stress at the clamped end of this structural element could be rather high.

Various dynamic systems subjected to periodic shock loads [8] including thin PCBs [9] were addressed by Suhir. Periodic (repetitive) loading could occur, for instance, during operation of some military systems subjected to excitations caused by rapidly shooting artillery or machine-gun shooting.

The dynamic response of a thin PCB subjected to a suddenly applied constant acceleration was addressed in [10]. Such loading can occur, for instance, during take-off of a jet-fighter or a space vehicle.

Various aspects of drop impact in application to portable electronics products were analyzed by Lim and Low [11], Lim et al. [12], Zhu and Marcinkiewicz [13], Luan and Tee [14].

The role of viscous damping in a single-degree-of-freedom system experiencing a drop impact was investigated in [15]. It has been shown that although elevated damping always leads to lower maximum displacements ("breaking distances"), it can result in accelerations (decelerations) that are significantly larger than the accelerations in a damping-free system. It has been demonstrated also that there is a certain limit to what could be achieved, as far as minimizing the "breaking distance" is concerned, by optimizing viscous damping in a single-degree-of-freedom system.

Shock tests are often used in addition to, or even instead of, drop tests for electronic and photonic equipment. Drop tests are usually more complicated than shock tests, especially when there is a need to measure dynamic accelerations and/or the induced stresses. In many cases, however, drop test conditions can be adequately substituted by shock tests, provided that these conditions are correctly predicted, and the shock tester is appropriately "tuned," in terms of the duration of the loading and maximum acceleration of the shock impulse [16].

Substantial improvement in shock protection of a portable electronic device can be achieved by employing a multi-degree of freedom system, with or without appreciable damping. This has been demonstrated [17] for a two-degree-of-freedom system ("box-in-a-box" type). It has been shown how to select the masses and spring constants of the structural elements, so that to avoid resonance conditions that might lead to a highly undesirable

"rigid impact" for the vulnerable element to be protected. Goyal et al. [18] evaluated the shock response spectrum of a shock protection system in a portable electronic product.

Many recent studies were devoted to the investigation of the effect of a short-term loading on the reliability of solder joints in portable products, with an emphasis on lead-free solders and on drop test conditions [19–35]. Ong et al. [19] addressed some general aspects of dynamic testing of materials in application to solder interconnections. Zhu [20], Sogo and Hara [21], Yi et al. [22], and Tan [23] examined the mechanical behavior of ball-grid-arrays (BGA) in chip-size packages (CSP). Arra [24] and Date et al. [25] investigated the behavior and reliability of lead-free solders under dynamic loading. Drop impact and drop test conditions were addressed also by Wu [26], Mishiro et al. [27], Wong et al. [28], Xie et al. [29], Yeh and Lai [30], Luan and Tee [30], Chiu et al. [32], and Syed et al. [33].

Predictive modeling and especially computer-aided evaluations (simulations) play an important role in the analysis and design of microelectronic materials and structures subjected to dynamic loading [34–49]. Dynamic response of some fiber-optics systems was addressed, based on predictive modeling, in [50–52].

Random vibrations are important in the evaluation of the dynamic response of micro- and opto-electronic systems to dynamic loading. Some general approaches to the analysis of random vibrations were described in [53]. Huang, Kececioglu and Prince [54] carried out a simplified analysis for portable electronic products subjected to random vibrations. It should be pointed out that random vibrations should be and will be widely employed to weed out infant mortalities in micro- and opto-electronic systems, as well as in their physical design and reliability evaluations. One of the major challenges is the selection of the width and the intensity of the input power spectrum: it should contain the frequencies of interest and should be, on one hand, strong enough to produce meaningful results, but, on the other hand, weak enough not to cause permanent damage.

16.3. EXTENSION: QUALITY OF SHOCK PROTECTION WITH A FLEXIBLE WIRE ELEMENTS

The analysis that follows uses analytical, rather than computer-aided, modeling [55]. We consider initially curved and compressed (as a result of the application of impact loading) wire-type elements (beams, rods) as suitable shock-protection means. Particularly, these means could be of nanoscale.

It has been shown [56,57], using examples of nonlinear springs with rigid or soft cubic characteristics of the restoring force, that non-linear springs, and particularly springs with soft characteristics of nonlinearity, offer certain advantages as effective shock protective elements. The wire-type structural elements behave, even when subjected to small deflections, as nonlinear springs with soft characteristics of the restoring force. We consider, as suitable idealized examples, cantilever wires with a lumped mass attached to their free ends. We evaluate the dynamic response of this mass to a drop impact applied to the wire base (substrate). An equivalent loading situation is shown schematically in Figure 16.1. We consider small deflections of initially curved wires, as well as large deflections of initially straight wires. We use the ratio of the initial velocity squared to the product of the maximum displacement and the maximum acceleration (deceleration) during the impact process as a suitable criterion (figure of merit) of the quality of the dynamic system, i.e., its ability to minimize the maximum displacement (the "breaking distance") and/or the maximum acceleration/deceleration.

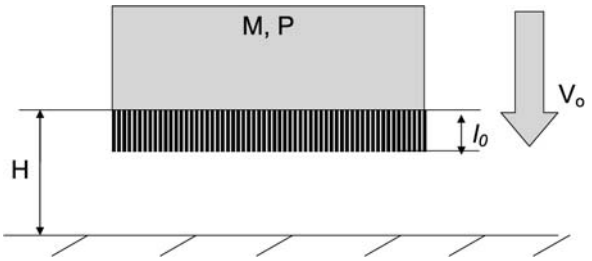


FIGURE 16.1. A nano-wire array (NWA) subjected to drop impact In the axial (through thickness) direction.

16.4. ANALYSIS

16.4.1. Pre-Buckling Mode: Small Displacements

16.4.1.1. Spring Constant of a Compressed Wire It is desirable (and practically inevitable) that a wire has an initial curvature, so that it performs as a spring, even at very small axial displacements (loading). Let a cantilever wire has an initial curvature

$$w_0(x) = f_0 \sin \frac{\pi x}{2l_0}, \tag{16.1}$$

where f_0 is the maximum initial deflection of the wire (at its free end), and l_0 is the wire's span, i.e., the distance measured along the x axis from the wire's tip to its clamped end. The origin of the coordinate, x, is at the wire's free end. In the analysis carried out in this section we assume that the compressive force, T, is appreciably smaller than the critical (Euler) force

$$T_e = \frac{\pi^2 EI}{4l_0^2}, \tag{16.2}$$

and that the induced axial displacement

$$\delta = l_0 - l_t \tag{16.3}$$

is significantly smaller than the initial, l_0, and final, l_t, wire spans. Typically such a linear approach is considered accurate enough, if the angle of rotation of the wire-cross section at its free end does not exceed 20°. In the Equation (16.2), EI is the flexural rigidity of the wire, E is Young's modulus of the wire material, and $I = \pi d^4/64$ is the moment of inertia of the cross-section for the case of a wire with a circular cross-section of the diameter d.

When the wire experiences small axial displacements, the induced deflections can be sought in a form similar to (16.1):

$$w_1(x) = f_1 \sin \frac{\pi x}{2l_t}, \tag{16.4}$$

where f_1 is the maximum induced deflection (at the free end of the wire).

The functions (16.1) and (16.4) satisfy the following zero boundary conditions

$$w_0(0) = w_1(0) = 0,\ w_0''(0) = w_1''(0) = 0,\ w_0'(l_0) = w_1'(l_t) = 0,\ w_0'''(l_0) = w_1'''(l_t) = 0. \tag{16.5}$$

The first conditions indicate that the deflection at the origin is zero for both the stress-free and deflected wire. The second conditions indicate that there is no bending moment at the free end, and therefore the corresponding curvatures are zero. The third conditions indicate that the rotation angle at the clamped end must be zero. The fourth conditions indicate that there are no lateral forces at the clamped end.

Since the axial displacement is small compared to the wire spans prior to, and after, the application of the compressive force, the final span, l_t, in the denominator of the argument in the Equation (16.4) can be substituted by the initial span l_0:

$$w_1(x) = f_1 \sin \frac{\pi x}{2l_0}. \tag{16.6}$$

The relationship between the initial, f_0, and the force induced, f_1, deflections can be established using the following equilibrium equation (equation of bending)

$$EIw_1''(x) + T[w_1(x) + w_0(x)] = 0. \tag{16.7}$$

This equation simply states that the elastic (internal) bending moment expressed by the first term in (16.7) should be equal, for each cross-section, to the external bending moment due to the force, T. Introducing the Equation (16.1) and the sought solution (16.6) into the Equation (16.7), we obtain the following formula for the maximum induced deflection, f_1:

$$f_1 = \frac{f_0}{\dfrac{T_e}{T} - 1}, \tag{16.8}$$

were T_e is the critical force given by the Equation (16.2). The total maximum deflection can be found as

$$f_t = f_0 + f_t = \frac{f_0}{1 - \dfrac{T}{T_e}}, \tag{16.9}$$

and the total deflection function, which is due to both the initial and the induced curvatures, is

$$w_t(x) = f_t \sin \frac{\pi x}{2l_t}. \tag{16.10}$$

The length, s_t, of the compressed and deflected wire can be evaluated, for small enough curvatures, as

$$\begin{aligned} s_t &= \int_0^{l_{ty}} \sqrt{1 + [w_t'(x)]^2}\,dx \cong \int_0^{l_t} \left[1 + \frac{1}{2}(w_t'(x))\right]^2 dx = l_t + \frac{1}{2}\int_0^{l_t} \left[w_t'(x)\right]^2 dx \\ &= l_t + \frac{1}{2}\left(\frac{\pi f_t}{2l_t}\right)^2 \int_0^{l_t} \cos^2 \frac{\pi x}{2l_t}\,dx = l_t + \left(\frac{\pi}{4}\right)^2 \frac{f_t^2}{l_t} \cong l_t + \left(\frac{\pi}{4}\right)^2 \frac{f_t^2}{l_0}. \end{aligned} \tag{16.11}$$

Similarly, the length, s_0, of the stress-free wire can be found as

$$s_0 = l_0 + \delta_0, \tag{16.12}$$

where

$$\delta_0 = \left(\frac{\pi}{4}\right)^2 \frac{f_0^2}{l_0} \tag{16.13}$$

is the difference between the total length of the wire and its span. From (16.11) and (16.12) we find that the difference between the initial total length of the wire and its stress-induced length is:

$$s_0 - s_t = \delta - \delta_0\left(\frac{f_t^2}{f_0^2} - 1\right) = \delta - 2\delta_0 \frac{T}{T_e}\frac{1-(T/2T_e)}{[1-(T/T_e)]^2}. \tag{16.14}$$

As evident from this formula, the difference in the total lengths of the wire prior to, and after the application of the axial force, T, can be computed as the difference between the total axial displacement and the axial displacement due to bending. On the other hand, the difference between the total lengths of the wire prior to and after the application of the compressive loading is due to the elastic axial contraction of the wire. This contraction can be found on the basis of the Hooke's law as follows:

$$s_0 - s_t = \frac{Ts_0}{E\dfrac{\pi d^2}{4}} \cong \frac{Tl_0}{E\dfrac{\pi d^2}{4}} = \left(\frac{\pi d}{8l_0}\right)^2 l_0 \frac{T}{T_e}, \tag{16.15}$$

where the Equations (16.2) and (16.4) are considered. Equating the right parts of the Equations (16.14) and (16.15) we obtain the following formula for the axial displacement:

$$\delta = 2\delta_0 \frac{T}{T_e}\left[\frac{1-\dfrac{T}{2T_e}}{\left(1-\dfrac{T}{T_e}\right)^2} + \left(\frac{d}{f_0}\right)^2\right]. \tag{16.16}$$

The first term in the brackets in this formula is due to bending and the second term is due to the axial elastic compression. The first term is the smallest at $T = 0$ and is equal to 1. Hence, if the d/f_0 ratio is significantly smaller than 1, then the second term in the brackets does not have to be considered, the s_t value is not different than s_0, and the axial displacement can be evaluated, taking into account the bending deformations only:

$$\delta = 2\delta_0 \frac{T}{T_e}\frac{1-(T/2T_e)}{[1-(T/T_e)])^2}. \tag{16.17}$$

It is this case that will be considered in the further analysis. Then the axial spring constant can be found as

$$K_T = \frac{T}{\delta} = K_0 \chi\left(\frac{\delta}{\delta_0}\right), \tag{16.18}$$

where

$$K_0 = \frac{T_e}{2\delta_0} = \frac{2EI}{f_0^2 l_0} \tag{16.19}$$

is the initial spring constant (prior to the application of the compressive force), and the factor

$$\chi(\delta/\delta_0) = 2\frac{\delta_0}{\delta}\left(1 - \frac{1}{\sqrt{1+\dfrac{\delta}{\delta_0}}}\right) \tag{16.20}$$

considers the effect of the axial displacement. As evident from this formula, the axial spring "constant" of the wire rapidly decreases with an increase in the induced displacement: the initial spring constant decreases by the factor of three, when the induced axial displacement is three times larger than the δ_0 value.

16.4.1.2. Equation of Motion Let a single-degree-of-freedom system with the mass M, which is protected by a spring element of the type considered in the previous sections, be dropped from the height H on a non-deformable (hard) floor. If the mass M is supported by a cantilever wire or an array of such wires, then the equation of motion of the system can be written as

$$\ddot{\delta}(t) + \frac{T_e}{M}\left(1 - \frac{1}{\sqrt{1+\dfrac{\delta(t)}{\delta_0}}}\right) = g, \tag{16.21}$$

where T_e is the critical force and g is the acceleration due to gravity. Since $\ddot{\delta} = \dot{\delta}(d\dot{\delta}/d\delta)$ and

$$\int_0^{\delta} \frac{d\delta}{\sqrt{1+\dfrac{\delta}{\delta_0}}} = 2\delta_0\left(\sqrt{1+\frac{\delta}{\delta_0}} - 1\right), \tag{16.22}$$

the Equation (16.21) has the following first integral:

$$\dot{\delta}^2 + 2\frac{T_e}{M}\left[\delta - 2\delta_0\left(-1 + \sqrt{1+\frac{\delta}{\delta_0}}\right)\right] = 2g(\delta + H) \cong 2gH. \tag{16.23}$$

16.4.1.3. Maximum Displacement and Maximum Compressive Force The maximum displacement can be found from (16.23) by putting $\dot{\delta} = 0$. This yields:

$$\delta_{\max} = \delta_0\left[\left(1 + \sqrt{\frac{MgH}{T_e\delta_0}}\right)^2 - 1\right]. \tag{16.24}$$

Solving the Equation (16.17) for the T/T_e ratio and using the Equation (16.24), we obtain the following formula for the ratio of the maximum force to the critical force:

$$\frac{T_{\max}}{T_e} = \frac{1}{1+\sqrt{\dfrac{T_e\delta_0}{MgH}}}, \tag{16.25}$$

where $T_{\max} = T|_{\delta=\delta_{\max}}$ is the value of the compressive force that corresponds to the maximum axial displacement.

16.4.1.4. Maximum Acceleration (Deceleration) and System's Quality Introducing the Equation (16.24) into the Equation (16.21), we obtain the following formula for the maximum acceleration (deceleration):

$$\ddot{\delta}_{\max} = g - \frac{T_e}{M}\frac{1}{1+\sqrt{\dfrac{T_e\delta_0}{MgH}}} = g - \frac{T_{\max}}{M} \cong -\frac{T_{\max}}{M}. \tag{16.26}$$

Since both the maximum displacement and the maximum acceleration (deceleration) should be made as small (low) as possible, the ability of a dynamic system to provide the most effective protection to the vulnerable mass M could be assessed on the basis of the following dimensionless parameter (quality):

$$Q = -\frac{2gH}{\delta_{\max}\ddot{\delta}_{\max}} = 2\frac{1+\sqrt{MgH/(T_e\delta_0)}}{2+\sqrt{MgH/(T_e\delta_0)}} = \frac{1}{1-\dfrac{1}{2}\dfrac{T_{\max}}{T_e}}. \tag{16.27}$$

In a linear system with the natural frequency ω_0 the maximum displacement and the maximum acceleration are $\delta_{\max} = \sqrt{2gH}/\omega_0$ and $\ddot{\delta}_{\max} = \omega_0\sqrt{2gH}$, respectively, so that the quality of a linear system is always 1 and is independent of the system's characteristics. As to the quality of an ideal/perfect protection system, it is as high as 2. Indeed, such system should be characterized by a constant acceleration (deceleration) that is "turned on" at the moment of the beginning of breaking and is turned off at the moment when the velocity of the mass to be protected becomes equal to zero. Then the corresponding breaking distance can be found as

$$\delta_{\max} = -\frac{v_0^2}{2\ddot{\delta}_{\max}}, \tag{16.28}$$

where v_0 is the initial velocity. From this equation, using notation (16.27), we find $Q = 2$.

The quality of a nonlinear system can be, as evident from the Equation (16.27), considerably higher than 1. For the maximum force value approaching the critical force, the system's quality can be almost twice as high as the quality of a linear system, i.e., will not be very far remote from the ideal situation.

16.4.1.5. Numerical Example Let a weight of $P = 0.1$ kg is dropped from the height $H = 1.5$ m (its potential energy is $MgH = 0.15\,\text{kg} \times \text{m} = 150\,\text{kg} \times \text{mm}$) and is protected by an array of $l_0 = 100$ μm $= 0.1$ mm long cantilever nano-wires of the diameter $d = 200\,\text{nm} = 0.0002\,\text{mm}$ (Figure 16.1). Young's modulus of the material is

$E = 20000\,\text{kg/mm}^2$. The maximum initial deflections of the stress-free nano-wires are $f_0 = 0.01$ mm, i.e., significantly larger than the wire diameter. The objective of the calculation that follows is to establish the required number of the nano-wires to provide adequate shock protection for the mass M.

The flexural rigidity of a single nano-wire is

$$EI = \frac{\pi d^4}{64} = 20000 \times \frac{3.14159 \times 0.0002^4}{64} = 1.5708 \times 10^{-12}\ \text{kg mm}^2.$$

Its critical force for a single nano-wire is

$$T_e = \frac{\pi^2 EI}{4l_0^2} = \frac{3.14159^2 \times 1.5708 \times 10^{-12}}{4 \times 0.1^2} = 3.8758 \times 10^{-10}\ \text{kg}.$$

Let us select the maximum force that the nano-wire will experience as, say, 75% of the critical force, i.e., $T_{\max} = 2.9068 \times 10^{-10}$ kg. This corresponds to the "quality" $Q = 1/[1 - (0.75/2)] = 1.6$. The δ_0 value for the nano-wires is

$$\delta_0 = \left(\frac{\pi}{4}\right)^2 \frac{f_0^2}{l_0} = \left(\frac{3.14159}{4}\right)^2 \frac{0.01^2}{0.1} = 0.00061685\ \text{mm}.$$

Then the Equation (16.30) suggests that a single nano-wire is able to effectively provide protection to an external energy of

$$MgH = \frac{T_e \delta_0}{\left(\dfrac{T_e}{T_{\max}} - 1\right)^2} = \frac{3.8758 \times 10^{-10} \times 0.00061685}{\left(\dfrac{1}{0.75} - 1\right)^2} = 2.1517 \times 10^{-12}\ \text{kg mm}.$$

Since, however, the potential energy that should be 'fought against" is as high as 150 kg mm, the required number of the nano-wires that should be grown (employed, manufactured) is as high as

$$\frac{150}{2.1517 \times 10^{-12}} \cong 7 \times 10^{13}.$$

The maximum acceleration (deceleration) is also high:

$$\ddot{\delta}_{\max}/g = \frac{T_{\max}}{P} = \frac{2.9068 \times 10^{-10} \times 7 \times 10^{13}}{0.1} = 203476.$$

The maximum displacement ("breaking distance") is

$$\delta_{\max} = \delta_0 \left[\frac{1}{\left(1 - \dfrac{T_{\max}}{T_e}\right)^2} - 1 \right] = 0.00061685 \left[\frac{1}{(1 - 0.75)^2} - 1 \right] = 0.009253\ \text{mm}$$
$$= 9.253\ \mu\text{m},$$

i.e., about 9% of the wire span.

16.4.2. Post-Buckling Mode: Large Displacements

16.4.2.1. Spring "Constant" of a Compressed Wire In the analysis that follows we assume that the initial curvature, if any, has a negligible effect on the large deflections of a cantilever beam and need not be accounted for. We assume also that the elastic contraction of the wire is small and does not have to be considered.

In the well-known Euler's "elastica" solution for large deflections of a cantilever beam subjected to compression (see, for instance, [58,59]), the relationship between the applied force, T, and the response of the beam is established by the following relationship:

$$k = \sqrt{\frac{T}{EI}} = \frac{K(p)}{l}, \tag{16.29}$$

where l is the length of the wire,

$$K(p) = \int_0^{\frac{\pi}{2}} \frac{d\varphi}{\sqrt{1 - p^2 \sin^2 \varphi}} \tag{16.30}$$

is the complete elliptic integral of the first kind, $p = \sin(\alpha/2)$ is the module of the elliptic function, and α is the angle of rotation of the free end of the wire. The "elastica" solution gives also the following expression for the "span" of the buckled wire:

$$l_t = \frac{2}{k} E(p) - l, \tag{16.31}$$

where

$$E(p) = \int_0^{\frac{\pi}{2}} \sqrt{1 - p^2 \sin^2 \varphi} d\varphi \tag{16.32}$$

is the complete elliptic integral of the second kind. The axial displacement of the wire tip can be found as

$$\delta = l - l_t = 2l - \frac{2}{k} E(p) = l\Psi(p), \tag{16.33}$$

where the function

$$\Psi(p) = 2\left[1 - \frac{E(p)}{K(p)}\right] \tag{16.34}$$

is the loading-dependent ratio of the axial displacement to the wire length. The functions $K(p)$ and $E(p)$ can be represented as series (see, for instance, [60–62]):

$$K(p) = \frac{\pi}{2} \sum_{n=0}^{\infty} \frac{(2n)!(2n)!}{2^{4n}(n!)^4} p^{2n}, \qquad E(p) = \frac{\pi}{2}\left[1 - \sum_{n=1}^{\infty} \frac{(2n-2)!(2n)!}{2^{4n-1}(n-1)!(n!)^3} p^{2n}\right]. \tag{16.35}$$

For small enough p values, i.e., for small deflections,

$$K(p) \approx \frac{\pi}{2}\left(1 + \frac{p^2}{4}\right), \qquad E(p) \approx \frac{\pi}{2}\left(1 - \frac{p^2}{4}\right), \tag{16.36}$$

so that

$$\Psi(p) \approx p^2. \tag{16.37}$$

The spring "constant" can be found as

$$K_T = \frac{T}{\delta}, \tag{16.38}$$

where δ is given by the formula (16.33) and the applied force, T, is expressed through the parameter p value as

$$T = k^2 EI = \left(\frac{2}{\pi}K(p)\right)^2 T_e, \tag{16.39}$$

where T_e is the critical force. If N cantilever wires are used then the critical force should be multiplied by N.

16.4.2.2. Equation of Motion The equation of motion is

$$\ddot{\delta}(t) + \frac{T_e}{M}\left(\frac{2}{\pi}K(p)\right)^2 = g, \tag{16.40}$$

where g is the acceleration due to gravity. Since $\ddot{\delta}(t) = \dot{\delta}(d\dot{\delta}/d\delta)$ and $d\delta(t) = l(d\Psi(p)/dp)dp$, the Equation (16.40) has the following first integral:

$$\dot{\delta}^2 + 2\frac{T_e}{M} l\chi(p) \approx 2gH, \tag{16.41}$$

where the function

$$\chi(p) = \int_0^p \left[\frac{2}{\pi}K(p')\right]^2 \frac{d\Psi(p')}{dp'} dp' \tag{16.42}$$

considers the effect of the finite level of loading. This function is equal to zero for $\alpha = 0$ (or $p = 0$), and increases with an increase in the p value. Indeed, for small p values the Equations (16.36) and (16.37) yield: $K(p) \approx 1$ and $d\Psi/dp = 2p$, and therefore $\chi(p) \approx p^2$. The calculated values of the function $\chi(p)$ are given in Table 16.1.

16.4.2.3. Maximum Displacement and Maximum Compressive Force The maximum value $p_{\max}$ of the parameter p can be found from the Equation (16.41) by putting the velocity $\dot{\delta}(t)$ equal to zero. This leads to the equation:

$$\chi(p_{\max}) = \chi_0 = \frac{MgH}{T_e l}. \tag{16.43}$$

TABLE 16.1.

α	0	20°	40°	60°	80°	100°	120°
p	0	0.342	0.643	0.866	0.985	0.985	0.866
$K(p)$	$\pi/2$	1.620	1,786	2.156	3.150	3.153	2.156
$d\psi/dp$	0	0.706	1.483	2.630	8.020	8.020	−2.630
$\left(\frac{2}{\pi}K(p)\right)^2\frac{d\Psi(p)}{dp}$	0	0.751	1.917	4.955	32.252	32.313	−4.955
$\chi(p)$	0	0.128	0.345	1.111	3.325	3.325	4.953

After the maximum deformation, which is determined by the $p_{\max}$ value, is found, then the maximum displacement can be evaluated as

$$\delta_{\max} = l\Psi(p_{\max}) = 2l\left[1 - \frac{E(p_{\max})}{K(p_{\max})}\right], \tag{16.44}$$

and the maximum compressive force can be determined as

$$T_{\max} = T_e\left[\frac{2}{\pi}K(p_{\max})\right]^2. \tag{16.45}$$

For small displacements, when $\chi(p_{\max}) = \Psi(p_{\max}) = p_{\max}^2$, and

$$K(p_{\max}) \approx \frac{\pi}{2}\left(1 + \frac{p_{\max}^2}{4}\right),$$

the Equations (16.43), (16.44) and (16.45) yield:

$$p_{\max} = \sqrt{\chi_0}, \qquad \delta_{\max} = lp_{\max}^2 = l\chi_0 = \frac{MgH}{T_e}, \qquad T_{\max} \approx T_e. \tag{16.46}$$

16.4.2.4. Maximum Acceleration (Deceleration) and System's Quality Maximum acceleration (deceleration) can be found from the equation of motion (16.40) and the maximum compressive force value determined by the Equation (16.45) as follows:

$$\ddot{\delta}_{\max} = -\frac{T_{\max}}{M}. \tag{16.47}$$

The system's quality is

$$Q = -\frac{2gH}{\delta_{\max}\ddot{\delta}_{\max}} = \frac{2MgH}{\delta_{\max}T_{\max}} = \frac{2\chi_0}{\frac{\delta_{\max}}{l}\frac{T_{\max}}{T_e}}. \tag{16.48}$$

For small displacements, using the Equation (16.46), we obtain $Q \approx 2$. For large displacements, however, the quality will be substantially lower.

16.4.2.5. Numerical Examples The evaluation of the operational characteristics of the wire(s) can be carried out in the following sequence:

- For the given mass to be protected, the initial velocity (drop height), the buckling force and the wire length evaluate the parameter χ_0 by the Equation (16.43).
- Find the $p_{\max}$ value from the Table 16.1.
- Calculate the maximum displacement and the maximum force on the basis of the Equations (16.44) and (16.45), respectively.
- Calculate the maximum acceleration (deceleration) on the basis of the Equation (16.47) and the corresponding "quality" of the system by the Equation (16.48).

Let, for instance, the calculated value of the parameter χ_0 be $\chi_0 = 0.345$. Then

$$\alpha = 40^0, \qquad p_{\max} = 0.643, \qquad \frac{\delta_{\max}}{l} = \Psi(p_{\max}) = 0.440,$$

$$K(p_{\max}) = 1.786, \qquad \frac{T_{\max}}{T_e} = \left(\frac{2}{\pi} \times 1.786\right)^2 = 1.2928,$$

and, with $\chi_0 = 0.345$, the Equation (16.45) yields: $Q = 2 \times 0.345/(0.440 \times 1.2928) = 1.213$. This value is much lower than the next-to-ideal quality of 2 that can be achieved in the close-to-buckling condition.

For $\chi_0 = 1.111$, we obtain:

$$\alpha = 60^0, \qquad p_{\max} = 0.866, \qquad \frac{\delta_{\max}}{l} = \Psi(p_{\max}) = 0.880,$$

$$K(p_{\max}) = 2.156, \qquad \frac{T_{\max}}{T_e} = \left(\frac{2}{\pi} \times 2.156\right)^2 = 1.8839,$$

$$Q = \frac{2 \times 1.111.}{0.880 \times 1.8839} = 1.340.$$

This value is a little higher than the previously obtained one (because of somewhat lower accelerations, despite higher displacements), is still better than what could be achieved by a linear spring, but is much lower than two.

16.5. CONCLUSIONS

The following conclusions can be drawn from the performed analysis:

- A convenient criterion of a quality of a shock protection system is suggested;
- A simple calculation procedure (predictive model) has been developed for the evaluation of the shock protection capabilities of a cantilever wire array;
- Appreciable gain can be achieved, in comparison with linear springs (cushions), by employing wire elements subjected to compression and operated in the pre-buckling and close-to-buckling mode;
- The design with a flexible wire or a wire array as a shock protective means should be carried out in such a way that the wire(s) perform as close-to-buckling condition as possible, preferably in the pre-buckling mode. Pre-buckling mode should be preferred to a post-buckling mode (provided that the stress condition in the wire(s) is close to the buckling condition), both because in the post-buckling mode the quality

of the shock protection might bee too low, and because of a possibility of a rigid impact, when the induced displacements are significant. The operation of the wires in a "deep" post-buckling condition (i.e., at the deformations exceeding substantially the deformations-at-buckling) is not recommended: large displacements not only result in a low quality value, but, more importantly, if the initial impact energy exceeds the available elastic energy of the wire spring, can lead to an undesirable "rigid impact". Indeed, if the initial kinetic energy of the mass to be protected exceeds the available strain energy of the wire or the wire array, a highly undesirable "rigid impact" becomes possible. This possibility of such a phenomenon, which can occur in a spring with a soft characteristic of nonlinearity, should be taken into account and thoroughly avoided in the wire design and in the specified operation conditions.

REFERENCES

1. D. Steinberg, Vibration Analysis for Electronic Equipment, 2nd edn, John Wiley, 1988.
2. E. Suhir, Dynamic response of microelectronics and photonics systems to shocks and vibrations, International Conference on Electronic Packaging, INTERPack'97, Hawaii, June 15–19, 1997.
3. JEDEC Standard JESD22-B104-B, Mechanical Shock, 2001.
4. JEDEC Standard JESD22-B111, Board Level Drop Test Method of Components for Handheld Electronic Products, 2003.
5. S. Goyal, E.K. Buratynski, and G.W. Elko, Shock-protection suspension design for printed circuit board, Proceedings of SPIE, 4217, 2002.
6. S.K.W. Seah, C.T. Lim, E.H. Wong, V.B.C. Tan, and V.P.W. Shim, Mechanical response of PCBs in portable electronic products during drop impact, Proceedings 4th Electronics Packaging Technology Conference (EPTC 2002), Singapore, 10–12 Dec. 2002.
7. E. Suhir, Is the maximum acceleration an adequate criterion of the dynamic strength of a structural element in an electronic product?, IEEE CPMT Transactions, Part A, 20(4) (1997).
8. E. Suhir, Linear and nonlinear vibrations caused by periodic impulses, AIAA/ASME/ASCE/AHS 26th Structures, Structural Dynamics and Materials Conference, Orlando, FL, April 1985.
9. E. Suhir, Response of a flexible printed circuit board to periodic shock loads applied to its support contour, ASME Journal of Applied Mechanics, 59(2) (1992).
10. E. Suhir, Nonlinear dynamic response of a flexible thin plate to a constant acceleration applied to its support contour, with application to printed circuit boards used in avionic packaging, Int. Journal of Solids and Structures, 29(1) (1992).
11. C.T. Lim and Y.J. Low, Investigating the drop impact of portable electronic products, 52nd ECTC, 2002, Proceedings, 28–31 May 2002.
12. C.T. Lim, C.W. Ang, L.B. Tan, S.K.W. Seah, and E.H. Wong, Drop impact survey of portable electronic products. 53rd ECTC, IEEE, May 2003.
13. L. Zhu and W. Marcinkiewicz, Drop impact reliability analysis of CSP packages at board and product system levels through modeling approaches, The Ninth Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems, IEEE, Part Vol. 2, 2004.
14. J.E. Luan and T.Y. Tee, Analytical and numerical analysis of impact pulse parameters on consistency of drop impact results. The 6th Electronics Packaging Technology Conference, IEEE Cat. No. 04EX971, 8–9 December 2004.
15. E. Suhir, Dynamic response of a one-degree-of-freedom linear system to a shock load during drop tests: effect of viscous damping, IEEE CPMT Transactions, Part A, 19(3) (1996).
16. E. Suhir, Could shock tests adequately mimic drop test conditions? 52nd ECTC Proc., 2002.
17. E. Suhir and R. Burke, Dynamic response of a rectangular plate to a shock load, with application to portable electronic products, IEEE Trans. Components, Packaging, and Manufacturing Technology, Part B: Advanced Packaging, 17(3) (1994).
18. S. Goyal, J.M. Papadopoulos, and P.A. Sullivan, Shock protection of portable electronic products: shock response spectrum, damage boundary approach, and beyond, Shock and Vibration, 4(3) (1997).
19. K.C. Ong, et al., Dynamic materials testing and modeling of solder interconnects, 54th ECTC, 2004.

20. L. Zhu, Submodeling technique for BGA reliability analysis of CSP packaging subjected to an impact loading, InterPACK Conference Proc., 2001.
21. T. Sogo and S. Hara, Estimation of fall impact strength for BGA solder joints, ICEP Conference Proc., Japan, 2001.
22. Q. Yu, H. Kukuichi, S. Ikeda, M. Shiratori, M. Kakino, and N. Fujiwara, Dynamic behavior of electronics package and impact reliability of BGA solder joints, Intersociety Conf. on Thermal Phenomena, 2002.
23. L.B. Tan, Board level solder joint failure by static and dynamic loads, Proc. 5th EPTC, 2003.
24. M. Arra, et al., Performance of lead free solder joints under dynamic mechanical loading, Proc. 52nd ECTC, 2002.
25. M. Date, et al., Ductile-to-brittle transition in Sn-Zn solder joints measured by impact tests, Scripta Materialia, 51 (2004).
26. J. Wu, Global and local coupling analysis for small components in drop simulation, 6th International LS-DYNA Users Conference Proc., 2000.
27. K. Mishiro, et al., Effect of the drop impact on BGA/CSP package reliability, Microelectronics Reliability, 42 (2002).
28. E.H. Wong, C.T. Lim, N. Lee, S.K.W. Seah, C. Hoe, and J. Wang, Drop impact test—mechanics and physics of failure, 4th EPTC Proc., Singapore, 2002.
29. D. Xie, et al., Solder joint behavior of area array packages in board-level drop for hand held devices, 53rd ECTC, 2003.
30. C.-L. Yeh and Y.-S. Lai, Effect of solder joint shapes on free drop reliability of chip-scale packages, Proc. IMAPS Taiwan Tech. Symp., 2004.
31. J.E. Luan and T.Y. Tee, Effect of impact pulse on dynamic responses and solder joint reliability of TFBGA packages during board level drop test, 6th EMAP Conference, Malaysia, Dec. 2004.
32. T.C. Chiu, et al., Effect of thermal aging on board level drop reliability for Pb-free BGA package, 54th ECTC 2004.
33. A. Syed, et al., A methodology for drop performance prediction and application for design optimization of chip scale packages, 55th ECTC, 2005.
34. J. Wu, G. Song, C.-P. Yeh, and K. Wyatt, Drop/impact simulation and test validation of telecommunication products, Thermomechanical Phenomena in Electronic Systems, Proceedings of the Intersociety Conference, 1998, Conference: Proceedings of the 1998 6th Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems, ITHERM, May 27–30 1998, pp. 330–336.
35. L. Zhu, Submodeling technique for BGA reliability analysis of CSP packaging subjected to an impact loading, InterPACK Conference Proc., 2001.
36. C.T. Lim, Y.M. Teo, and V.P.W. Shim, Numerical simulation of the drop impact response of a portable electronic product, IEEE CPMT Transactions, 25(3) (2002).
37. T.Y. Tee, H.S. Ng, C.T. Lim, E. Pek, and Z.W. Zhong, Application of drop test simulation in electronic packaging, 4th ASEAN ANSYS Conf., 2002.
38. L. Xu, et al., Numerical studies of the mechanical response of solder joints to drop/impact load, Proc. EPTC 2003.
39. T.Y. Tee, H.S. Ng, C.T. Lim, E. Pek, and Z.W. Zhong, Board level drop tests and simulation of TFBGA packages for telecommunication applications, 53rd ECTC Proc., May 2003.
40. L. Zhu, Modeling technique for reliability assessment of portable electronic product subjected to drop impact loads, 53rd ECTC, pp. 100–104, 2003.
41. Y.Q. Wang, Modeling and simulation of PCB drop test, Proc. 5th EPTC, 2003, pp. 263–268.
42. J.E. Luan, T.Y. Tee, E. Pek, C.T. Lim, and Z.W. Zhong, Modal analysis and dynamic responses of board level drop test, 5th EPTC Conference4 Proc., Singapore, 2003.
43. L. Zhu, and W. Marcinkiewicz, Drop impact reliability analysis of CSP packages at board and product system levels through modeling approaches, Intersociety Conference on Thermal and Thermo-Mechanical Phenomena, 2004, pp. 296–303.
44. S. Irving, and Y. Liu, Free drop test simulation for portable IC package by implicit transient dynamics FEM, Proceedings Electronic Components and Technology Conference, Vol. 1, Proceedings 54th Electronic Components and Technology Conference, 2004.
45. T.Y. Tee, J.E. Luan, E. Pek, C.T. Lim, and Z.W. Zhong, Novel numerical and experimental analysis of dynamic responses under board level drop test, EuroSime Conference Proc., 2004.
46. T.Y. Tee, J.E. Luan, E. Pek, C.T. Lim, and Z.W. Zhong, Advanced experimental and simulation techniques for analysis of dynamic responses during drop impact, 54th ECTC Proc., 2004, pp. 1088–1094.
47. C.L. Yeh and Y.S. Lai, Transient simulation of solder joint fracturing under impact test, 6th EPTC, Singapore, 8–10 Dec. 2004.

48. P. Marjamaki, T. Mattila, and J. Kivilahti, FEA of lead-free drop test boards, 55th ECTC Proc., 2005.
49. E. Suhir, Predicted fundamental vibration frequency of a heavy electronic component mounted on a printed circuit board, ASME Journal of Electronic Packaging, 122(1) (2000).
50. E. Suhir, Free vibrations of a fused biconical taper lightwave coupler, Int. Journal of Solids and Structures, 29(24) (1992).
51. E. Suhir, Vibration frequency of a fused biconical taper (FBT) lightwave coupler, IEEE/OSA Journal of Lightwave Technology, 10(7) (1992).
52. E. Suhir, Elastic stability, free vibrations, and bending of optical glass fibers: the effect of the nonlinear stress-strain relationship, Applied Optics, 31(24) (1992).
53. E. Suhir, Applied Probability for Engineers and Scientists, McGraw-Hill, 1997.
54. W. Huang, D.B. Kececioglu, and J.L. Prince, A simplified random vibration analysis on portable electronic products, IEEE Transactions on Components & Packaging Technologies, 23(3) (2000).
55. E. Suhir, Analytical modeling in structural analysis for electronic packaging: its merits, shortcomings and interaction with experimental and numerical techniques, ASME Journal of Electronic Packaging, 111(2) (1989).
56. E. Suhir, Shock protection with a nonlinear spring, IEEE CPMT Transactions, Advanced Packaging, Part B, 18(2) (1995).
57. E. Suhir, Shock-excited vibrations of a conservative duffing oscillator with application to shock protection in portable electronics, Int. Journal of Solids and Structures, 33(24) (1996).
58. S.P. Timoshenko and J.M. Gere, Theory of Elastic Stability, 2nd edn, McGraw-Hill, 1988.
59. E. Suhir, Structural Analysis in Microelectronics and Fiber Optics, Van-Nostrand, 1997.
60. P. Appell and E. Lacour, Principes de la Theorie des Functions Elliptiques et Applications, 2nd edn, Paris, 1922 (in French).
61. F. Oberhettinger and W. Magnus, Anwendung der elliptischen Funktionen in Physiknund Technik, Springer Verlag, 1949 (in German).
62. J. Spanier and K.B. Oldham, An Atlas of Functions, Hemisphere Publ. Corp., 1987.

17

Dynamic Physical Reliability in Application to Photonic Materials

Dov Ingman[a], Tatiana Mirer[b], and Ephraim Suhir[c]

[a]*Technion, Israel Institute of Technology, Haifa, Israel*
[b]*Technion, Israel Institute of Technology, Quality Assurance and Reliability, Haifa, Israel*
[c]*University of California, Santa Cruz, CA, University of Maryland, College Park, MD, and ERS/Siloptix Co., Los Altos, CA, USA*

17.1. INTRODUCTION: DYNAMIC RELIABILITY APPROACH TO THE EVOLUTION OF SILICA FIBER PERFORMANCE

The widespread application of fiber light-guides in optical communication systems and in diverse physical field transducers has enhanced interest in their mechanical reliability (see, for instance, [1,28,35,36]). The latter is defined as the probability of the fiber rupture or cracking under load.

A defect-free silica glass fiber possesses higher strength than even a steel wire of the same diameter; however, a minimal contact of the silica core surface with a hard object or even with a dust particle gives rise to microcracks. These begin growing rapidly in the presence of moisture and on the application of a tensile load. As a result, the tensile strength of the silica fibers turns out to be much lower than its theoretical strength.

Silica glass is a brittle material, i.e., is a material that does not exhibit any appreciable plasticity under deformation. Reliability prognostication of the mechanical behavior of such materials causes significant challenges [1,24,29,35]. This is because of the high strength scatter and the absence of any preliminary "symptom" (such as, say, plastic deformation) preceding rupture.

As the microcracking process is random, the most suitable model for describing the fiber strength should be statistical [21,22,25,30,34]. The existing fracture mechanics based predictive models use mathematical/mechanical apparatus without any reference to the statistics of the crack propagation and statistical characteristics of the time to failure [23,26,27, 32,33]. The traditionally used combinations of two Weibull distributions are essentially static and do not account for the dynamic nature of the material strength deterioration, which is typically inherent to material under an external or an internal load. The equation for the evolution of the reliability distribution function, described in Section 17.1.1, is applied in our analysis in conjunction with a proposed new physical model of strength deterioration. This model is intended to adequately describe the process of damage accumulation.

The mechanism of the propagation of surface flaws in silica optical fibers is governed by a combination of stress, moisture, temperature, and time. In Section 17.1.2 we suggest a predictive model for time-to-failure when the material is subjected to the action of a three-dimensional mechanical-temperature-humidity load. The model is based on the dynamic physical damage accumulation model described in Section 17.1.1.

The statistical distribution of the tensile strength of optical silica fibers exhibits bimodality. This phenomenon is usually attributed to two different types of flaws in the fiber material. Section 17.1.3 addresses the effect of the bimodality based on an approach that uses the suggested model. We introduce a new micromechanical model of damage accumulation. This model is related to the parameters of the damage evolution equation.

17.1.1. Dynamic Physical Model of Damage Accumulation

The recently proposed dynamic reliability theory for the damage accumulation processes [2,3] is based on the energy aspect of the stress/element interaction. This interaction is reflected in the transitions between the consequent damage states represented by a Markovian stochastic process. The Markovian process describes the dynamic evolution of the system in terms of the change in its reliability distribution function over the range of sequential damage states. The interaction of the loaded specimen with the environment is treated in our approach as a combination of two distinct types. The first type is of a "potential" character determined by the elastic energy of the specimen as a whole, or, alternatively, by the probability density of the elastic energy distribution in the material. This probability is characterized by the likelihood of the material (its ability) to return to the previous damage state after the removal of the applied stress. The second type is of a "local" nature, which is characterized by the energy-dissipative mechanisms of microstructural rearrangement during the deformation process. This type of the interaction of the loaded specimen with the environment is always accompanied by the energy consumption and local structural changes, such as initiation and propagation of microvoids and microcracks. The accumulation process is considered as one of transfer by transition. This process is characterized by elastic potential curves that correspond to different damage states of the material. The transitions are due to the intrinsically stochastic mechanisms of formation and growth of micro-cavities and microcracks: vacancy diffusion in dislocation annihilation, climb-up, etc. Such transitions require local delivery of sufficient energy for the work of rupture of atomic bonds and separation of the atomic planes.

At each damage state, N, the elastic energy per unit volume for a simple tensile load is:

$$U_N = \frac{E_N \cdot \varepsilon^2}{2 \cdot (1 - 2 \cdot \nu_N)}, \tag{17.1}$$

where E_N is Young's modulus of the material at the state N, ε is the strain and ν_N is Poisson's ratio for state N. Figure 17.1 shows a series of elastic potential curves. Although the exact shape of a particular curve depends on temperature, the material microstructure, etc., its general features are assumed to represent a wide range of time-dependent processes of strength deterioration under stress loading.

Transition to the state N involves dynamic relaxation of the material to the state of mechanical equilibrium corresponding to this state. The maximum slope of the relevant

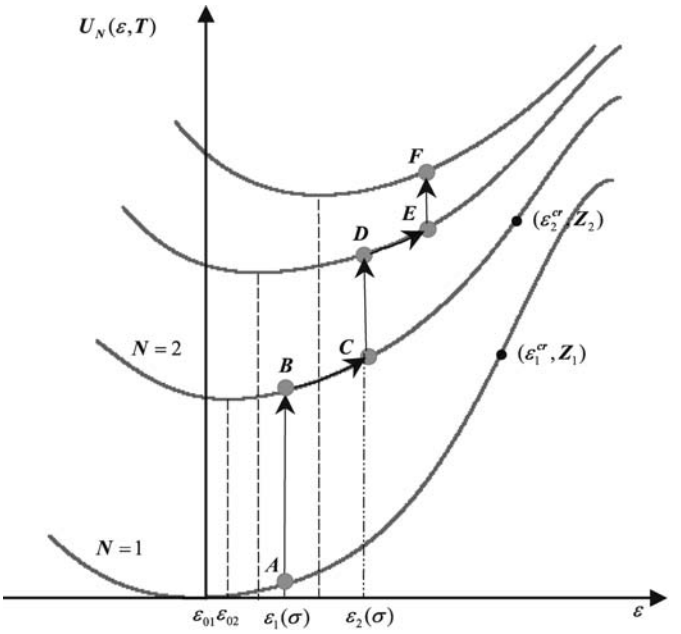


FIGURE 17.1. Elastic energy potential representation of damage accumulation process [1].

curve represents the corresponding maximum material strength Z_N, which decreases as the number N increases. The failure criterion in this model reads:

$$Z_{N_{cr}} = \sigma(t). \tag{17.2}$$

Figure 17.2 illustrates the implementation of this failure criterion through the material's stress–strain diagrams.

The state transitions are induced by fluctuations in the local factors of the element-stress interaction. Since the transition rate function, $\omega(N', N)$, depends on the onset of local fluctuations with sufficient amounts of energy, the transition rate function can be written, according to the theory of statistical fluctuations, as

$$\omega(N', N) = \omega_0(N', N) \cdot \exp[-\Delta S(N', N)]. \tag{17.3}$$

Here $\omega_0(N', N)$ should be viewed as a (temperature and damage state dependent) material constant, and $\Delta S(N', N)$ is the entropy change associated with the transition between the states N' and N with the corresponding microstructural rearrangement.

The damage accumulation rate, $\dot{N}$, is determined as:

$$\begin{aligned}
\dot{N} &= \dot{N}_+ + \dot{N}_- = 2\omega_0 \cdot \sinh\left(\frac{\Delta \cdot \sigma}{\rho \cdot R \cdot T/A}\right) \cdot \exp\left(-\frac{Q}{\rho \cdot R \cdot T/A}\right), \\
\dot{N}_+ &= \omega_0 \cdot \exp\left(-\frac{Q - \Delta \cdot \sigma}{\rho \cdot R \cdot T/A}\right), \\
\dot{N}_- &= \omega_0 \cdot \exp\left(-\frac{Q + \Delta \cdot \sigma}{\rho \cdot R \cdot T/A}\right).
\end{aligned} \tag{17.4}$$

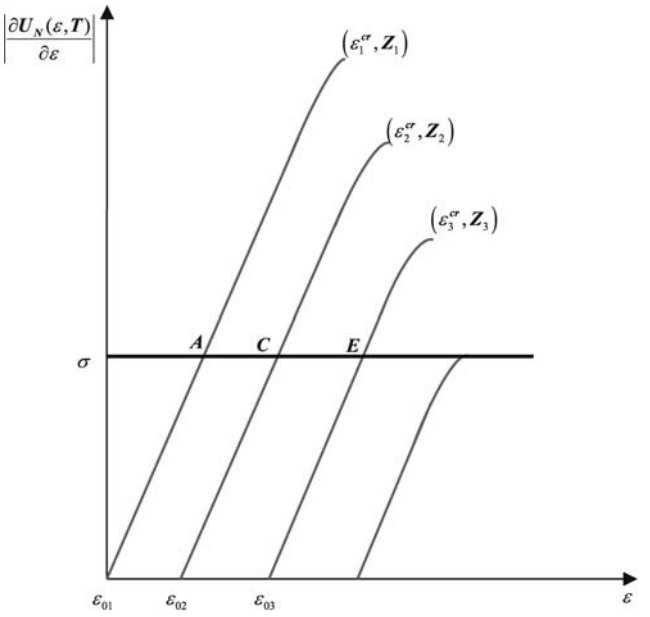


FIGURE 17.2. Stress-strain diagram representation of damage accumulation process [1].

Here $\dot{N}_+$ is the damage accumulation rate, $\dot{N}_-$ is the recovery rate, ρ is the material density; A is the atomic weight; R is the gas constant, Q is the activation energy represented by the ordinate increment, and Δ is the abscissa increment.

17.1.1.1. Reliability Distribution Function The reliability distribution function (RDF) [4] is defined as the probability density function, $r(x,t)$, for finding an element functioning at the moment of time t and having strength in the interval $(x, x+dx)$, x being a generalized strength parameter. The conventional reliability function, $R(t)$, can be expressed through the RDF as follows:

$$R(t) = \int_{(x)} r(x,t)dx. \tag{17.5}$$

The integral of the RDF over the strength space expresses the probability of finding an element functioning at time t, while being on a non-failed strength level. The Smoluchowski's probability of transition from the strength level, x, to a weaker state, x', should satisfy the normalization condition:

$$\int_{x' \le x} P(x \to x')dx' = 1. \tag{17.6}$$

The rate of change of the RDF is determined by the rates of three processes:

(1) The failure rate at the strength level x: $\lambda(x,t) \cdot r(x,t)$, where $\lambda(x,t)$ is the rate of events with a stress level above the element strength, x, and leading, therefore, to the element's failure;
(2) the rate of the strength deterioration: $\omega(x,t) \cdot r(x,t)$, where $\omega(x,t)$ is the rate of events below the element strength at time t;

(3) the integral rate of transition to the strength level x from the higher strength state x':

$$\int_{x'>x} \omega(x',t)\,p(x' \to x)r(x',t)dx'. \tag{17.7}$$

While the first two terms in (17.7) tend to reduce the RDF, the last term tends to increase it. The transfer equation for the RDF can be written therefore in the following form:

$$\frac{\partial r(x,t)}{\partial t} = -[\lambda(x,t) + \omega(x,t)] \cdot r(x,t) + \int_{x'>x} \omega(x',t)\,p(x' \to x)r(x',t)dx' + \delta(t) \cdot r(x,t). \tag{17.8}$$

The function $r(x,t)$ in the last term accounts for the initial values $r(x,0)$, and should satisfy the requirement:

$$R(0) = \int_{(x)} r(x,0)dx = 1. \tag{17.9}$$

The assumption that x_0 is the initial strength distribution of the "virgin" undamaged fiber, so that $r(x,0) = \delta(x - x_0)$, does not entail any lack of generality, since the solution obtained on this basis can serve as a Green function for any other initial distribution $f(x) = r(x,0)$. In such a case, $r(x,t)$ can be found as the following convolution:

$$r(x,t) = \int_0^\infty r_{x_0}(x',0) \cdot f(x - x')dx'. \tag{17.10}$$

The simplified model for the transfer function $p(x' \to x)$ is chosen in the following form:

$$p(x' \to x) = \beta \cdot \frac{x^{\beta-1}}{(x')^\beta}. \tag{17.11}$$

17.1.2. Impact of the Three-Dimensional Mechanical-Temperature-Humidity Load on the Optical Fiber Reliability

As it has been mentioned above, the basic factor of crack propagation in silica glass is the combination of stress, moisture, temperature, and time [16,18]. The Equation (17.4) represents the damage accumulation rate as a function of stress and temperature. Accordingly, the time to failure dependence on stress and temperature should be exponential as well. Assuming that for a relatively high load the recovery part is negligible, time to failure can be represented as follows:

$$t = t_0 \cdot \exp\left(-\frac{Q - \Delta \cdot \sigma}{\rho \cdot R \cdot T/A}\right). \tag{17.12}$$

TABLE 17.1.
Activation energy (eV) for some type of fibers for 45% and 100% humidity.

Sample 1	Stress, GPa					
	4.84		5.05		5.28	
	RH = 45%	RH = 100%	RH = 45%	RH = 100%	RH = 45%	RH = 100%
	3.896	3.88	3.896	3.88	3.891	3.868
Sample 2	Stress, GPa					
	4.55		4.74		4.94	
	RH = 45%	RH = 100%	RH = 45%	RH = 100%	RH = 45%	RH = 100%
	3.827	3.721	3.826	3.723	3.826	3.728
Sample 3	Stress, GPa					
	4.65			4.84		
	RH = 45%	RH = 100%		RH = 45%		RH = 100%
	3.755	3.566		3.755		3.564
Sample 4	Stress, GPa					
	4.65			4.84		
	RH = 45%	RH = 100%		RH = 45%		RH = 100%
	3.899	3.886		3.899		3.886

We proceed from the idea that the main reason of a slow crack growth is humid environment. Elevated relative humidity (RH) decreases the activation energy. With this in mind, Equation (17.12) could be written as:

$$t = t_0 \cdot \exp\left(-\frac{Q(\mathrm{RH}) - \Delta \cdot \sigma}{\rho \cdot R \cdot T/A}\right). \tag{17.13}$$

Here the function $Q(\mathrm{RH})$ is assumed to be linear, with intercept (0% humidity) equal to sublimation energy of the material. Table 17.1 shows the levels of the activation energy, which were evaluated for different fiber types and different loads for 45% and 100% humidity. It is clear that the activation energy does not depend on the applied load, but depends only on the relative humidity (RH) and the fiber type. Humid environments reduce the activation energy and therefore accelerate the crack grows and, consequently, the process of the damage accumulation.

17.1.3. Effect of Bimodality and Its Explanation Based on the Suggested Model

The standard analysis of the strength data is based on the Weibull distribution:

$$F(x) = 1 - \exp\left[-\left(\frac{x}{\theta}\right)^{\eta}\right], \tag{17.14}$$

where θ is the scaling parameter, η is the shape parameter, and x is strength. The function $\ln[-\ln(1 - F(x))]$ is a linear function of x, with η as a slope and $-\eta \cdot \ln(\theta)$ as an intercept of the line [34].

The strength distribution for optical fibers is usually represented on a Weibull scale. According to most references and the authors' own experimental data, the obtained slope

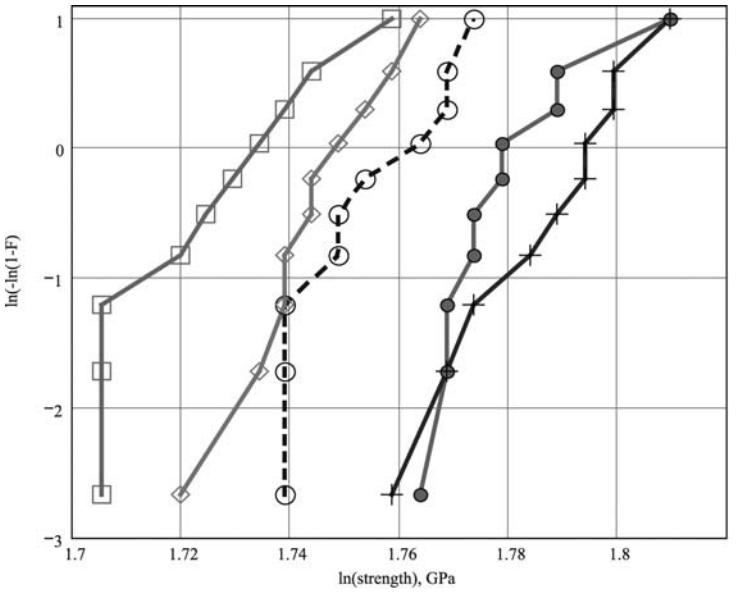


FIGURE 17.3. Examples of optical fiber strength distribution (Weibull Plot).

(the shape parameter in the Weibull distribution) comprises, actually, two different components [5]. Some examples of the strength probability distribution (cumulative distribution function) are shown in Figure 17.3. If the application of the Weibull distribution is justified, i.e., if this distribution adequately reflects the physics of rupture, then the plots should be straight lines. Deviations from such straight lines cannot be explained by statistical uncertainties, but clearly indicate a need for corrections in the model itself. The S-shape line behavior of the test data indicates that the statistical strength distribution could be bi- or even multimodal.

Most references [6,7] assume that bimodality can be explained from the standpoint of the interaction of the "low strength" due to the "external" defects (caused by various mechanical and chemical factors), and "high strength" which is due to various "internal" defects, such as production process defects, thermal fluctuations etc. Consequently, some researchers claim that the Weibull distribution is, in general, unsuitable. They suggest a superposition of two such distributions with two different shape- and scale parameters. Note that, unlike the "regular" Weibull distribution, the bimodal Weibull model is, actually, a *five*-parameter distribution.

In our concept, the two-slope phenomenon in the fiber strength distribution is explained not as an intrinsic feature of the fibers, but rather as a combination of the original strength distribution of the "virgin" (undamaged) material ("strong mode") and the secondary distribution ("weak mode"), developed in the process of damage accumulation under an external and/or internal load. The high-slope portion describes the "original," "young," "non-aged" strength distribution and the low-slope portion of the distribution describes the performance of the damaged population [8].

The tensile test [9] is viewed as a suitable test valid for comparative analysis of the initial and the post-loading distributions of the fiber. In these experiments, 5 m long silica glass fiber specimens of outer diameter 300 μm were divided into five groups. We prepared and tested at least 60 specimens. The first group was subjected to the standard test

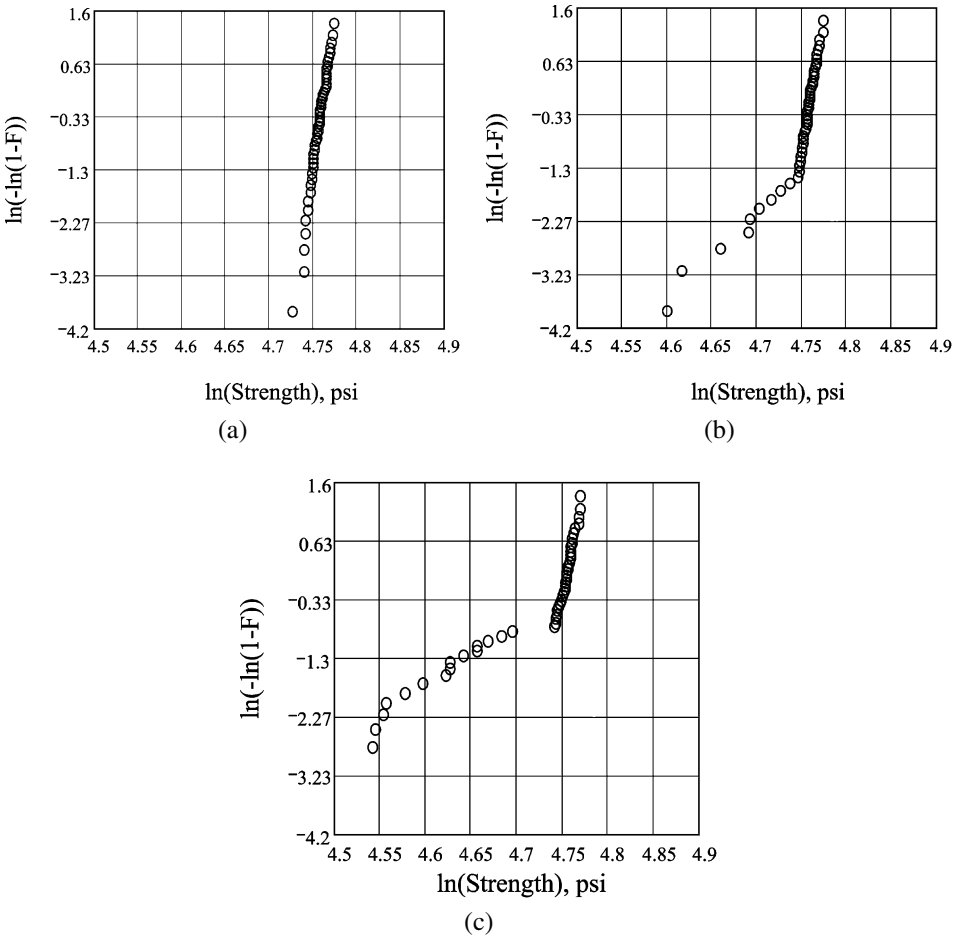


FIGURE 17.4. Strength Weibull plot of (a) undamaged fibers; (b) fibers exposed to low load dose; (c) fibers exposed to high load dose.

(with a constant strain rate), and the other specimens were preliminarily subjected to different "doses" of load, and only after that were subjected to the tensile test. Figure 17.4(a) shows the Weibull chart of the initial fiber. It is clear that the distribution is unimodal. The post-loading distributions begin to look like bimodal [Figures 17.4(b), 17.4(c)] and exhibit several behavioral patterns. Initially, the fiber strength can be described by a homogeneous high-slope distribution with a small standard deviation [Figure 17.5(a)] and a low-slope component (if any) is integrated into the high-slope distribution as a part of it. Under the applied load the high-slope component retains its initial mean and the standard deviation, but the low-slope component becomes more strongly pronounced [Figure 17.5(b)]. Here is the explanation of this phenomenon.

The population of fibers with the low strength increases because of the damage accumulation process. The low-slope part represents the strength distribution of the damaged parts and their source in the initial fibers (high-slope distribution). Area under the low-slope distribution increases at the account of the area of the high-slope part as a result of additional loading [Figures 17.5(b), 17.5(c) and 17.5(d)]. By the end of the process the part

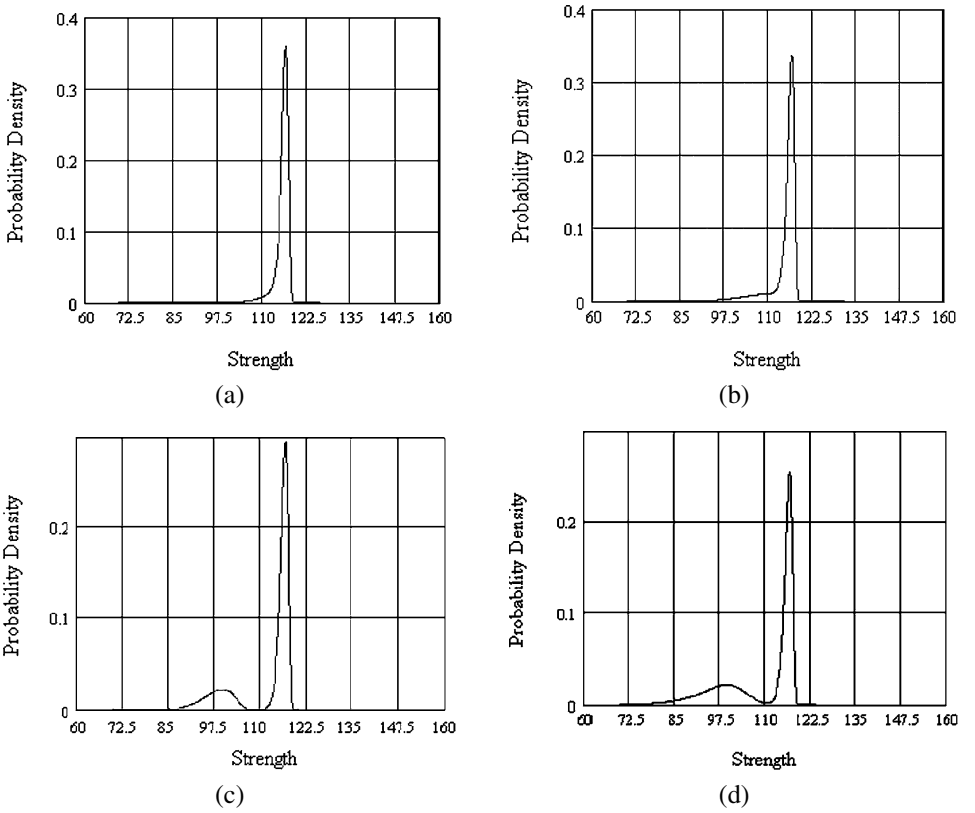


FIGURE 17.5. Strength distribution of (a) undamaged fiber; (b) fibers exposed to low load dose; (c) fibers exposed to medium load dose; (d) fibers exposed to high load dose.

of fibers with the initial strength will vanish (due to the transition to the subpopulation with lower strength and, less probable, due to failure), and only the low-slope distribution with a small mean and a large standard deviation will remain.

In characterizing the damage accumulation process in brittle solids, many researchers agree that such a process starts with the random formation of numerous micron-sized cracks (microcracks) throughout the solid [10–12,17,20]. The size of these microcracks is comparable to that of the defects that are initially present in the "virgin" unloaded material. When the characteristic distance between the adjacent microcracks (correlation length) reduces to a certain threshold value, the microcracks begin to coalesce, forming microcrack clusters. One of these clusters eventually develops into a macroscopic crack, which continues growing until the solid ruptures [31]. As to the time scale, the first stage of the damage accumulation process is, on the average, much longer than the second stage, while the last (third) stage is usually very short compared to the previous stages. The strength deterioration at the first stage is mainly determined by a typical size of the microcrack and is approximately constant (within the statistical "noise").

In the evolution Equation (17.8) it is assumed that the load on the fiber is relatively small, i.e., $x \gg \sigma$, and, accordingly, the failure rate is negligible. The governing Equation (17.8) reduces to

$$\frac{dr(x,t)}{dt} = -\omega(x,t) \cdot r(x,t) + \int_{x'>x} \omega(x',t) \cdot p(x' \to x) \cdot r(x',t) dx'. \tag{17.15}$$

The rate, $\omega(x,t)$, of the strength deterioration at the first stage of the damage accumulation process ($x = x_0$) is merely the inverse of the mean stage duration and, therefore, may be by several orders of magnitude lower than that at the second stage ($x < x_0$). The second stage is generally a time- and strength-dependent process controlled by the frequency of the microcracks formation in the vicinity of a cluster. In many practical situations, however, it may be approximated by a constant [10,12,13]. As the third stage is negligibly short in time, the transition rate may be approximated for our purposes as follows:

$$\omega(x,t) = \begin{cases} \omega, & x < x_0 \\ \Omega, & x = x_0 \end{cases}, \quad \omega \gg \Omega. \tag{17.16}$$

The strength transition probability, $P(x \to x')$, during the second stage of damage accumulation can be estimated based on the following reasoning. Consider a solid that comprises both discrete microcracks and microcrack clusters. The strength of such a solid is defined by the size of the largest cluster. Let us enclose the largest cluster into a 3D lattice such that its links are small relatively to a typical microcrack [see Figure 17.6]. The strength of the solid is related to the size of the largest cluster as [12,14]

$$x = B \cdot N^{-v}, \tag{17.17}$$

where B and v are positive material constants, and N is the number of the lattice sites in the cluster. We attribute the growth of a cluster to thermal-fluctuation formation of new microcracks in the vicinity of the cluster. The probability of nucleation of a new microcrack

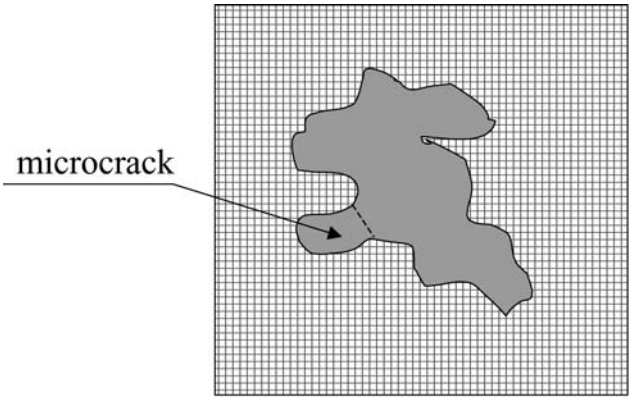


FIGURE 17.6. The largest microcracks cluster circumscribed by a percolation lattice. The lattice shown is a 2D lattice, but generally it can be 3D. The lattice links are small relatively to a typical microcrack size.

in the vicinity of the cluster, i.e., the probability of N increasing by increment ΔN, is then given by

$$P(N \to N + \Delta N) \sim \exp\left(-\frac{\Delta S}{k_B \cdot T}\right), \tag{17.18}$$

where k_B is the Boltzmann's constant, T is the absolute temperature, and S is the configuration entropy. It can be shown that for $N \gg 1$ and $\Delta N \ll N$, the entropy increment is approximately $\Delta S \approx \Delta N/N$, leading to

$$P(N \to N + \Delta N) \approx A \cdot \exp\left(-C \cdot \frac{\Delta N}{N \cdot k_B \cdot T}\right). \tag{17.19}$$

From (17.17) and (17.19) we have

$$P(x \to x') = A \cdot \exp\left(-\frac{v \cdot B \cdot C}{k_B \cdot T} \cdot \frac{x - x'}{x}\right). \tag{17.20}$$

This expression, for $x - x' \ll x$, is identical to the empirical relation (17.11) in Section 17.1.1. Indeed,

$$\exp\left(-\beta \cdot \frac{x - x'}{x}\right) = \left[\exp\left(\frac{x' - x}{x}\right)\right]^{\beta} \approx \left(1 + \frac{x' - x}{x}\right)^{\beta} = \left(\frac{x'}{x}\right)^{\beta}. \tag{17.21}$$

Therefore, the empirical parameter β is given by

$$\beta = \frac{v \cdot B \cdot C}{k_B \cdot T}. \tag{17.22}$$

The parameter B is proportional to the material's fracture toughness [12,13,19]: the tougher the material, the higher the β value and the narrower is the transition probability distribution. The form (17.11) of the transition probability, although approximate, is nevertheless preferable it since (as is shown in the next section), it enables one to proceed with an analytical solution to the evolution Equation (17.15). The corresponding single-parameter transition probability density function can be found as

$$p(x \to x') = (\beta + 1) \cdot \frac{(x')^{\beta}}{x^{\beta+1}}. \tag{17.23}$$

17.1.3.1. Time Behavior of the Reliability Distribution Function [8] The Laplace transform

$$\tilde{r}_{x_0}(x, s) = \int_0^{\infty} \exp(-s \cdot t) r_{x_0}(x, t) dt \tag{17.24}$$

of the solution of Equation (17.8) with the obtained relationships (17.16) and (17.23) has the following form:

$$\tilde{r}_{x_0}(x, s) = (\beta + 1) \cdot \frac{\Omega}{(\Omega + s) \cdot (\omega + s)} \cdot \frac{1}{x} \cdot \left(\frac{x}{x_0}\right)^{\frac{(\beta+1)\cdot s}{\omega+s}} + \frac{\delta(x - x_0)}{\Omega + s}. \tag{17.25}$$

In order to obtain the inverse transform in the analytic form, let us consider the first term of the Equation (17.25) and represent it as follows:

$$\frac{(\beta+1)\cdot\Omega}{(\Omega+s)\cdot(\omega+s)}\cdot\frac{1}{x}\cdot\left(\frac{x}{x_0}\right)^{\frac{(\beta+1)\cdot s}{\omega+s}} = \frac{(\beta+1)\cdot\Omega}{x\cdot(\Omega+s)}\cdot\frac{1}{(\omega+s)}\cdot\exp\left[\log\left(\frac{x}{x_0}\right)\cdot(\beta+1)\cdot\left(1-\frac{\omega}{\omega+s}\right)\right]. \tag{17.26}$$

Expanding the exponent into a series we obtain:

$$\tilde{r}_{x_0}(x,s) = \frac{\Omega\cdot(\beta+1)}{x\cdot(\Omega+s)}\cdot\left(\frac{x}{x_0}\right)^{\beta+1} \times\sum_{i=0}^{\infty}\frac{(-1)^i\cdot\omega^i\cdot(\beta+1)^i\cdot\left[\log\left(\frac{x}{x_0}\right)\right]^i}{i!\cdot(\omega+s)^{i+1}}+\frac{\delta(x-x_0)}{\Omega+s}. \tag{17.27}$$

Thus, the inverse transform is the series:

$$r_{x_0}(x,t) = \delta(x-x_0)\cdot\exp(-\Omega\cdot t)+\frac{\Omega\cdot(\beta+1)\cdot\exp(-\Omega\cdot t)}{x}\left(\frac{x}{x_0}\right)^{\beta+1} \times\sum_{i=0}^{\infty}\left\{\frac{(\omega)^i\cdot(\beta+1)^i\cdot\left[\log\left(\frac{x_0}{x}\right)\right]^i}{(i!)^2}\int_0^t(t')^i\cdot\exp(\Omega-\omega)\cdot t'dt'\right\} \tag{17.28}$$

where

$$\int_0^t(t')^i\cdot\exp(\Omega-\omega)\cdot t'dt' = \frac{i!}{(\omega-\Omega)^{i+1}}-\frac{i!\cdot\exp(\Omega-\omega)\cdot t}{(\omega-\Omega)^{i+1}}\sum_{j=o}^{i}\frac{(\omega-\Omega)^j\cdot t^j}{j!}. \tag{17.29}$$

The first term in the obtained distribution (17.28) is the contribution of the "virgin" (original, "undamaged") material and the second term represents the secondary distribution caused by the damage accumulation process. The proportion of the original distribution in the composite structure decreases exponentially with time. Figure 17.7 describes the time evolution of the secondary distribution for a loaded silica fiber. As the central trend of the strength distribution decreases with time, the distribution spreads wider. After a time interval of about $5/\Omega$, the original distribution has practically vanished (no undamaged specimens remain), and the secondary distribution approaches the Weibull curve. The shape parameter of this distribution is governed by the parameter $\beta(15)$, i.e., by the fracture toughness of the material and temperature.

17.1.3.2. Dynamics of the Optical Fiber Strength In order to validate the obtained model let us return to the experiment described at the beginning of this section.

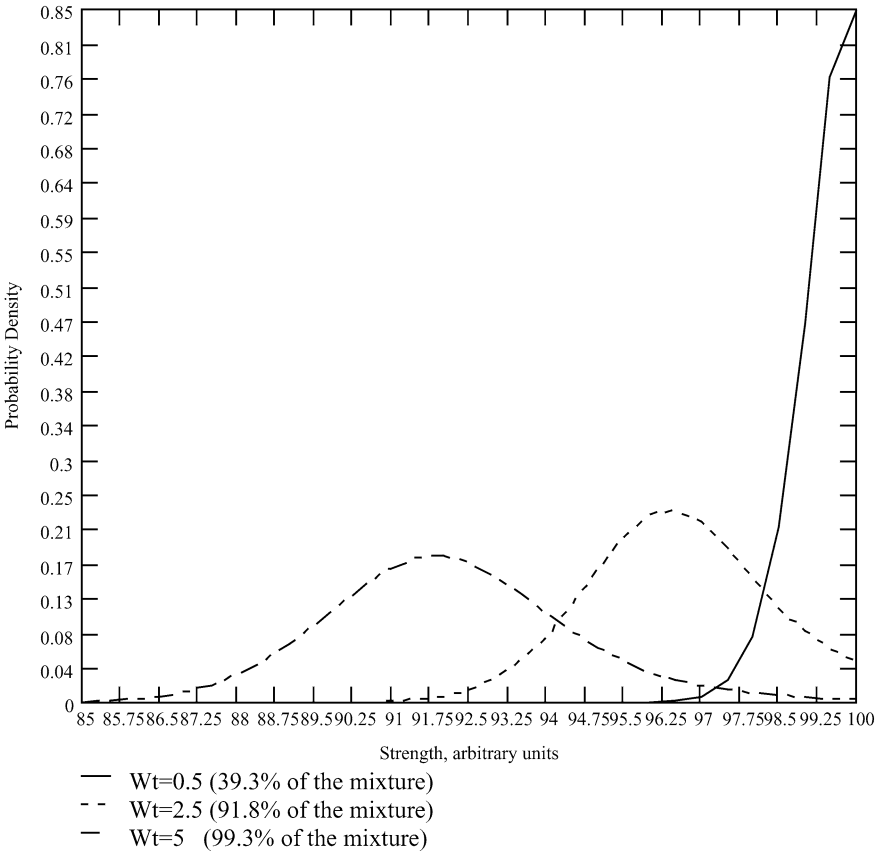


FIGURE 17.7. The Daughter Distribution Dynamics as a solution of the evolution equation.

To estimate the parameters of the predicted distribution (17.28) on the basis of the experimental data, a minimal distance scheme [15] was used. In this scheme, the distance between the empirical Cumulative Distribution Function (CDF) and the analytical one is minimized within the parameter space. The only three parameters in the predicted distribution (17.28) are Ωt, ω and β. The latter two are not supposed to change during the process of damage accumulation. The constancy of the parameters ω and β was used as a convenient condition in the estimation procedure. The delta-function was approximated by a normal distribution, which had an excellent goodness-of-fit for the first-group sample.

Figures 17.8 and 17.9 represent the estimation results for low and high doses of the preliminary stressing, respectively. The agreement between the estimated distributions and experimental results is satisfactory, especially having in mind that only one estimated parameter (Ωt) accounts for the damage dose. For the low preliminary damage the resultant distribution is close to the initial delta-function with a rather small contribution of the "weak mode." For the high preliminary damage the resultant strength distribution exhibits bimodality. Since the initial flaw distribution is apparently the same for all the tested fiber specimens, this bimodality can be explained by the damage accumulated in the preliminarily stressed specimens.

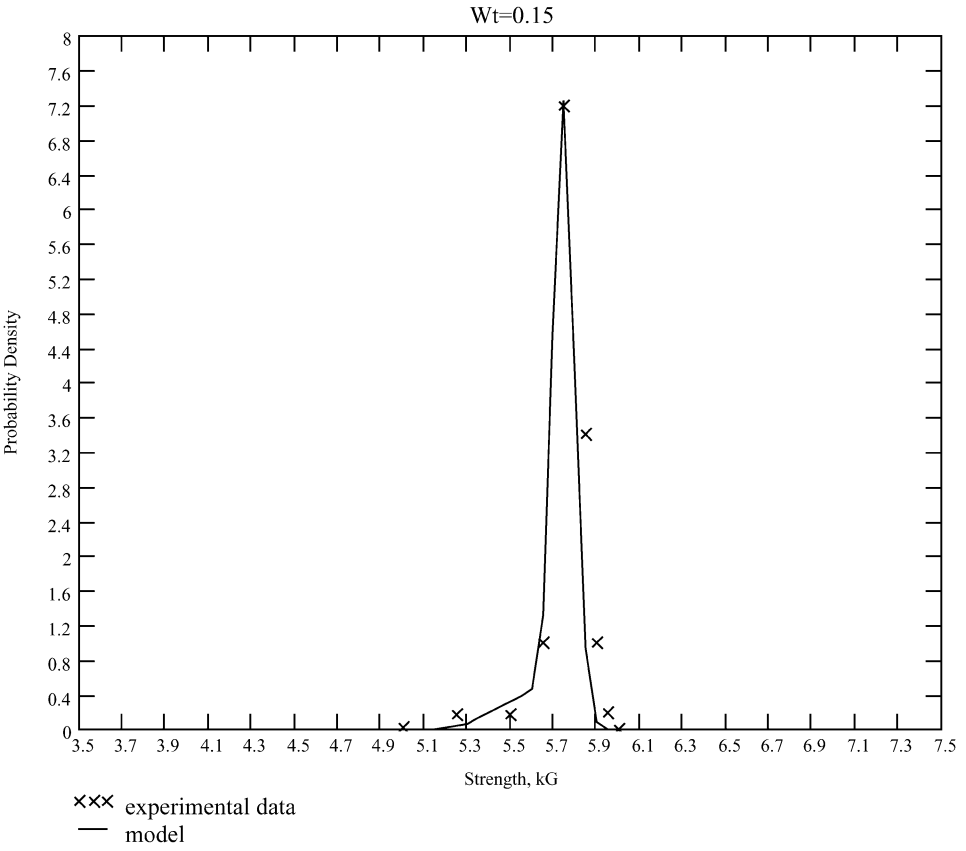


FIGURE 17.8. Experimental strength distribution of optical fibers exposed to a relatively low load dose and the model prediction.

Unlike the traditional approach to statistical modeling of fracture, the present method describes the phenomenon in its dynamics. The proposed approach links the thermal-fluctuation damage events with the corresponding strength deterioration, and thereby establishing an evolution equation of the time-dependent strength distribution whose solution describes the above pattern, which is confirmed experimentally.

The actual strength of the "virgin" material is by several orders of magnitude lower than the theoretical strength calculated on the basis of the material intermolecular forces [10]. This discrepancy is usually attributed to the defects present even in the "virgin" material. Since these defects are not inherent to the material, but formed during the production process and the subsequent use conditions, the proposed approach is quite general and is not limited to the pre-stressed specimens.

The resultant combined distribution has only three parameters, two of which (Ω and ω) characterize the strength deterioration dynamics, and the remaining parameter (β) characterizes the "daughter distribution" shape. The two former parameters are controlled by the rate of the microcrack nucleation, which is governed by the material's fracture toughness and the temperature.

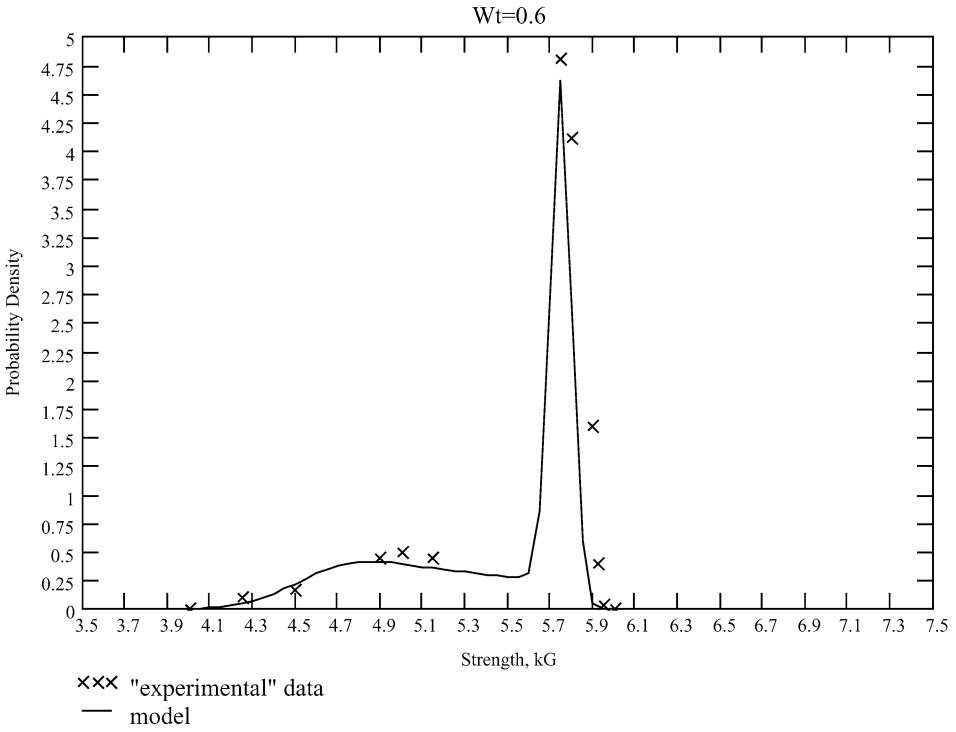


FIGURE 17.9. Experimental strength distribution of optical fibers exposed to a relatively high load dose and the model prediction.

17.2. RELIABILITY IMPROVEMENT THROUGH NPM-BASED FIBER STRUCTURES

17.2.1. Environmental Protection by NPM-Based Coating and Overall Self-Curing Effect of NPM Layers

In a fiber under load, a surface defect acts as a stress concentrator with the maximum stress at its tip, which is the preferential site of moisture attack and rupture of the silica bonds. This causes the defect to spread, and, consequently, the stress to increase in magnitude until the fiber catastrophically fails [28]. Fibers require long-term protection against moisture and oxygen, as well as mechanical and thermal protection. Most existing coatings for optical fiber are polymer-based. These are moisture-sensitive ("non-hermetic"). Although the reliability of polymer coatings has improved considerably during the last decade, it is not as high as necessary for particular applications.

The surface of most polymers is covered with protrusions, typically 0.1–0.3 micron high, their density being of the order of $0.25–1 \times 10^8$ peaks/cm^2 (meaning average spacing of the peaks about 1–2 microns). Under temperature cycles and deformation of the material between the peaks, significant stress concentrations are created in the "gullies" resulting in inter-granular microcracks or pinholes. These defects destroy the continuity of the plastic coating and seem to be the main reason for its permeability to gases.

Nanoparticle-type coating materials (NPM's) could be used to fill the gullies, microcracks and other defects in the plastic. The most important physical property of these materials is their thixotropy, which can be defined as fluidity under stress. This property is

exhibited by some gels that become fluid when stirred or shaken, and revert to a semi-solid state, when stirring or shaking ends. Therefore the same stress that creates or aggravates the defect, makes the NPM less viscous and thus enables it to penetrate (fill in) and "heal" the defect.

Thanks to this effect, plastics no longer have to be designed for a long lifetime (which is usually the main problem). The plastic material may deteriorate: the NPM coating will take care of improving the situation. Another important aspect of the NPM coatings is their moisture-proofing action. The hydrophobic nanoparticles tend to fill in the spaces between the peaks, but do not necessarily follow the profile, so that this part of the film is somewhat flattened. The particles interconnect among themselves and with the coated plastic through the links of the same or a congeneric polymer, whose presence was predetermined in the original coating material. The contents of hydrophobic nanoparticles and of the polymer in the coating material have to correspond to the amounts needed to fill the spaces between the peaks. The characteristic dimension of the spaces between the peaks is about 3–6 Å in a well-packed structure. This narrows tremendously the passages for water molecules. The hydrophilic nanoparticles (one of their functions is to provide the affinity with the coated molecules) form the outer layer of the coating.

The space between the peaks, which are about 0.15 micron in height, is covered by the hydrophobic nanoparticles on both sides of the sheet. Their total volume of these nanoparticles on both sides of the sheet is

$$V_{phob} = \frac{2}{3} \cdot S \cdot h, \tag{17.30}$$

where S is the area of the sheet. The coefficient in front of h is based on the simplifying assumption that the peaks are conical.

Let H be the thickness of the hydrophilic NPM layer. Then,

$$V_{phil} = S \cdot H, \tag{17.31}$$

and, for the same packing density, the concentration ratio between the hydrophilic and hydrophobic nanoparticles is:

$$\frac{V_{phil}}{V_{phob}} = \frac{3 \cdot H}{2 \cdot h}. \tag{17.32}$$

For $S = 1$ m^2, $H = 3$ μm and $h = 1.5 \times 10^{-7}$ m, the volumes of the hydrophobic and hydrophilic parts will be, respectively, $V_{phob} = 10^{-7}$ m^3 (0.07 grams) and $V_{phil} = 3 \times 10^{-6}$ m^3 (2 grams). This correspond to the nanoparticle surface of each kind of $S_{phob} = 20$ m^2 and $S_{phil} \sim 600$ m^2. This amount of hydrophilic particles is able to absorb up to 3 grams of water. This means that even in the case of a plastic with water permeability of 10^{-2} g/m^2/day it will take 300 days for the coating to be completely penetrated, not to mention the "healing" effect of the NPM, nor the another NPM layer on the other side. Due to the extremely low modulus of elasticity of these layers, the flexibility of the systems remains practically unchanged. In some application, it might be a very useful property.

On the other hand, the filling effect of the nanoparticles creates a 3D network of narrow passages with the characteristic width of 3–10 Å, i.e., the one of molecular size (the typical size (diameter) of the water molecule is about 3 Å). If we idealize the packing structure as large balls with the spaces between them filled by smaller balls (Figure 17.10),

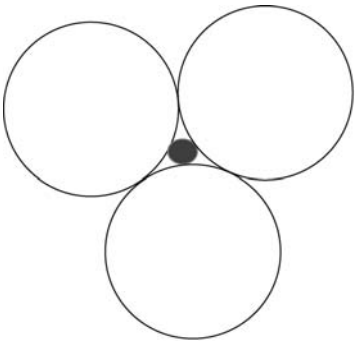


FIGURE 17.10. Packing structure of nanoparticles.

then the smallest remaining spaces between the smallest balls will be about 0.155 of the smallest ball radius. Assuming the smallest nanoparticles to be in the range of 4–10 nm in diameter, the above mentioned passages could be readily evaluated.

This dramatic change in the passage pattern has a drastic effect on the diffusion of oxygen and water molecules through the system. The diffusion coefficient before filling corresponds to that in gases, while after the filling it is expected to be closer to that of solids. The molecule free path lengths in gases are on the order of $\sim$100 nm, while in solids these lengths are on the order of 3–10 nm. This alone is expected to reduce the diffusion coefficient by about 3 orders of magnitude. Another physical mechanism, resulting in a reduced diffusion coefficient is the number of "jumps" per second. In gases, this number is the ratio of the molecule velocity to its free length, while in our, NPM "filled," case, the number of "jumps" per second depends on the activation energy. Our estimation is that this mechanism can add another order of magnitude to the above effect, so that the total reduction in permeability thus being improved by about 4 orders of magnitude.

Thus, the combination of the hydrophilic-hydrophobic layers is expected to provide the necessary protection against the water vapor and oxygen penetrator for an acceptable lifetime.

17.2.2. Improvement in the Reliability Characteristics by Employing NPM Structures in Optical Fibers

In order to compare the mechanical and the environmental characteristics of the NPM-based and "regular" fibers under different loading and environmental conditions, two experiments described below were carried out.

The specimens were obtained from light guides consisting of silica cores and polymer coatings, either regular ("reference" fibers) or modified (by adding the NPM to the fiber coating).

- **Time-to-failure (TTF): delayed fracture**

  In this experiment, we applied mechanical stresses to the fiber specimens ("two-point bending" condition [36]) and measured the time-to-failure (TTF) for each case under two relative humidity (RH) conditions—RH = 35% ("normal") and RH = 100%. All the specimens were subjected to the predetermined humidity for at least 48 hours.

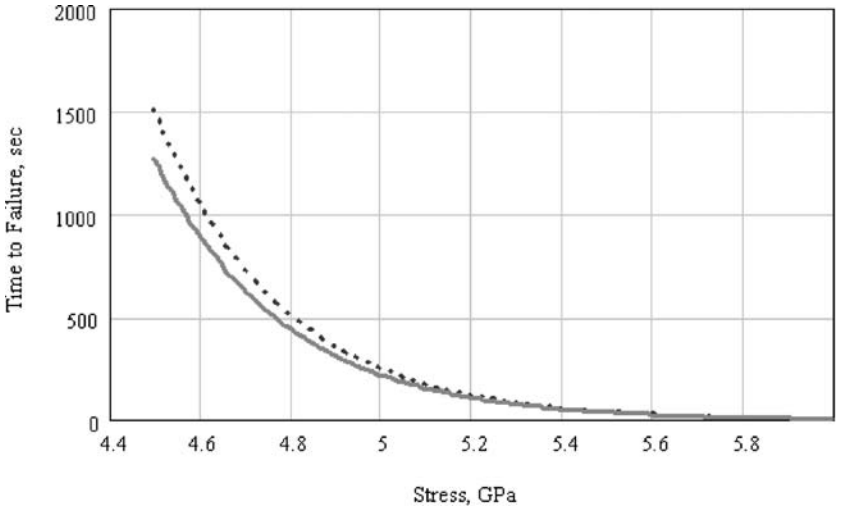


FIGURE 17.11. Time-to-failure vs. bending stress (35%RH). NPM-based sample (dotted curve) and the reference samples strength (solid curve).

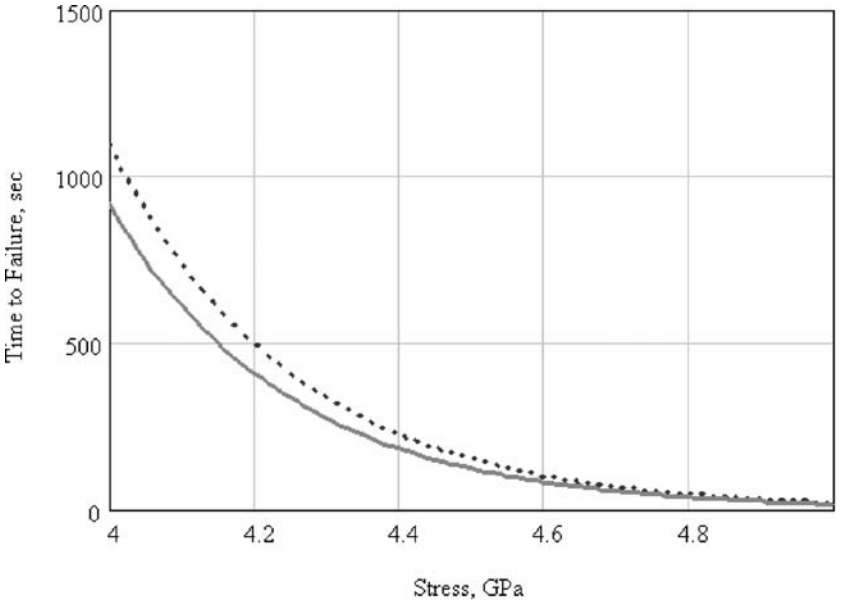


FIGURE 17.12. Time-to-failure vs. bending stress (100%RH). NPM-based sample (dotted curve) and the reference samples strength (solid curve).

- **Ultimate strength: "immediate" rupture**

  This experiment was carried out under the same RH and loading conditions as the previous ones, but the applied stress being such as to cause "immediate" rupture. The faceplates were brought together at the rate of 50 microns/sec up to the very moment of rupture.

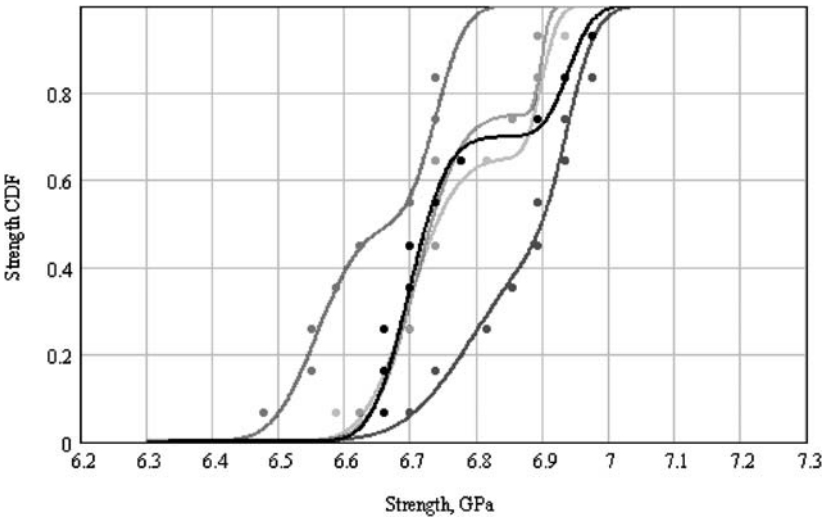


FIGURE 17.13. Strength probability for all tested at 35%RH samples.

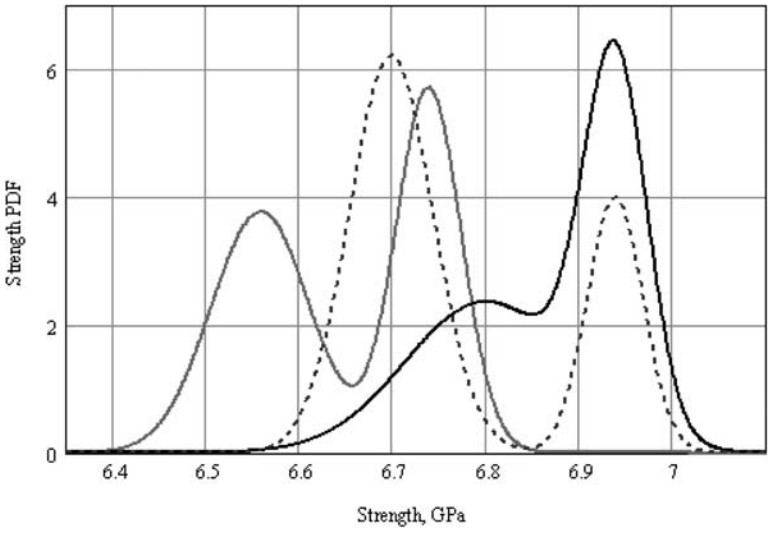
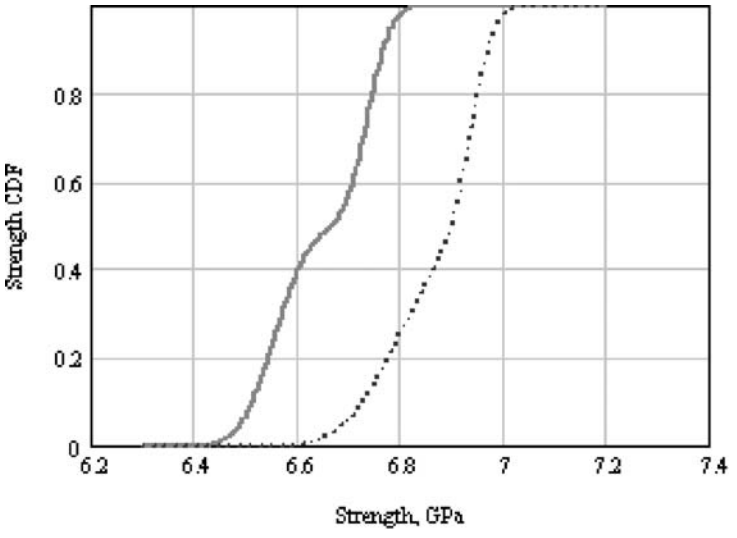


FIGURE 17.14. Strength PDF for samples tested at 35%RH.

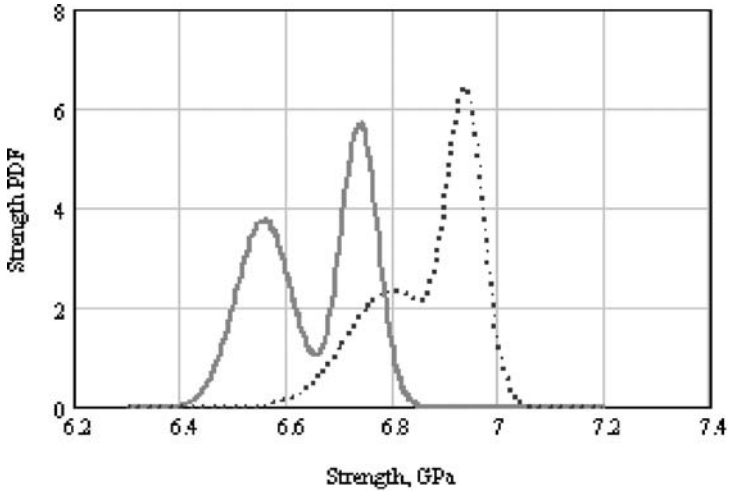
In order to compare the TTF for different specimens, we approximated the relationships between the TTF and the applied stress, x, by a power law:

$$\mathrm{TTF}(x) = \frac{\tau}{x^n},$$

where τ is an empirical constant, representing the time to failure (rupture) under a unity stress level, and n is an empirical (material's)exponent. This relationship is shown for the two types of specimens in Figures 17.11 (35%RH) and 17.12 (100%RH). The obtained data clearly demonstrate the favorable effect of the NPM-based coating.



(a)



(b)

FIGURE 17.15. NPM-based sample (dotted curve) and the reference samples strength (solid curve) distributions—35%RH. (a) CDF, (b) PDF.

In order to analyze the results of the ultimate strength tests, the strength probability distributions (cumulative distribution functions, CDF) and their Weibull approximations [Equation (17.14)] were constructed for all the specimens (Figure 17.13). Figure 17.14 represents the derivatives of these functions, i.e., the probability density functions (PDF) for the fiber strength. Note that all the plots exhibit the double-slope pattern mentioned above. It is easy to see two explicit extremes on the strength distributions—the most deteriorated strength population corresponds to the reference case (no NPM-based fibers), and the least deteriorated strength has been demonstrated by the sam-

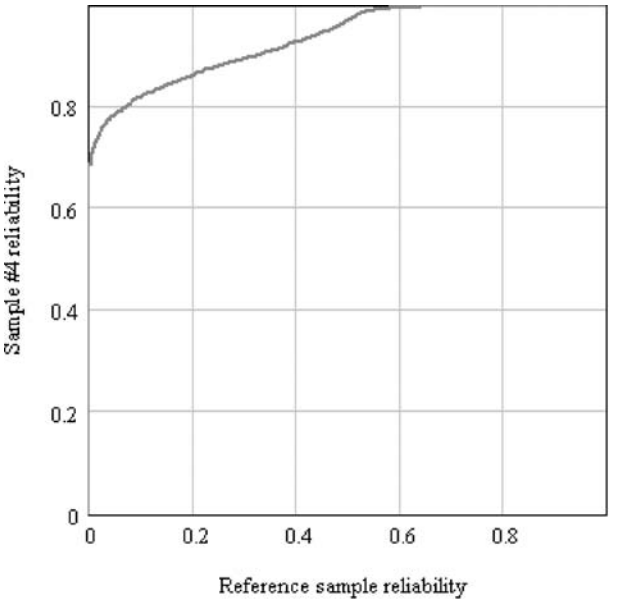
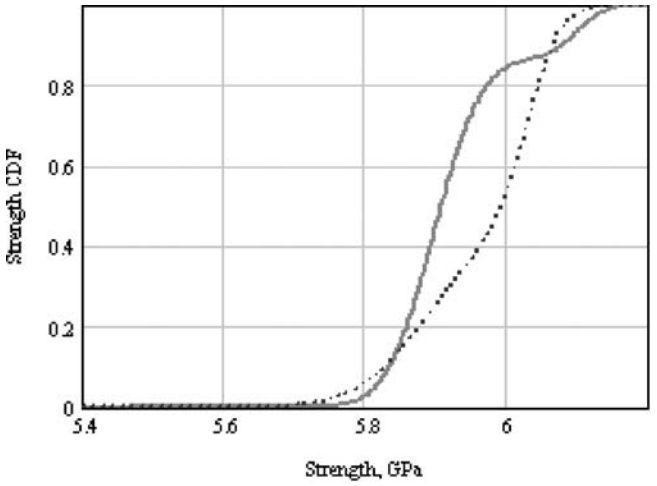


FIGURE 17.16. Reliability comparison—35%RH. NPM-based sample reliability vs. reference sample reliability.

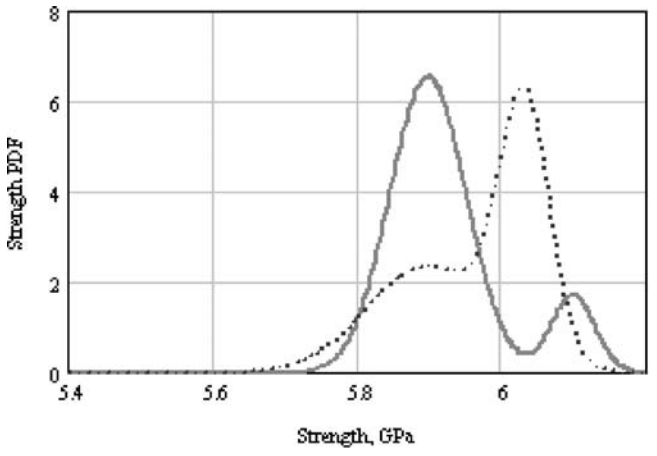
ple with NPM containing coating. These two extreme cases are shown also in Figure 17.15.

The plot in Figure 17.16 demonstrates the reliability aspect of the differences in the strength distributions for the two extreme cases under the action of the same stress. The reliability range of the reference case is zero up to about 0.55, while the NPM-based case begins at about 0.7 and ends very close to 1, a fact which speaks for itself. The two cases demonstrated considerably the different mechanical and environmental performance, although the specimens were taken out of the same preform with the same core material, manufactured at the same drawing facility under identical conditions, and tested at the same time and at the same "age," by the same testing equipment and the same personnel. The improvements can be explained, first of all, by the "curing" abilities of the thixotropic NPM compound and the NPM-based coating. In addition, the coating, when subjected to bending deformations, has the ability to "slide" along the silica surface. This have to a situation, when the silica fiber and its coating behave (mechanically) "independently". This advantage was, in effect, anticipated in the basic considerations underlying the introduction of the NPM-based coating.

Figure 17.17 shows the strength and the reliability characteristics of the two cases at 100%RH. Unlike the 35% tests, where the undamaged and damaged subpopulations are divided almost evenly, the situation here is different. The NPM case exhibits the same proportion between the subpopulations—the point, at which the curve changes its shape, falls in the vicinity of 6 GPa, corresponding to the 0.5 probability. By contrast, for the reference case, this point falls at approximately 6.05 GPa, which corresponds to the probability of almost 90%, indicating that almost 90% of the reference population is damaged. The reference distribution is shifted considerably to the left (in the direction of lower strengths) in comparison with the NPM case. The maximum value of the right distribution for the ref-



(a)



(b)

FIGURE 17.17. NPM-based sample (dotted curve) and the reference samples strength (solid curve) distributions—100%RH. (a) CDF, (b) PDF.

erence PDF covers a much smaller area than the high CDF values describing the damaged population.

The NPM sub-populations are separated less than those of the reference case. This indicates that in the NPM-based case we have both less damaged areas and a lesser damage (in the damaged specimens) than in the reference case. In the 35%RH test we observe, for the NPM case, much less damage and almost non-separated subpopulations. This subdivision most likely characterizes the micro-cracks and micro-damage of the core due to the manufacturing, winding and other fiber making processes. The healing action of the NPM slows down further deterioration of the fiber due to these defects, when in contact with water OH groups, and keeps the damaged population closer to the strength of the undamaged one. The reference case, therefore, represents growth of the damaged subpopulation for the

elevated humidity conditions, as well as the more strongly pronounced impact of damage on the specimen strength.

17.3. CONCLUSIONS

The following major conclusions can be drawn from the performed analysis:

- Application of the NPM results in substantial slowing-down of fiber aging and in the rate of damage accumulation;
- NPM is able to "heal" small core-surface defects;
- NPM effectively protects the silica surface from damage by water vapor;
- Application of the NPM improves substantially the fiber strength and reduces its variability;
- Application of the NPM coating is a promising way to improve optical, mechanical, environmental, reliability and even economic characteristics of silica light guides.

REFERENCES

1. E. Suhir, M. Fukuda, and C.R. Kurkjian, Eds., Reliability of photonic materials and structures, Materials Research Society Symposia Proceedings, Vol. 531, 1998.
2. D. Ingman and L.A. Reznik, Dynamic reliability model for damage Accumulation Processes, J. Nucl. Technol, 75, pp. 261–282 (1986).
3. D. Ingman and L.A. Reznik, Dynamic character of failure state in damage accumulation processes, Nuclear Science and Engineering, 107, pp. 284–290 (1991).
4. D. Ingman and L.A. Reznik, A dynamic model for element reliability, Nuclear Engineering and Design, 70, pp. 209–213 (1982).
5. K.K. Phani, Strength of long optical glass fibers, Journal of Applied Physics, 62, pp. 719–720 (1987).
6. D.S. Durham and W.J. Padgett, Cumulative damage models for system failure with application to carbon fibers and composites, Technometrics, 39, pp. 34–44 (1997).
7. H.M. Taylor, The Poisson-Weibull flaw model for brittle fiber strength, in J. Galambos, et al., Eds., Extreme Value Theory, Amsterdam, Kluwer, 1994, pp. 43–59.
8. M.I. Zeifman and D. Ingman, A dynamic view of strength bimodality of optical fibers, 2003.
9. A. Dumai, Reliability prediction on tensile loaded fiber optics, Research Thesis, Technion, 1991.
10. V.A. Petrov, A.Ya. Bashkarev, and V.I. Vettegren, Physical Basis of Durability Forecasting for Engineering Materials, Politekhnika, Sankt-Petersburg, 1993 (in Russian).
11. V.S. Kuksenko and V.P. Tamuz, Fracture Micromechanics of Polymer Materials, Martinus Nijhoff, Dordrecht, 1981.
12. M.I. Zeifman and D. Ingman, A percolation model for lifetime variability in polymeric materials under creep conditions, Journal of Applied Physics, 88, pp. 76–87 (2002).
13. M.I. Zeifman, A coarse semi-analytical lattice model of time-dependent failure of hierarchical materials, Europhysics Letters, 63, pp. 333–339 (2003).
14. G.P. Cherepanov, et al., Fractal fracture mechanics—a review, Engineering Fracture Mechanics, 51, pp. 997–1033 (1995).
15. B.S. Everitt and D.J. Hand, Finite Mixture Distributions, Chapmen and Hall, Cambridge, GB (1981).
16. D.B. Barker and Y. Yang, Effect of proof testing on optical fiber fusion splices, http://www.nepp.nasa.gov
17. P. Beumont and H. Sekine, Physical modelling of engineering problems of composites and structures, Applied Composite Materials, 7, pp. 13–37 (2000).
18. R.J. Castilone and T.A. Hanson, Strength and dynamic fatigue characteristic of aged fiber, Corning Inc., Corning, NY.
19. J.A. Collins, Failure of materials and Mechanical Design, 2nd edition, John Wiley&Sons, New York, 1981.
20. F. Desrumaux, F. Meraghni and M.L. Benzeggagh, Micromechanical modeling coupled to a reliability approach for damage evolution prediction in composite materials, Applied Composite Materials, 7, pp. 231–250 (2000).

21. X. Diao, A statistical equation of damage evolution, Engineering Fracture Mechanics, 52, pp. 33–42 (1995).
22. X. Diao and X. Xing, Nonequilibrium statistical theory of damage fracture for quasi-brittle materials, Engineering Fracture Mechanics, 56, pp. 321–330 (1997).
23. A. Ekberg, Fracture mechanics—some notes, http://www.solid.chalmers.se/~anek/research/fm.pdf.
24. M. Fukuda, Historical overview and future of optoelectronics reliability for optical fiber communication systems, Microelectronics Reliability, 40, pp. 27–35 (2000).
25. R. Ganesan, A stochastic modeling and analysis methodology for quantification of fatigue damage, Comput. Methods Appl. Mech. Engrg., 190, pp. 1005–1019 (2000).
26. G.S. Glasemann, The mechanical behavior of large flaws in optical fiber and their role in reliability prediction, Corning, New York. http://www.corning.com/opticalfiber/pdf/tr3268.pdf
27. W. Griffioen, et al., Evaluation of optical fiber lifetime models, Proc. SPIE, Vol. 1791, Optical Materials Reliability and Testing, 1992.
28. S.L. Semenov, Physical process determining strength and durability of optic fibers, Moscow, 1997.
29. D. Sauvage, D. Laffitte, J. Perinet, Ph. Berthier, and J.L. Gourdard, Reliability of optoelectronic components for telecommunications, Microelectronics Reliability, 40, pp. 1701–1708 (2000).
30. G.S. Wang, A probabilistic damage accumulation solution based on crack closure model, International Journal of Fatigue, 21, pp. 531–547 (1999).
31. D.S. Wilkinson, E. Maire, and R. Fougeres, A model for damage in a clustered particulate composite, Materials Science and Engineering, A262, pp. 264–270 (1999).
32. M.P. Wnuk, Constitutive modeling of damage accumulation and fracture in multiphase materials, Comput. Methods Appl. Mech. Engrg., 151, pp. 587–591 (1998).
33. J. Jireh, Yue Energy Concepts for Fracture, http://www.sv.vt.edu/classes/MSE2094_NoteBook/97ClassProj/anal/yue/energy.html
34. E. Suhir, Applied Probability for Engineers and Scientists, McGraw Hill, New York, 1997.
35. E. Suhir, Polymer coated optical glass fibers: review and extension, Proceedings of the POLYTRONIK'2003, Montreaux, October, 21–24, 1997.
36. E. Suhir, V. Ogenko, and D. Ingman, Two-point bending of coated optical fibers, Proceedings of the PhoMat'2003 Conference, San-Francisco, CA, August 2003.

18

High-Speed Tensile Testing of Optical Fibers— New Understanding for Reliability Prediction

Sergey Semjonov[a] and G. Scott Glaesemann[b]

[a]*Fiber Optics Research Center at the General Physics Institute RAS, Moscow, Russia*
[b]*Corning Incorporated, Corning, NY, 14831 USA*

18.1. INTRODUCTION

Mechanical reliability of silica-based optical fibers in an optical communication system is limited by the fatigue effect. Flaws in glass subjected to tensile stress in the presence of moisture grow subcritically prior to failure [1]. The crack growth rate and respective time-to-failure depend on the initial crack size, applied tensile stress, and environment (temperature, humidity, pH) [2–4]. Estimating fiber lifetime for a commercial installation consists of two important steps:

- The determination of the flaw size distribution. In strength terms, this is the inert (no fatigue) strength distribution prior to the fatigue events that follow. The minimum inert strength represents the largest flaw in the distribution and, in the case of a proof tested fiber, is often used in the most conservative estimates of fiber lifetime.
- Calculation of time-to-failure for fiber with this starting strength distribution and stressed in fatigue environments typical of fiber and cable manufacturing, installation and in-service life.

A minimum strength is often established through tensile proof testing. The fiber is loaded to a proof stress, held for a short period of time and then unloaded [5]. Fatigue occurs even during short-term stress events like proof testing, and therefore, the minimum surviving strength after proof testing is determined by the proof stress and the time it takes to unload from the proof stress. Because of this, the minimum inert strength after proof testing can be less than the proof test stress [6]. It is important that the strength distribution used in fiber lifetime models incorporate fatigue during short-term events like proof testing.

The predicted failure rate during proof testing, processing, installation, and in-service lifetime are sensitive to the crack growth model for silica glass under stress. The most

frequently used model is a simple power law that relates crack growth rate to both the crack size and the applied stress [1]. Experiments using relatively large flaws in bulk silica show that at comparatively high crack growth rates (∼1–100 μm/s) the fatigue behavior transitions from a strong dependence on the stress intensity factor ($n_1 \sim 20$–40) to a much smaller dependence ($n_2 \sim 3$–10) [7,8]. This effect is attributed to crack growth that is reaction rate limited transitioning to growth that is diffusion limited. In other words, water molecules, responsible for rupture of stressed Si−O bonds in the crack tip, cannot penetrate to the crack tip fast enough to support higher growth rates. For long-term stress events it is well recognized that diffusion limited crack growth is too short to consider. However, one is justified in accounting for both regions for very short-term stress events [9].

Diffusion limited crack growth occurs at comparatively high crack growth rates. For fiber, where strength testing is generally used as a indirect measure of crack growth, the effects of region II-like crack growth require loading rates greater than 10^4 GPa/s. In the case of pristine high strength fibers, measuring the second region of fatigue through strength testing is practically unfeasible [10]. Such testing speeds are unattainable with the existing test equipment [11]. Fortunately, this region can be observed at slower testing speeds using weak fibers. Using test equipment capable of loading rates of 10 to 100 GPa/s tensile strengths close to those in liquid nitrogen have been achieved with n values of approximately 5 at the faster rates [12].

In addition to modeling high speed events characterizing the behavior of proof stress level flaws has been an area of study. Lifetime modeling of optical fiber has historically been based on the results and parameters obtained from strong, "flaw-free" fibers. The fatigue parameter $n \sim 20$ obtained for strong fibers is generally used for both calculation of the fiber strength after proof testing as well as the evaluation of strength degradation during the in-service lifetime. However, the reliability of optical fibers in most communication systems depends on relatively large (∼1 μm) flaws that survive proof testing at ∼0.7 GPa. Only a few flaws of this kind exist on multi-kilometer fiber lengths. Due to their rarity fatigue of these flaws is best characterized with the use of artificially induced flaws. In order to draw meaningful results the flaw inducement method has to be reproducible, resulting in a narrow strength distribution. Furthermore, it is well known that fiber has multiple flaw types ranging from mechanical damage to partially submerged particulate on the surface. It is important that the mechanical behavior of various flaw types be understood and accounted if the lifetime model is to be of practical use.

This publication contains both an overview of a five-year collaboration between Corning Inc. (USA) and Fiber Optics Research Center (Russia) on testing of weak fibers [13–16] and a discussion about the effect of the multi-region crack growth model on the fiber lifetime prediction in optical communication systems.

18.2. THEORY

18.2.1. Single-Region Power-Law Model

According to the widely used single-region power-law model, a crack of depth a grows under applied stress σ in the following way [1]:

$$\frac{da}{dt} = AK_{\mathrm{I}}^{n}, \tag{18.1}$$

where K_I is the stress intensity factor:

$$K_I = Y\sigma(a)^{1/2}, \tag{18.2}$$

A and n are fatigue parameters related to the environment; Y is the geometry factor ($\sim$1) and dependent on the crack geometry.

The strength for any crack size is:

$$S = \frac{K_{IC}}{Y\sqrt{a}}, \tag{18.3}$$

where K_{IC} is the critical stress intensity factor (for silica glass $K_{IC} = 0.79$ MN/m$^{3/2}$). The initial strength, S_i is the inert strength prior to the fatigue event being modeled. "Inert" refers to a test where no crack growth takes place during the test itself. In the case of optical fiber, vacuum and liquid nitrogen have been used as inert test conditions as well as extremely high testing speeds. The strength after a crack growth event is denoted as S_f. From Equations (18.1)–(18.3) it can be derived that the decrease in strength from an event consisting of stress $\sigma(t)$ over time t, is [6]:

$$B\left(S_i^{n-2} - S_f^{n-2}\right) = \int_0^t \sigma^n(\tau)d\tau, \tag{18.4}$$

where

$$B = \frac{2}{(n-2)AY^2K_{IC}^{n-2}}. \tag{18.5}$$

In the case of an applied static stress σ_s, the time-to-failure is deduced from Equations (18.4) and (18.5):

$$t_s = \frac{BS_i^{n-2}}{\sigma_s^n}, \quad \text{assuming } \sigma_s^{n-2} \ll S_i^{n-2}. \tag{18.6}$$

Thus, a flaw with initial strength S_i will grow and fail in time t_s under static stress σ_s. In the case where one conservatively designs around the weakest flaw in proof tested fiber, the minimum initial strength is taken to be the proof stress level in Equation (18.6). However, there is the chance that a flaw that just passes the proof stress grows during unloading and survives to be weaker than the proof stress [6]. The probability of this happening depends on the quality of the strength distribution. It is, therefore, important to characterize the strength distribution near the proof stress level. The following equation is frequently used to estimate the fiber lifetime t_s under service stress σ_s with some failure probability F provided that pre-proof flaw statistics are known [17]:

$$\sigma_s^n t_s = \sigma_p^n t_p \left\{ \left[1 - \frac{\ln(1-F)}{N_p L} \right]^{\frac{n-2}{m}} - 1 \right\}, \tag{18.7}$$

where t_p is time during which each point of the fiber experiences the proof stress, L is the fiber length, N_p is the mean number of breaks per length during proof testing, m is Weibull parameter characterizing the flaw statistics of the fiber before proof testing.

Other lifetime models are widely used that, similar to Equation (18.7), are based on the single-region power-law and pre-proof initial strength described by Weibull parameters [5]. They give realistic requirements on the service stress in a cable for a 25 years service time ($\sigma_s \sim 0.2$–$0.25\,\sigma_p$) for a low failure probability ($F \sim 0.01$–0.001). Also, some do not use the fatigue parameter B, which is difficult to measure. On the other hand, this set of statistical lifetime models have some disadvantages:

1. The Weibull parameter m, should be obtained for flaws near the proof stress level. This requires a commitment to strength testing many kilometers of fiber [18].
2. The fiber break rate N_p is usually confidential. Also, some have questioned the assumption that the distribution of failing defects describes those that survive [19].
3. Experiments with large flaws in bulk silica and with weak fibers show that the crack growth rate demonstrate a complex multi-region behavior for high speed events [7,8]. This can have a significant influence on the determination of B [20] and the estimated post-proof strength distribution.

For these reasons, another approach to lifetime estimation (so-called safe stress model) has been developed [21]. It is based on a two-region power-law model.

18.2.2. Two-Region Power-Law Model

According to this model, the first and the second regions have their own fatigue parameters n_1, A_1 and n_2, A_2, respectively (Figure 18.1):

$$\frac{da}{dt} = A_1 \left(\frac{K_\mathrm{I}}{K_\mathrm{IC}} \right)^{n_1} \quad \text{for } K_\mathrm{I} \le r K_\mathrm{IC}, \tag{18.8}$$

$$\frac{da}{dt} = A_2 \left(\frac{K_\mathrm{I}}{K_\mathrm{IC}} \right)^{n_2} \quad \text{for } K_\mathrm{I} \ge r K_\mathrm{IC}, \tag{18.9}$$

where r is the parameter characterizing the point of transition from one region to the other.

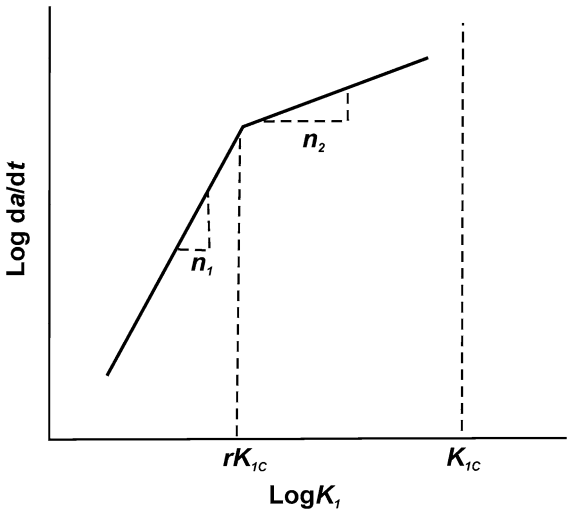


FIGURE 18.1. Two-region dependence of the crack growth rate on stress intensity factor [21].

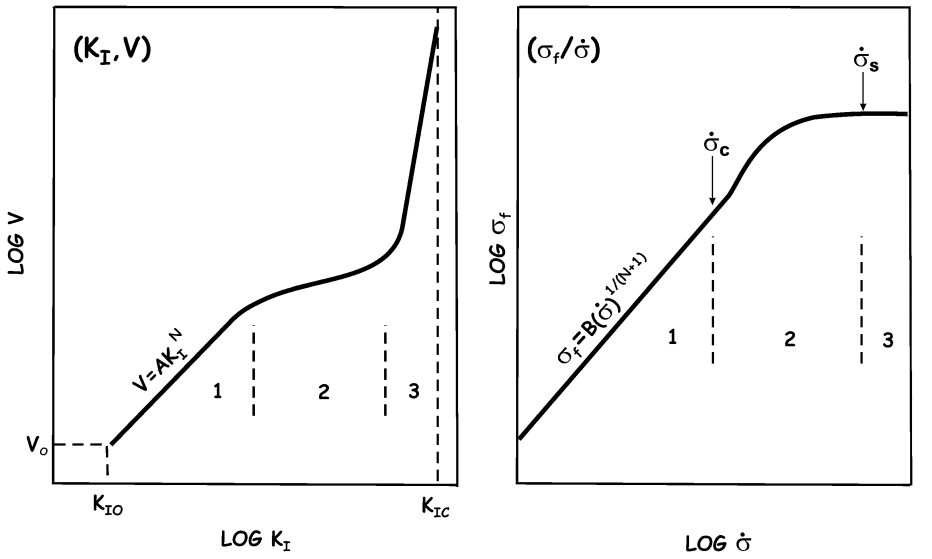


FIGURE 18.2. Influence of Regions II and III on the dependence of the breaking strength on loading speed (dynamic fatigue) [22].

These equations can be written for the case of dynamic fatigue tests in terms of the changing inert strength, S, and applied stress σ, using the usual relations, $K_{\mathrm{I}} = Y\sigma a^{1/2}$ and $K_{\mathrm{IC}} = YSa^{1/2}$:

$$S_i^{n_1-2} = S_r^{n_1-2} + \frac{\sigma_r^{n_1+1}}{B_1(n_1+1)\sigma'} \quad \text{for flaw growth through Region I,} \tag{18.10}$$

$$S_r^{n_2-2} = \sigma_d^{n_2-2} + \frac{\sigma_d^{n_2+1} - \sigma_r^{n_2+1}}{B_2(n_2+1)\sigma'} \quad \text{for flaw growth through Region II,} \tag{18.11}$$

where S_i is the initial strength before loading, S_r and σ_r are the inert strength and applied stress when $K_{\mathrm{I}}/K_{\mathrm{IC}} = \sigma_r/S_r = r$, σ_d is the dynamic fatigue strength for stress rate σ', $1/B_2 = \nu_2(Y/K_{\mathrm{IC}})^2(n_2-2)/2$ and $B_1 = B_2 \cdot r^{(n_1-n_2)}(n_2-2)/(n_1-2)$. Unfortunately, the fatigue parameters must be deduced from by a numerical search where one varies all the fatigue parameters to find the best fit to experimental data.

Calculations based on the multi region model [22] show that the presence of region two type growth results in an increased slope of the dynamic fatigue curve at fast rates (Figure 18.2). For pristine fibers, this slope change would begin at loading rates approaching 10^4 GPa/s and 10–100 GPa/s in a case of weak fibers.

18.2.3. Universal Static and Dynamic Fatigue Curves

The shape of the fatigue curve depends on the strength regime of the flaws under study and this complicates the comparison of fatigue behavior of samples with different initial strength. For this reason, a method for comparing fatigue data of weak and strong fibers was developed [14].

In earlier papers [23–25], a so-called "universal fatigue curve" was used to correlate the results of static fatigue measurements for different kinds of samples. Coordinates $(t/t_{0.5}) - (\sigma/S_i)$ were used for this purpose, where t is the time-to-failure under stress σ, S_i is the initial inert strength of the sample, $t_{0.5}$ is the time-to-failure at $\sigma = 0.5S_i$. There were several reasons to continue development in this direction:

- the above "universal" presentation was deduced using simple power or exponential laws of crack growth;
- it is useful only for static fatigue measurements;
- it requires knowledge of an additional parameter $t_{0.5}$.

Therefore, it was desirable to solve the problem of correlating fatigue measurements of different fibers in a more general form.

The effect of subcritical crack growth or fatigue in brittle materials, such as glass or ceramics, is usually described as a dependence of crack growth rate on stress concentration at the crack tip with fracture mechanics as a framework:

$$\frac{da}{dt} = \Phi\left(\frac{K_{\mathrm{I}}}{K_{\mathrm{IC}}}\right), \tag{18.12}$$

where K_{IC} is the critical stress intensity factor and is a material property. Power and exponential laws are often used in Equation (18.12), but we will deal with an arbitrary positive function defined in the range $0 < K_{\mathrm{I}} < K_{\mathrm{IC}}$.

Two types of test methods are widely used to observe fatigue effects; namely, static and dynamic. In the first case, constant stress σ_s is applied to the sample, and time-to-failure t_s is measured. In the second test, an increasing load with a constant loading rate $\sigma' = d\sigma/dt$ is applied to the sample, until it breaks at some load σ_d. The dependence of time-to-failure on the applied stress is called static fatigue; the dependence of breaking strength on the loading speed is called dynamic fatigue. Both dependencies can be derived from Equation (18.12) with $K_{\mathrm{I}} = K_{\mathrm{IC}}$ defining failure.

Combining Equations (18.2) and (18.3),

$$k = \frac{K_{\mathrm{I}}}{K_{\mathrm{IC}}} = \frac{\sigma}{S}, \tag{18.13}$$

and thus,

$$a = \left(\frac{k \cdot K_{\mathrm{IC}}}{Y \cdot \sigma}\right)^2. \tag{18.14}$$

For *static fatigue* ($\sigma = const. = \sigma_s$) Equation (18.12) can be transformed into,

$$d\left(\frac{k \cdot K_{\mathrm{IC}}}{Y \cdot \sigma_s}\right)^2 = \Phi(k) \cdot dt. \tag{18.15}$$

At the beginning of the test the initial crack size is a_i, the corresponding initial inert strength is S_i, and the following relation is created:

$$k_i = \frac{\sigma_s}{S_i}. \tag{18.16}$$

Thus,

$$dt = \left(\frac{K_{IC}}{Y \cdot k_i \cdot S_i}\right)^2 \frac{dk^2}{\Phi(k)}. \tag{18.17}$$

The solution of Equation (18.17) is

$$t_s \cdot S_i^2 = \left(\frac{K_{IC}}{Y \cdot k_i}\right) \cdot \int_{k_i}^{1} \frac{2k \cdot dk}{\Phi(k)} = \Psi(k_i) = \Psi\left(\frac{\sigma_s}{S_i}\right). \tag{18.18}$$

In the case of a single-region power law, the well-known solution of Equation (18.17) for $S_i^{n-2} \gg \sigma_s^{n-2}$ is

$$S_i^2 \cdot t_s = B \cdot \left(\frac{\sigma_s}{S_i}\right)^{-n}, \tag{18.19}$$

where n and B are the fatigue parameters.

For *dynamic fatigue* ($\sigma = t \cdot \sigma'$), Equation (18.12) can be transformed into

$$d\left(\frac{k \cdot K_{IC}}{Y \cdot \sigma}\right)^2 = \Phi(k) \cdot \frac{d\sigma}{\sigma'}. \tag{18.20}$$

By introducing a term $\nu = (\sigma / S_i)$, we obtain

$$\left(\frac{K_{IC}}{Y \cdot S_i}\right)^2 \cdot d\left(\frac{k}{\nu}\right)^2 = \Phi(k) \cdot \frac{S_i}{\sigma'} \cdot d\nu, \tag{18.21}$$

or

$$\frac{\sigma'}{S_i^3} \cdot d\left(\frac{k^2}{\nu^2}\right) = \left(\frac{Y}{K_{IC}}\right)^2 \Phi(k) \cdot d\nu, \tag{18.22}$$

where k lies in the range from 0 to 1, and ν from 0 to (σ_d / S_i). The analytical or numerical solution of Equation (18.22) will give the following dependence:

$$\frac{\sigma_d}{S_i} = \chi\left(\frac{\sigma'}{S_i^3}\right). \tag{18.23}$$

For a single-region power law, the solution for $S_i^{n-2} \gg \sigma_d^{n-2}$ is

$$\frac{\sigma_d}{S_i} = \left(B \cdot (n+1) \cdot \frac{\sigma'}{S_i^3}\right)^{\frac{1}{n+1}}. \tag{18.24}$$

Thus, if cracks of different size have the same dependence of the crack growth rate on the stress intensity factor [Equation (18.12)], the static fatigue results from tests of flaws with different initial strengths would lie on one curve in $(t_s S_i^2)$–(σ_s / S_i) coordinates. Accordingly, all of the dynamic fatigue results would lie on one curve in (σ_d / S_i)–(σ' / S_i^3) coordinates. In other words, universal static and dynamic fatigue curves can be drawn.

By the following substitution

$$\begin{cases} x = \sigma'/S_i^3 \\ y = S/S_i. \end{cases} \tag{18.25}$$

Equations (18.10) and (18.11) describing multi-region strength degradation can be transformed into a system of equations:

$$\begin{cases} 1 = y_r^{n_1-2} + \dfrac{r^{n_1+1} y_r^{n_1+1}}{B_1(n_1+1)x} \\ y_r^{n_2-2} = y_d^{n_2-2} + \dfrac{y_d^{n_2+1} - r^{n_2+1} y_r^{n_2+1}}{B_2(n_2+1)x}, \quad \text{where } y_d < y_r < 1. \end{cases} \tag{18.26}$$

The solution of this system $y_d(x) = \sigma_d/S_i(\sigma'/S_i^3)$ describes the dynamic fatigue behavior of the fiber in universal coordinates. Unfortunately, like the case of Equations (18.10) and (18.11), there is no exact analytical solution for this function. Numerical simulation of this function by varying all the fatigue parameters has to be performed to achieve the best fit to experimental data.

18.3. EXPERIMENTAL

18.3.1. Sample Preparation

The first step of the experimental work was to develop methods for producing weak fibers with defects of different origin having a size of ~1 μm and a narrow strength distribution. Three techniques were used to create defects on the fiber surface—abrasion, seeding the preform with ZrO_2 particles and indentation. For comparison, pristine fibers were tested as well. At ultra-high loading rates the failure load on standard 125 mm diameter pristine fiber can be in excess of 15 kg. Maintaining a grip on the fiber at these loads can be difficult; and therefore, specially made 40 μm fiber were used to avoid this issue.

18.3.1.1. Abrasion The silica-clad fiber was manufactured in a conventional manner with the exception that it was intentionally abraded during draw to a strength level near typical proof stress levels. The abrasion technique consisted of three stationary silica rods in contact with the fiber above the polymer applicator during the drawing process (Figure 18.3). The polymer coating was a commercial available one. The abrasion process (position of the rods) was optimized to yield a Weibull modulus of 20 on 10-cm long samples.

18.3.1.2. Seeding by Particles Weak fibers were created by drawing a preform whose surface had been seeded with zirconia powder (~1 μm particle size). To make the distribution of particles on the preform surface more homogeneous, the preform was washed in an alcohol solution containing the zirconia powder. SEM study of the ends of broken samples show that fragments of aggregated particles are transformed by the drawing process into a long groove on the fiber surface. These grooves contain a series of melted-in particles (Figure 18.4) with one of the particles providing the source for fracture. The density of particles in the solution was systematically changed to optimize the strength distribution for 10-cm long samples. A Weibull modulus, m, of approximately 20 was obtained.

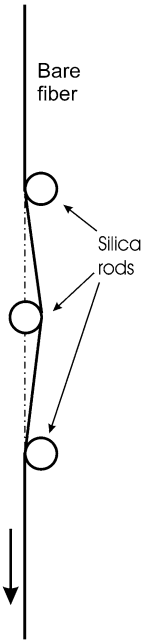


FIGURE 18.3. Process of abrasion during the fiber drawing.

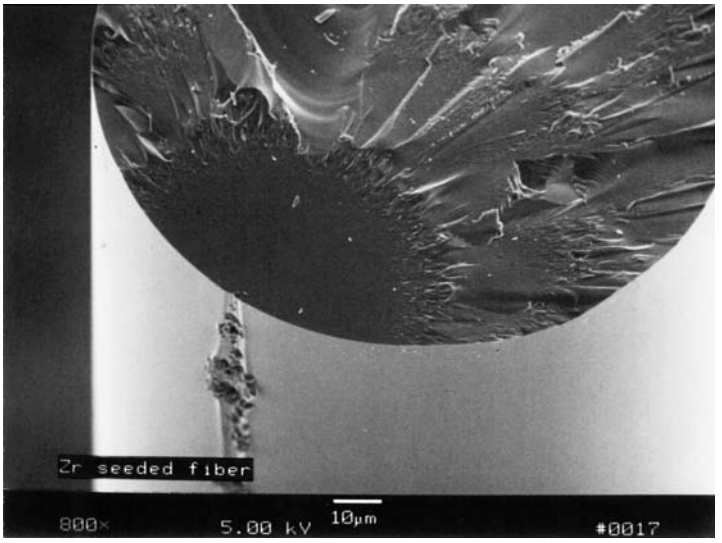


FIGURE 18.4. SEM micrograph of the end of broken fiber with seeded particles.

18.3.1.3. Indentation The reproducibility of indentation flaws is difficult and has been the subject of several investigations [26–28]. Factors that affect the reproducibility of Vickers indentation flaws in glass are the repeatability of the indentation loading cycle, the presence of secondary cracking (lateral and Hertzian), the crack initiation threshold, fatigue during indentation and loading to failure, as well as specimen and indenter cleanliness

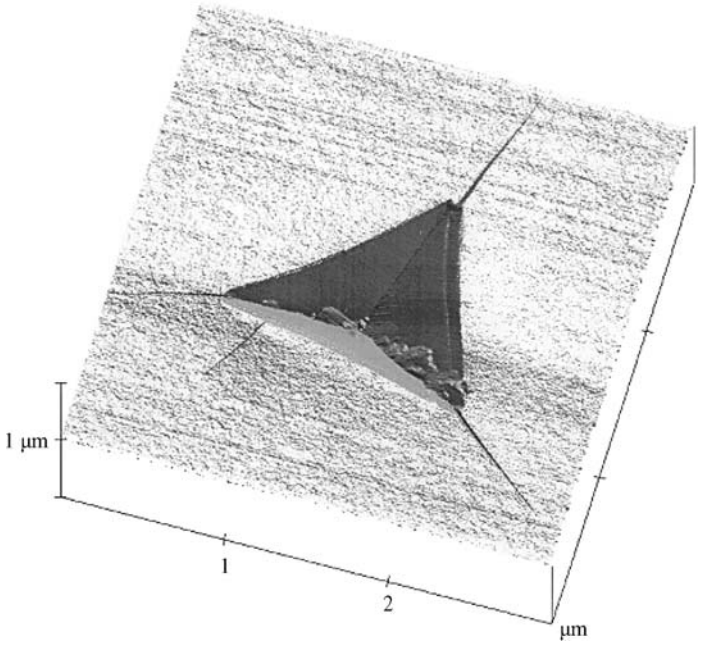


FIGURE 18.5. AFM image of 1 gram indent with a cube-corner indenter.

[29–32]. For this reason flaws in fiber made with Vickers Indentation are not ideal model flaws for the study of fatigue of proof stress level flaws in fiber (∼0.7 GPa strength and ∼1 μm long cracks). Recent work [13,28,33,34] has shown that micron size crack lengths can be produced using a cube-corner-shaped diamond and indentation loads of ∼ 1 gram. A cube-corner indenter displaces three times as much volume for a given contact area than a Vickers indenter, and thus, generates greater contact residual stress. A typical AFM image of 1 gram indent is presented in Figure 18.5. Using this technique a highly reproducible flaw size was created and the needed narrow strength distribution was achieved [13,28,34].

To prepare samples for indentation, a 25 mm length of polymer coating was removed from the middle of a meter length of a standard 125 micron diameter single-mode silica clad fiber. The coating was removed by exposing the fiber to 180 to 200°C sulfuric acid for 10 to 20 seconds. The fiber was then rinsed in distilled water for 5 seconds before indentation. To stabilize the fiber during the indentation process, it was fixed in a grooved glass substrate [34]. A nano-indentation apparatus (Nano-indenter II, MTS Systems Corp.) equipped with a cube-corner indenter was used to create proof-stress level flaws on the fiber. A Weibull modulus of ∼50 is typical with this indentation method and is close to that of pristine fibers.

18.3.2. Dynamic Fatigue Tests

In this study three loading techniques were used to generate strength values over nearly eight decades of stressing rates. For the slower speed tests (3×10^{-5} to 0.7 GPa/s), a conventional universal testing machine (lnstron Corp.) was adapted with eight load cells and corresponding capstans. Thus, instead of the usual single fiber testing method, eight fibers could be tested in tension simultaneously. This greatly decreased the overall experi-

mental time at the slower rates. A belt slide apparatus (Parker Hannifin Corp.) was used to generate stressing rates ranging from 1 to 100 GPa/s. For the highest load rates, a pneumatic piston (Airmatic-Allied, Inc.) was fitted with a gas reservoir at the inlet to provide a ready volume of air for the piston chamber. Nitrogen pressure levels ranging from 40 to 95 psi (280 to 665 kPa) were used to achieve stressing rates ranging from 100 to 10,000 GPa/s).

Of primary importance to high speed testing is the mass of the fiber attachment system and the method of data acquisition. All tests on the belt slide and air piston were performed with two load cells in place, a conventional lightweight strain gauge load cell (Interface, Inc., Scottsdale, AZ) and a piezoelectric load cell (Kistler Instrument Corp.). The piezoelectric load cell was chosen such that drift and resonant frequency problems were minimized for the range of failure times used in this study.

The data acquisition rate for the strain gauge load cell at the highest speeds was 40,000 Hz. At the highest loading rate the failure times were on the order of 10^{-3} seconds; and therefore, the number of data points using this load cell was sufficient. The signal from the piezoelectric load cell, on the other hand, was monitored at 5×10^6 Hz using a digital oscilloscope. This yielded several thousand data points per test at the highest speed. The stressing rate was taken from the last 20% of the loading curve.

Fiber was attached to both load cells by carefully taping the fiber to a nylon screw that was threaded directly into the load cell. The total weight of the screw and tape was approximately 1 gram. Fiber pullout from the tape was not an issue since the maximum loads were sufficiently low with the weak fiber.

Mathematical simulation (see Appendix A) demonstrated that the observed results at the highest stressing rates could be incorrect due to oscillation inside the load cell during high-speed testing. Taking into account the parameters of actual oscillation observed in high-speed experiments, the results of the tests at the highest possible speed of the air piston were excluded from further consideration if testing time was less than $\sim$150 μs.

In this study, 10 cm long coated fiber samples were tensile tested. To avoid premature failure where the fiber is gripped, weak fiber ends were glued to pieces of strong fibers by epoxy. Only failures in the gauge length were accepted. 20 samples were usually tested at each stressing rate in the case of abraded and Zr-seeded fibers, and only 10 specimens were tested in case of indented samples owing to their narrow strength distribution. A small amount of slack in the fiber gauge length was introduced at the higher loading rates to allow the test device to reach its maximum speed before fiber loading. All the tests were performed in an environment controlled to 23°C and 50% RH.

In addition to fatigue testing the inert strength was measured in liquid nitrogen. A small horizontal thermo-isolated bath, with narrow slots for introducing the fiber sample was filled with liquid nitrogen. The whole length of the weak sample was immersed in the liquid, and only strong fibers glued to the weak sample extended out of the bath. The fiber's acrylate coating becomes stiff and brittle at such low temperatures, and thus, was removed with hot sulfuric acid prior to testing. A silicone rubber sealant was used to glue pieces of weak fiber to strong fiber. Too soft for testing at room temperature, this sealant became hard but not brittle in liquid nitrogen. As with the fatigue testing, only failures in the gauge length were accepted.

18.3.3. Static Fatigue Tests

To compare dynamic fatigue results with longer-time static fatigue measurements, indented fibers were also tested in static fatigue. In this case, the ends of the fiber sample

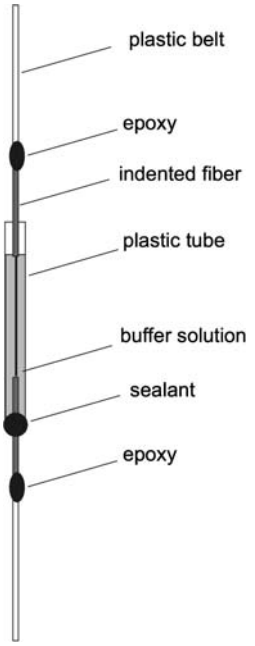


FIGURE 18.6. Schematic of static fatigue measurements in buffer solution.

were glued to polymer belts with epoxy. The fiber sample was fixed to a holding frame using a top belt. The free end of the bottom belt was attached to a loading weight. The reduced loads for large flaws simplifies the experimental setup compared to the testing of strong fibers, where special holding capstans are required to apply static weight to the fiber. Static tests were performed in a room with stabilized temperature and humidity (23°C; 50% RH). To test in various pH solutions, the fiber was inserted into a small polymer tube that covered the indented region of the fiber. The bottom part of the tube was sealed and the tube was filled with the desired pH solution (Figure 18.6). The total weight of the tube and solution did not exceed 1 gram. Samples prepared for testing in this manner were kept overnight in the pH solution before load application. The pH level in the tubes was checked periodically during static testing. The time to failure was recorded on videotape. Ten samples were tested in the same environment at the same load.

18.4. RESULTS AND DISCUSSION

18.4.1. High-Speed Testing

Typical strength distributions for each stressing rate for the zirconia-seeded weak fiber are shown in Figure 18.7. This rather narrow strength distribution enabled us to accurately view the strength variation with stressing rate.

Plots of the median strength of seeded and abraded samples tested in laboratory environment (50% RH, 23°C) are shown in Figure 18.8. Both plots demonstrate a nonlinear strength versus stressing rate behavior when plotted in the usual power-law fashion. A direct comparison of the fatigue behavior of the two fibers in Figure 18.8 is hampered by the difference in initial strength between the two fibers.

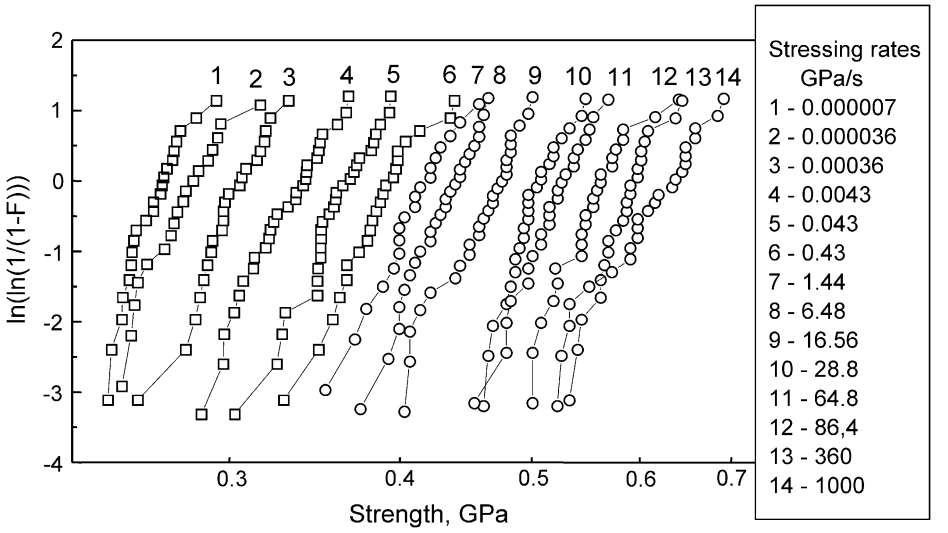


FIGURE 18.7. Weibull plots of zirconia seeded fiber strength measured at different stressing rates in laboratory environment.

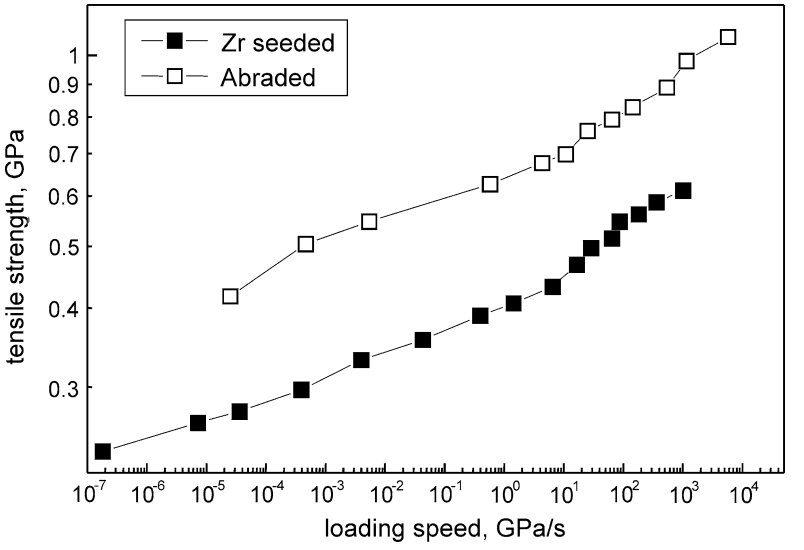


FIGURE 18.8. Median strength of seeded and abraded samples tested in laboratory environment (50% RH, 23°C).

It has been shown in Section 18.2.3 that $(\sigma_d/S_i) - (\sigma'/S_i^3)$ are universal coordinates for presentation of dynamic fatigue data. Here σ_d is the median strength of a fiber at stressing rate σ', and S_i is the initial (inert) strength of the same fiber. With these coordinates one can compare fatigue data for any initial defect size. The strength measured in liquid nitrogen (Figure 18.9) was used for determinating S_i. The dynamic fatigue data for seeded and abraded fibers is shown in universal coordinates in Figure 18.10. Both fibers exhibit very similar non-linear fatigue behavior despite the difference in flaw type and initial strength.

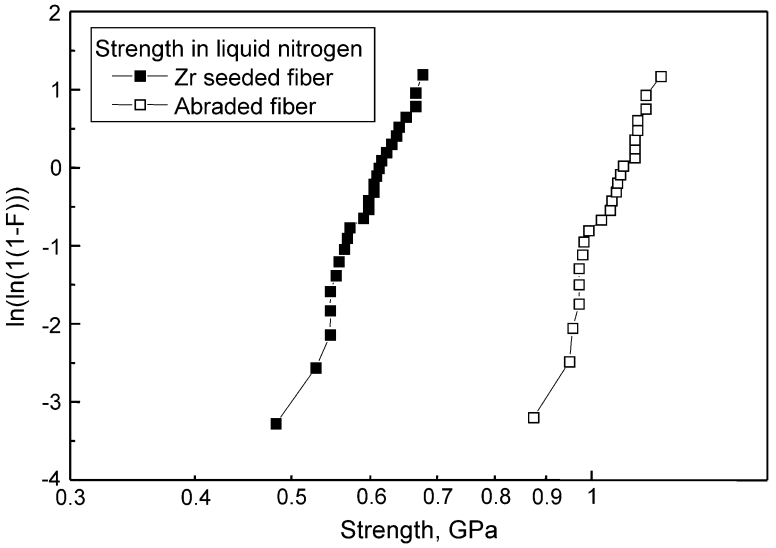


FIGURE 18.9. Weibull plots of strengths of zirconia seeded fiber and abraded fiber measured in liquid nitrogen.

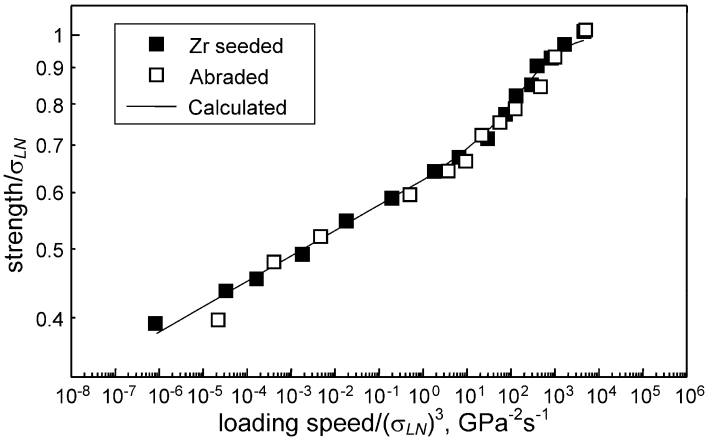


FIGURE 18.10. Median strength of seeded and abraded samples tested in laboratory environment (50% RH, 23°C) replotted in "universal" coordinates.

The abraded and contaminated fiber results are compared with similar test results on indented and pristine fibers in Figure 18.11. Unfortunately, measuring the liquid nitrogen strength of strong pristine fiber was problematic owing to problems with gripping the fiber. Thus, previously published values of 12 GPa and 14 GPa were used [35,36].

Comparison with high strength fiber is somewhat limited due to testing difficulties of pristine fibers at high speeds, but overall there is good agreement between fibers of various strength. At slower speeds there is some discrepancy between high strength and low strength results. It has been hypothesized that the non-linear fatigue behavior at the fast stressing rates can be attributed to the influence of region II crack growth at short times to failure. The multi-region crack growth parameters are given in Table 18.1.

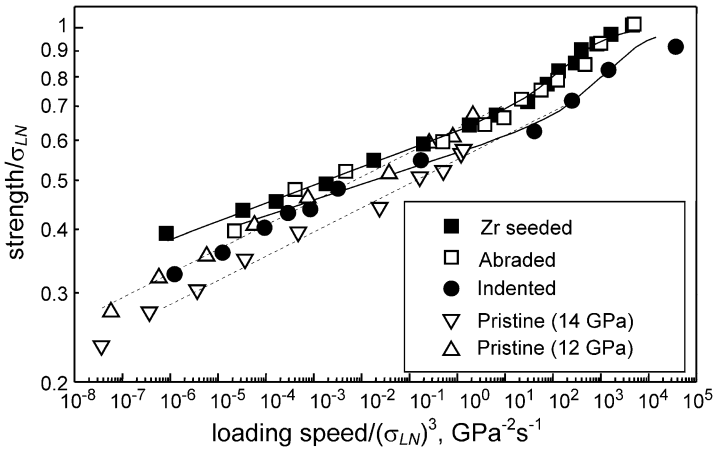


FIGURE 18.11. Joint plot of dynamic fatigue results obtained for different types of weak fibers and pristine fiber.

TABLE 18.1.
Fatigue parameters for different types of weak fibers and a pristine fiber.

Seeded and abraded	Indented fiber	Pristine fiber
$n_1 = 27$	$n_1 = 31$	$n_1 = 20$
$B_1 = 0.7 \times 10^{-7}$ GPa^2 s	$B_1 = 0.4 \times 10^{-9}$ GPa^2 s	$B_1 = 3.1 \times 10^{-6}$ GPa^2 s
$r = 0.757$	$r = 0.70$	($S_i = 12$ GPa)
$n_2 = 0.5$	$n_2 = 3.0$	$B_1 = 1.6 \times 10^{-7}$ GPa^2 s
$B_2 = -1.9 \times 10^{-3}$ GPa^2 s	$B_2 = 2.5 \times 10^{-4}$ GPa^2 s	($S_i = 14$ GPa)
		$r =$ not available
		$n_2 =$ not available
		$B_2 =$ not available

The shape of the fatigue curve generated by the indented samples in Figure 18.11 is close to that of the abraded and contaminated fibers. This suggests that the fatigue of flaws in silica is independent of the flaw introduction mechanism and that the form of the crack velocity function for surface flaws in silica-clad optical fiber is independent of flaw type. This is important since it simplifies reliability modeling of the multiple flaw populations known to exist in optical fiber. One does not have to have a separate model for each flaw type.

In universal coordinates the indented strength is somewhat lower than that of the contaminated and abraded specimens. There are several possible factors that could contribute to this difference. The three factors discussed here represent areas of further investigation. First, the role of the contact-induced residual stress of the indented specimens was not taken into account in determining the failure stress. The residual stress is known to promote subcritical crack growth in fatigue environments, and therefore, yield lower fatigue strength. The mechanics have been developed for such determinations [37–39], but accurate quantification of residual stress levels is lacking. It should be noted that the partially submerged zirconia particles on the contaminated fiber also create significant residual stresses in the host silica [40]. Unfortunately, crack growth mechanics that

incorporates the influence of residual stresses associated with contaminants is not fully developed.

The second explanation for the lower universal fatigue strength for the indented specimens is crack geometry. An indirect method for assessing the role crack geometry would be to perform the same tests using cube corner indents [13,28,34]. The flaws associated with cube corner indentation are on the order of 1 μm in size with virtually no residual stress.

Thirdly, the indented specimens are not covered with polymer coating. The local environment for the coated flaws would certainly be different than that of uncoated flaws. Even if the local environment of a coated flaw were known, the influence of this environment on crack growth kinetics is unresolved. An experimental approach would at least yield empirical understanding of the coating influence for the data presented here.

Finally, the strength at the fastest stressing rates is the same as that measured in liquid nitrogen. This result suggests that the effect of low temperature on the mechanical properties of silica does not effect the strength significantly. Fracture toughness of bulk glass is known to be effected by temperature [41,42] and the liquid nitrogen strength of pristine fibers is known to differ from that measured in liquid helium [32].

18.4.2. Static Fatigue

Fatigue parameters obtained from high stressing rate experiments are believed to provide a more accurate prediction of the post-proof strength needed in reliability predictions (the first step). For the second step of lifetime estimation, information about fatigue behavior of relatively large flaws at very slow loading speed or during long-term static tests is required. For this aim, long-term static fatigue testing of low-strength (indented) optical fibers was performed in an effort to quantify the long-term fatigue parameters. Typical Weibull plots of static fatigue results are shown in Figure 18.12.

Results of static fatigue and dynamic fatigue (high-speed) tests were combined in one graph using consideration on recalculating static fatigue data into dynamic fatigue data

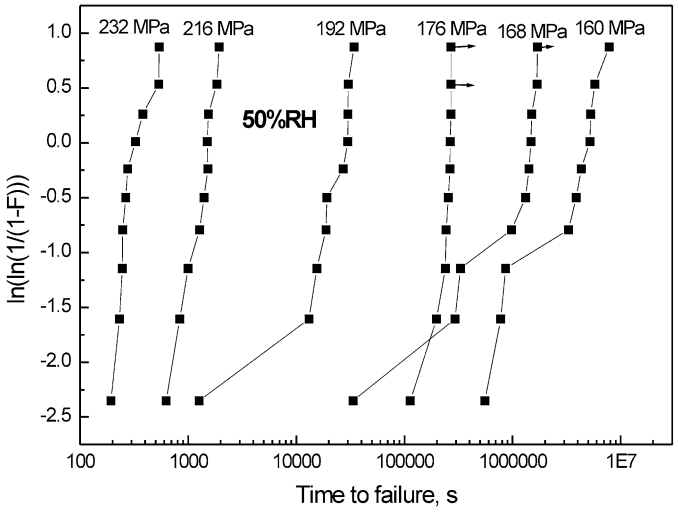


FIGURE 18.12. Weibull plots of time to failure of indented fibers at different loads (static fatigue tests) at 50% RH.

from Appendix B. Figure 18.13 demonstrates very good correlation between both types of tests.

The test environment of 50% RH and 23°C is typical of proof testing. Thus, for determining the minimum surviving strength of the fiber after proof testing, only results of high-speed dynamic tests obtained in this environment are required. For predictions for long-term in-service conditions, it is common to quantify fatigue in more severe environments. For many applications the worst case environment is 100% relative humidity combined with different pH levels. To demonstrate changes in fatigue parameters related with changes in environment, static fatigue testing of indented fibers in different pH-solutions was performed.

The results of static fatigue testing of indented fibers at 50% RH and in different pH solutions are presented in Figure 18.14. Calculation of the n-value was performed using the following equation:

$$\ln\left(\frac{t_1}{t_2}\right) = n \ln\left(\frac{\sigma_2}{\sigma_1}\right), \tag{18.27}$$

where t_1, t_2 and σ_1, σ_2 are the times to failure and applied stresses, respectively. The n-value was calculated to be 20.0 ± 1.0 for all solutions.

To calculate the other fatigue parameter, B, the following equation was used:

$$B \approx \frac{\sigma_s^n t_s}{\sigma_i^{n-2}}, \tag{18.28}$$

where t_s and σ_s are the results of a static fatigue test and σ_i is the initial or inert strength of the fiber, which is close to the strength in liquid nitrogen. The calculated fatigue parameters for the various test environments are given in Table 18.2.

As shown in Figure 18.14, the time to failure for a given load decreases when the conditions change from 50% RH to the water solutions. If the pH parameter of the solution changes from acidic to base (a pH increase), the failure time also decreases. This

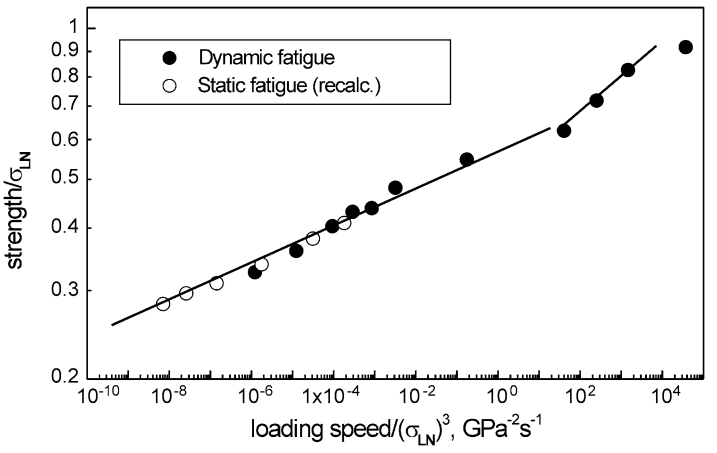


FIGURE 18.13. Combined plot of dynamic fatigue and recalculated static fatigue results obtained for indented fibers in a laboratory environment (50% RH, 23°C).

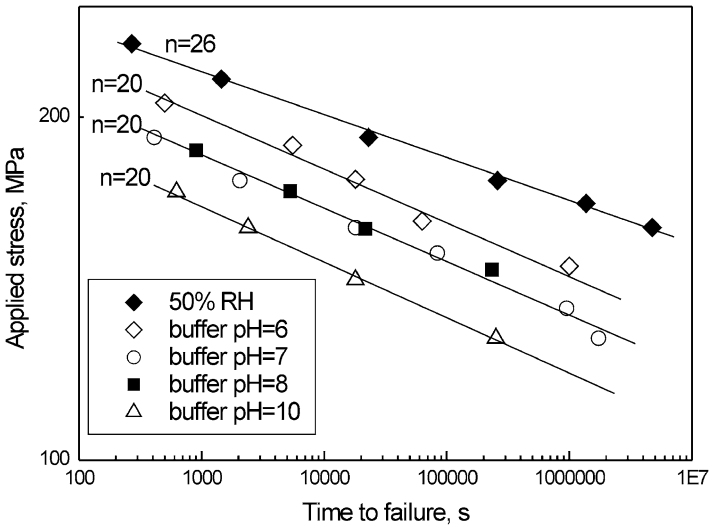


FIGURE 18.14. Dependencies of median time to failure on applied load (static fatigue) for indented fiber samples in different environments.

TABLE 18.2.
Fatigue parameters of indented fiber.

Conditions	n	$\log_{10} B$ (GPa^2 s)	B (GPa^2 s)
50% RH	26 ± 1	-8.15 ± 0.5	7.1×10^{-9}
pH = 6	20 ± 1	-6.6 ± 0.5	2.5×10^{-7}
pH = 7 & pH = 8	20 ± 1	-7.3 ± 0.5	5.0×10^{-8}
pH = 10	20 ± 1	-8.05 ± 0.5	8.9×10^{-9}

corresponds with the general opinion that an increase in the concentration of OH groups activates the reaction between water and silica glass. There are two detailed publications [43,44] on the fatigue of bare fibers in different pH solutions. In both studies, a reduction of the n-value with increasing pH was observed. In another study [45] the behavior of crack velocity in various pH conditions was investigated in bulk glass samples. The n-value from these crack velocity experiments decreased from 47 to 22 when the pH increased from 2 to 12.

In contrast to the crack velocity studies on bulk glass, the n-value in this study showed no dependence on pH. Furthermore, the n-value of 26 at 50% RH from this study is higher than typically measured on pristine or abraded fibers in dynamic fatigue [46]. However, the n-value of pristine fiber is known to increase from about 20 in short-term dynamic fatigue tests to the mid 20's at very slow stressing rates in dynamic fatigue [47,48]. When tested on the same time scale, fatigue of fiber is independent of the test method. For the test conditions in this study, the measured n is lower than that for bulk samples in similar conditions. All lifetime reduction with pH increase was due to a change of the B-value. The nature of this difference is not quite clear and requires additional investigation.

TABLE 18.3.
Relationship between the service stress and the initial strength required for a 30-year lifetime.

Conditions	σ_s/σ_i ($\sigma_i = 0.7$ GPa or 1% elongation)	σ_s/σ_i ($\sigma_i = 1.4$ GPa or 2% elongation)
50% RH	0.225	0.213
pH = 6	0.174	0.162
pH = 7 and pH = 8	0.161	0.151
pH = 10	0.147	0.137

18.4.3. Influence of Multiregion Model on Lifetime Prediction

The static fatigue results describe fatigue behavior of bare indented fibers in different environments for failure times in the range of 10^3–10^7 seconds. Figure 18.14 shows that the fatigue curves demonstrate quite linear behavior in this range; and therefore, Equations (18.27) and (18.28) were used to describe their behavior. Supposing that the behavior of the fatigue curves maintain this trend until 10^9 seconds (∼30 years), one can make an evaluation of the fiber lifetime in the investigated conditions directly using Equation (18.27). The relationship between the service stress, σ_s, and the initial strength, σ_i, required for a 30-year lifetime depends slightly on the level of the initial strength. The difference can be seen in Table 18.3.

Abrasions and melted-in particles are the most common flaws on the fiber surface. Proof testing of these flaws is usually performed under ambient conditions and so the fatigue parameters obtained at 50% RH and 23°C can be directly used for prediction of the inert strength after proof testing and, importantly, the minimum surviving strength. The minimum surviving strength can be close to the proof test stress, σ_p, or be significantly less than that level depending on whether unloading is "fast" or "slow" [6,11]. The requirement for the case of "fast" unloading is simply [17]:

$$t_u < t_c = \frac{(n_2 - 2) \cdot B_2}{\sigma_p^2}, \tag{18.29}$$

where t_u is the time of unloading. Using the parameters obtained for abraded and contaminated fibers and $\sigma_p = 0.7$ GPa, one obtains $t_c = 5.8 \times 10^{-3}$ s. The same parameter for indented fiber is 0.51×10^{-3} s. These values are more achievable than the 0.58×10^{-5} s calculated in the case of single-region model and fatigue parameters obtained for strong fiber. Based on the results and analyses of this study the actual minimum surviving strength after proof testing is closer to the proof stress level than what one would predict using the single region power law and data obtained from high strength fiber. In other words, if the time of unloading is shorter than this value, the minimum surviving strength of the fiber after proof testing can be close to the proof stress (90% or more) [21]. Thus, Equation (18.27) and data from Table 18.3 can be used directly in some cases for lifetime estimations.

18.5. CONCLUSION

Fiber lifetime predictions in communication networks can be made using the safe stress approach for mechanical reliability. It incorporates multi-region behavior of the crack

growth rate and is confirmed by the results of high-speed tensile testing of weak fibers. The proof stress can be used in the minimum strength design if the unloading time is shorter than 0.5×10^{-3} s.

The safe stress approach to the fiber lifetime based on a two-region model of crack growth requires additional information about crack growth. To measure the fatigue parameters for all regions, the usual testing machines for dynamic and static fatigue measurements are insufficient. Special equipment for high-speed testing is required. Fatigue measurements are best performed on fiber with a narrow strength distribution near the proof stress level. Defects on these fibers should be similar to the defects responsible for the fiber lifetime in a communication line. With weak fiber the requirements on high-speed testing machines are somewhat minimized, because all the regions are observed at significantly lower testing speeds.

Fatigue parameters obtained for weak fibers at 50% RH and 23°C can be directly used for prediction of the inert strength after proof testing and, importantly, the minimum surviving strength (the first step of lifetime prediction). For the second step of lifetime estimation, information about fatigue behavior of relatively large flaws at very slow loading speed or during long-term static tests is required. We demonstrated that fatigue behavior of weak fibers is sensitive to experimental conditions. Thus, further detailed experiments are required to solve the following problems:

1. To make an estimate of the fiber lifetime in a communication line, the fatigue curve parameters in the region of low crack growth rates (the slowest testing speeds) should be studied for environmental conditions typical of communication cables. There is no agreement concerning the quantitative influence of the environmental factors (humidity, pH, temperature, etc.) on the fatigue parameters. Applicability of the power law or the exponential law to this region remains an open question.
2. Another open question is the fatigue behavior of defects of different size and nature in the region of low crack growth rates. Can the fatigue results of an indented fiber be applied to a fiber with melted-in particles or to an abraded fiber? Can the results for a high-strength fiber be correlated with a weak fiber?
3. The fiber polymer coating can effect the local environment at the fiber surface. The influence of the coating on fiber fatigue has been studied extensively and coatings that affect fatigue tests have been commercialized. The effect of coating behavior on fatigue at elevated stressing rates was not addressed in this study and there is still much work to be done in the area of coatings and their effect on long-term fatigue behavior.

REFERENCES

1. R.J. Charles, Static fatigue of glass, J. Applied Physics, 29(11), pp. 1549–1560 (1958).
2. S. Sakaguchi and T. Kimura, Influence of temperature and humidity on dynamic fatigue of optical fibers, J. Amer. Ceram. Soc., 64(5), pp. 259–262 (1981).
3. W.J. Duncan, P.W. France, and S.P. Craig, The effect of environment on the strength of optical fiber, in C.R. Kurkjian, Ed., Strength of inorganic glass, Plenum Press, New York, 1985, pp. 309–326.
4. M.J. Matthewson and C.R. Kurkjian, Environmental effects on the static fatigue of silica optical fiber, J. Amer. Ceram. Soc., 71(3), pp. 177–183 (1988).
5. TIA/EIA-455-31-C (FOTP-31, Proof Testing Optical Fiber by Tension.

6. A.G. Evans and S.M. Wiederhorn, Proof testing of ceramic materials—an analytical basis for failure prediction, Int. J. Fracture, 10, pp. 379–392 (1974).
7. S.M. Wiederhorn, Influence of water vapor on crack propagation in soda-lime glass, J. Amer. Ceram. Soc., 50(8), pp. 407–414 (1967).
8. P. Chanticul, B.R. Lawn, H. Richter, and S.W. Freiman, Relation between multiregion crack growth and dynamic fatigue of glass using indentation flaws, J. Amer. Ceram. Soc., 66(7), pp. 515–518 (1983).
9. A.G. Evans, Slow crack growth in brittle materials under dynamic loading conditions, Inter. J. of Fracture, 10(2), pp. 251–259 (1974).
10. T. Svensson, Evaluation of B-values of telecom fibers, objectives and methods, Mater. Research Soc. Symp. Proc., 531, pp. 47–52 (1998).
11. S.L. Semjonov and M.M. Bubnov, On the concept of multiregion crack growth, Mater. Research Soc. Symp. Proc., 531, pp. 243–248 (1998).
12. P.T. Garvey, T.A. Hanson, M.G. Estep, and G.S. Glaesemann, Mechanical reliability predictions: an attempt at measuring the initial strength of draw abraded optical fiber using high stress rates, 46th Int. Wire & Cable Symp. Proc., pp. 883–887 (1997).
13. S.L. Semjonov, G.S. Glaesemann, C.R. Kurkjian, and M.M. Bubnov, Modeling of proof test level flaws using cube corner indents, 47th Int. Wire & Cable Symp. Proc., 1998, pp. 928–932.
14. S.L. Semjonov, G.S. Glaesemann, and M.M. Bubnov, Fatigue behavior of silica fibers of different strength, Proc. SPIE, 3848, pp. 102–107 (1999).
15. S.L. Semjonov, G.S. Glaesemann, D.A. Clark, and M.M. Bubnov, Fatigue behavior of silica fibers with different defects, Proc. SPIE, 4215, pp. 28–35 (2000).
16. S.L. Semjonov, G.S. Glaesemann, D.A. Clark, and M.M. Bubnov, Effect of environmental conditions on fatigue of weak silica-clad optical fibers, Proc. SPIE, 5465, pp. 61–67 (2004).
17. Y. Mitsunaga, Y. Katsuyama, H. Kobayashi, and Y. Ishida, Failure prediction for long length optical fiber based on proof testing, J. Applied Phys., 53(7), pp. 4847–4853 (1982).
18. G.S. Glaesemann, Optical fiber failure probability predictions from long-length strength distributions, 40th Int. Wire & Cable Symp. Proc., 1991, pp. 819–825.
19. A. Paul and G.S. Glaesemann, An appraisal of mechanical reliability predictions for optical fibers based on break rates, 46th Int. Wire & Cable Symp. Proc., 1997, pp. 896–901.
20. M.M. Bubnov and S.L. Semjonov, B-value and optical fiber lifetime, Mat. Res. Soc. Symp. Proc., 531, pp. 231–241 (1998).
21. T.A. Hanson and G.S. Glaesemann, Incorporation multi-region crack growth into mechanical reliability predictions for optical fiber, J. Materials Science, 32, pp. 5305–5311 (1997).
22. H.C. Chandan, R.C. Bradt, and G.E. Rindone, Dynamic fatigue of float glass, J. Amer. Ceram. Soc., 61(5-6), pp. 207–210 (1978).
23. R.E. Mould and R.D. Southwick, Strength and static fatigue of abraded glass under controlled ambient conditions: II, J. Amer. Ceram. Soc., 42, pp. 582–592 (1959).
24. S.M. Wiederhorn and L.H. Bolz, Stress corrosion and static fatigue of glass, J. Amer. Ceram. Soc., 53, pp. 543–548 (1970).
25. J.E. Ritter, Jr. and C.L. Sherburne, Dynamic and static fatigue of silicate glasses, J. Amer. Ceram. Soc., 54, pp. 601–605 (1971).
26. J. Gomg, Y. Chen, and C. Li, Statistical analysis of fracture toughness of soda-lime glass determined by indentation, J. Non-Cryst. Solids, 279, pp. 219–223 (2001).
27. G.S. Glaesemann, K. Jakus, and J.E. Ritter, Jr., Strength variability of indented soda-lime glass, J. Amer. Ceram. Soc., 70(6), pp. 441–444 (1987).
28. S.L. Semjonov and C.R Kurkjian, Strength of silica optical fibers with micron size flaws, J. Non-Cryst. Solids, 283, pp. 220–224 (2001).
29. B.R. Lawn, and R. Wilshaw, Indentation fracture: principles and applications, J. Mater. Sci., 10, pp. 1049–1081 (1975).
30. K. Jakus, J.E. Ritter, and Jr., S.R. Choi, and T.J. Lardner, Failure of fused silica fibers with subthreshold flaws, J. Non-Cryst. Solids, 102, pp. 82–87 (1988).
31. S. Lathabai, J. Rodel, T. Dabbs, and B.R. Lawn, Fracture mechanics model for subthreshold indentation flaws, J. Mater. Sci., 26, pp. 2157–2168 (1991).
32. B. Lin and M.J. Matthewson, Inert strength of sub-threshold and post-threshold Vickers indentations on fused silica fibres, Phil. Mag. A, 74(5), pp. 1235–1244 (1996).
33. G.M. Pharr, D.S. Harding, and W.C. Oliver, Measurement of fracture toughness in thin films and small volumes using nanoindentation methods, in M.A. Nastasi, D.M Parkin, and H. Gleiter, Eds., Mechanical Proper-

ties and Deformation Behavior of Materials Having Ultrafine Microstructures, Kluwer Academic Publishers, The Netherlands, 1993, pp. 449–461.
34. G.S. Glaesemann, D.A. Clark, and J.J. Price, An indentation method for creating reproducible proof-stress level flaws in commercial optical fiber, Proc. SPIE, 4639, pp. 21–29 (2002).
35. C.R. Kurkjian, D. Biswas, and H.H. Yuce, Intrinsic Strength of Lightguide Fibers, Proc. SPIE, 2611, pp. 56–63 (1995).
36. B.A. Proctor, I. Whitney, and J.W. Johnson, The strength of fused silica, Proc. Roy. Soc. 297A, p. 534 (1967).
37. S.R. Choi, J.E. Ritter, Jr., and K. Jackus, Failure of glass with subthreshold flaws, J. Amer. Ceram. Soc., 73(2), pp. 268–274 (1990).
38. D.H. Roach and A.R. Cooper, Effect on contact residual stress relaxation on fracture strength of indented soda-lime glass, J. Amer. Ceram. Soc., 68(11), pp. 632–636 (1985).
39. T.P. Dabbs and B.R. Lawn, Strength and fatigue properties of optical glass fibers containing microindentation flaws, J. Amer. Ceram. Soc., 68(11), pp. 563–569 (1985).
40. D.J. Wissuchek, Effect of refractory particles on the strength of optical fibers, Mat. Res. Soc. Symp. Proc., 531, pp. 187–192 (1998).
41. S.M. Wiederhorn, Fracture surface energy of glass, J. Amer. Ceram. Soc., 65(8), pp. 99–105 (1969).
42. G.S. Glaesemann and J.D. Helfinstine, Measuring the inert strength of large flaws in optical fiber, Proc. SPIE, 2074, pp. 95–107 (1993).
43. A.T. Taylor and M.J. Matthewson, Effect of pH on the strength and fatigue of fused silica optical fiber, 47th Int. Wire & Cable Symp. Proc., 1998, pp. 874–880.
44. S.L. Semjonov, M.M. Bubnov, and O.V. Khleskova, Susceptibility of static fatigue parameters of optical fibers to environmental conditions, Proc. SPIE, 2611, pp. 49–54 (1995).
45. S.M. Wiederhorn and H. Johnson, Effect of electrolyte pH on crack propagation in glass, J. Amer. Ceram. Soc., 56(4), pp. 192–197 (1973).
46. G.S. Glaesemann, The mechanical behavior of large flaws in optical fiber and their role in reliability predictions, 41st Int. Wire & Cable Symp. Proc., 1992, pp. 698–704.
47. G.S. Glaesemann, Assessing the long-term reliability of optical fiber, Proc. National Fiber Optic Engineers Conference, 1994, p. 297.
48. G.S. Glaesemann and S.T. Gulati, Dynamic fatigue data for fatigue resistant fiber in tension vs. bending, Proc. Optical Fiber Communication Conference, Technical Digest Series, Vol. 5, 1989, p. 48.
49. M.J. Matthewson and C.R. Kurkjian, Environmental effects on the static fatigue of silica optical fiber, J. Amer. Ceram. Soc., 71(3), pp. 177–183 (1988).

APPENDIX 18.A. HIGH SPEED AXIAL STRENGTH TESTING: MEASUREMENT LIMITS

Equipment for axial strength testing consists of a load sensor and a type of mandrel device for wrapping and gripping the fiber at both ends. This is shown schematically in Figure 18A.1 where M is the mass of mandrel with connection to the load sensor; $F(t)$ is a pulling force (the actual force experienced by the fiber). For obvious reasons, it is important that the force experienced by the fiber be detected by the sensor. There is concern that at ultra-fast stressing rates the act of moving the mass of the mandrel limits one's ability to accomplish this. As a first approach, a linear dependence of loading force on time is used:

$$F(t) = \begin{cases} F' \cdot t, & \text{if } t \le t_0 \\ 0, & \text{if } t > t_0, \end{cases} \tag{18A.1}$$

where F' is a loading rate, t_0 time to the fiber failure. Using Newton's laws one can write

$$F(t) - k \cdot x(t) - \mu \frac{dx(t)}{dt} = M \frac{d^2x(t)}{dt^2}, \tag{18A.2}$$

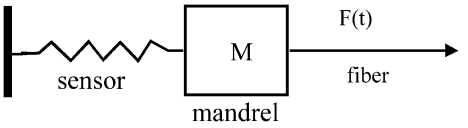


FIGURE 18A.1. Scheme of sensitive element.

where k is a joint elasticity coefficient of the load cell sensor; μ—the damping coefficient. We assume that both displacement and $F_{\text{meas}} = kx(t)$ are measured.

The solution of Equation (18A.2) is

$$x(t) = \frac{F' \cdot t}{k}\left[1 - \frac{2 \cdot \delta}{1+\delta^2} \cdot \frac{1}{\omega \cdot t} + \frac{\exp(-\delta \cdot \omega \cdot t)}{\omega \cdot t}\left(\frac{2 \cdot \delta}{1+\delta^2} \cdot \cos(\omega \cdot t) + \frac{\delta^2 - 1}{1+\delta^2}\sin(\omega \cdot t)\right)\right] \quad \text{for } t < t_0 \tag{18A.3}$$

and

$$x(t - t_0) = x_0 \exp[-\delta \cdot \omega \cdot (t - t_0)] \times \left\{\left(\frac{v_0}{\omega \cdot x_0} + \delta\right) \cdot \sin[\omega \cdot (t - t_0)] + \cos[\omega \cdot (t - t_0)]\right\} \quad \text{for } t > t_0, \tag{18A.4}$$

where

$$\omega = \sqrt{\frac{k}{M} - \frac{\mu^2}{4 \cdot M^2}}, \tag{18A.5}$$

$$\omega_0 = \sqrt{\frac{k}{M}}, \tag{18A.6}$$

$$\delta = \frac{\sqrt{\omega_0^2 - \omega^2}}{\omega} = \frac{\mu}{2 \cdot M \cdot \omega}, \tag{18A.7}$$

$$x_0 = x(t_0), \tag{18A.8}$$

$$v_0 = x'(t_0). \tag{18A.9}$$

We can find maximum value of the displacement measured

$$x_{\max} = x_0 \cdot \exp(-\delta \cdot \varphi) \cdot \left[\left(\frac{v_0}{\omega \cdot x_0} + \delta\right) \cdot \sin(\varphi) + \cos(\varphi)\right], \tag{18A.10}$$

where

$$\varphi = \operatorname{arctg}\left(\frac{v_0}{x_0 \cdot \omega + \delta \cdot v_0 + \delta^2}\right). \tag{18A.11}$$

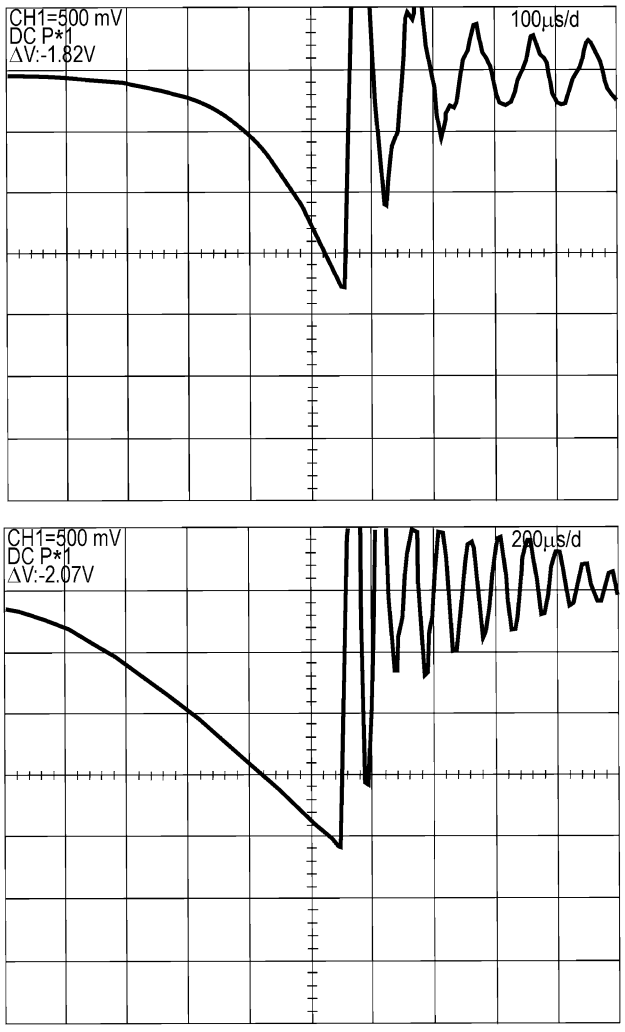


FIGURE 18A.2. Signal from a piezoelectric load cell.

Parameters ω and δ can be calculated from the load trace of fiber failure. As an example, Figure 18A.2 presents a typical signal from the load cell for different loading speeds. The parameters $\omega \sim 6 \times 10^4$ s^{-1} and $\delta \sim 0.05$ are determined from these plots. Figure 18A.3 demonstrates a mathematical simulation of the load cell behavior using these parameters.

To visualize the discrepancy between the real and measured strength, the dependence of $F_{\max}/F_0$ on ωt_0 was plotted in Figure 18A.4. It demonstrates that for a long time-to-failure, slow loading rate, i.e., $t_0 \gg (20/\omega) \sim 300$ μs, the result registered by the load sensor is practically equal to the actual fiber breaking force. But, if the rate of loading increases and respectively time to failure decreases to $t_0 \leq 20/\omega$, the sensor data can be incorrect. In this case an increase or decrease of the fiber strength with increasing of loading rate can be observed.

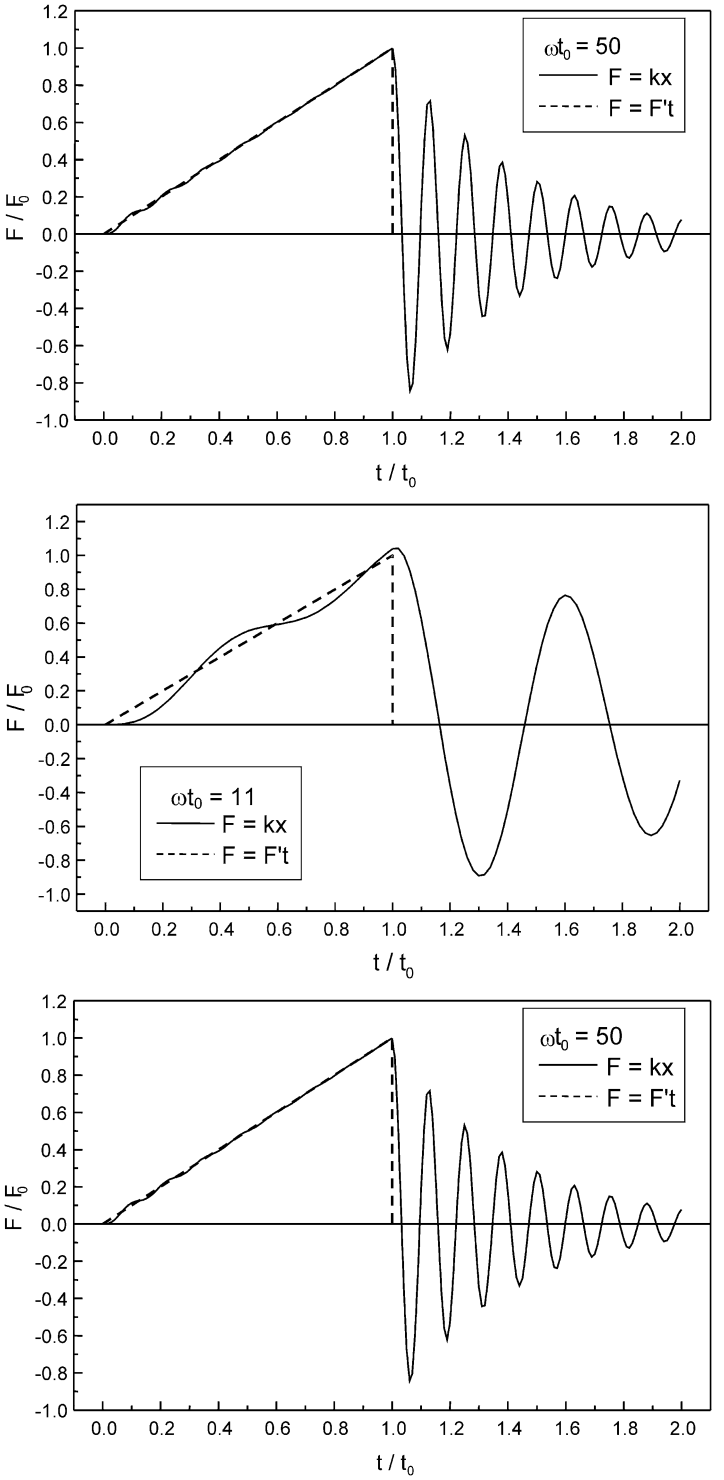


FIGURE 18A.3. Results of mathematical simulation.

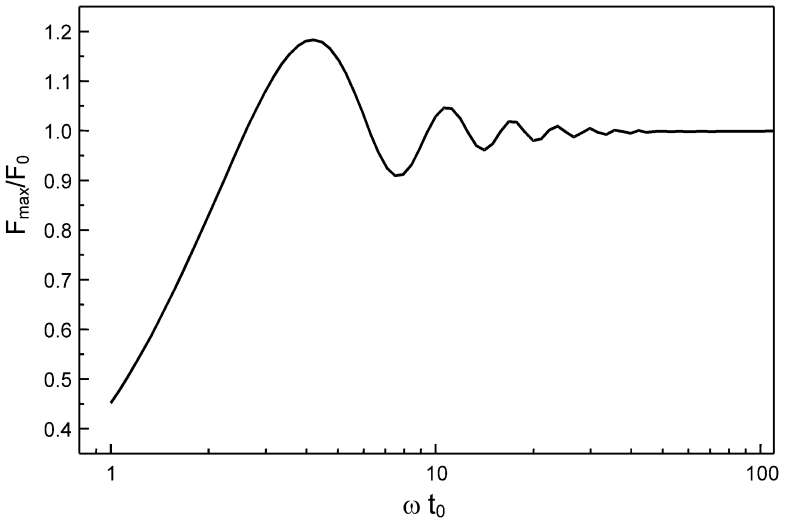


FIGURE 18A.4. Discrepancy between real and measured strength.

APPENDIX 18.B. INCORPORATING STATIC FATIGUE RESULTS INTO DYNAMIC FATIGUE CURVES

When a sample of fiber is loaded under constant stress in a static fatigue test or the ever-increasing stress of a dynamic test, its strength degrades from its initial value. The crack responsible for strength grows slowly at first and accelerates throughout the test. The end result is a final velocity that is orders of magnitude faster than at the beginning of the test. A single power-law model enables us to evaluate this behavior using simple mathematics.

18.B.1. Static Fatigue Test

In this test, a constant load σ_s is applied to a specimen, and the time to failure t_s is measured. The strength begins at some initial value, S_i and degrades to some lower value S after time t according to,

$$B(S_i^{n-2} - S^{n-2}) = \sigma_s^n t. \tag{18B.1}$$

The sample breaks when the final strength $S = S_f$ equals the applied static stress, $S = S_f = \sigma_s$. Thus, the time to failure $t = t_s$ is

$$t_s = \frac{B}{\sigma_s^2}\left[\left(\frac{S_i}{\sigma_s}\right)^{n-2} - 1\right]. \tag{18B.2}$$

For our further needs, Equation (18B.1) can be transformed into the following:

$$\frac{S_f}{S_i} = \left(1 - \frac{t}{t_{0s}}\right), \tag{18B.3}$$

where

$$t_{0s} = \frac{B}{\sigma_s^2}\left(\frac{S_i}{\sigma_s}\right)^{n-2}, \tag{18B.4}$$

is a time very close to the time-to-failure, t_s in Equation (18B.2).

Using Equations (18.3) and (18B.3), the following relation describing crack growth during the test can be obtained:

$$a_f = a_i\left(1 - \frac{t}{t_{0s}}\right)^{-\frac{2}{n-2}}. \tag{18B.5}$$

Thus, the crack growth rate is

$$V = \frac{da_f}{dt} = V_s\left(1 - \frac{t}{t_{0s}}\right)^{-\frac{n}{n-2}}, \tag{18B.6}$$

where

$$V_s = \frac{a_i}{t_{0s}}\left(\frac{2}{n-2}\right). \tag{18B.7}$$

18.B.2. Dynamic Fatigue Test

In this case, a linearly increasing stress ($\sigma = \sigma' t$) is applied to a specimen, and the stress at failure, σ_d, is measured. The following dependence between strength, S, and test time, t, is

$$B\left(S_i^{n-2} - S^{n-2}\right) = \frac{(\sigma')^n t^{n+1}}{n+1}. \tag{18B.8}$$

The sample breaks when strength $S = S_f$ becomes equal to applied stress σ_d ($S_f = \sigma_d = \sigma' t_d$). Thus, time to failure, t_d is

$$t_d = \left\{\frac{(n+1)BS_i^{n-2}}{(\sigma')^n}\left[1 - \left(\frac{\sigma_d}{S_i}\right)^{n-2}\right]\right\}^{\frac{1}{n+1}}. \tag{18B.9}$$

For our further needs, Equation (18B.8) can be transformed into the following:

$$\frac{S_f}{S_i} = \left[1 - \left(\frac{t}{t_{0d}}\right)^{n+1}\right]^{\frac{1}{n-2}}, \tag{18B.10}$$

where

$$t_{0d} = \left(\frac{(n+1)BS_i^{n-2}}{(\sigma')^n}\right)^{\frac{1}{n+1}}, \tag{18B.11}$$

is a time that is very close to time to failure t_{d} [Equation (18B.9)].

Using Equations (18.3) and (18B.10), the following relation describing crack growth during dynamic test can be obtained:

$$a_f = a_i \left[1 - \left(\frac{t}{t_{0d}} \right)^{n+1} \right]^{-\frac{2}{n-2}}. \tag{18B.12}$$

Thus, the crack growth rate is

$$V = \frac{da_f}{dt} = V_d \left[1 - \left(\frac{t}{t_{0d}} \right)^{n+1} \right]^{-\frac{n}{n-2}} \left(\frac{t}{t_{0d}} \right)^{n}, \tag{18B.13}$$

where

$$V_d = \frac{a_i (n+1)}{t_{0d}} \left(\frac{2}{n-2} \right). \tag{18B.14}$$

For the case of a dynamic fatigue testing, it is also useful to consider the dependence of crack growth rate on strength:

$$V = V_d \left[1 - \left(\frac{S}{S_i} \right)^{n-2} \right]^{\frac{n}{n+1}} \left(\frac{S_f}{S_i} \right)^{-n}. \tag{18B.15}$$

18.B.3. Discussion

Strength degradation and crack velocity curves for the static fatigue test are presented in Figures 18B.1 and 18B.2. In Figure 18B.2 it can be seen that the crack starts to grow at a rate V_s and for most of the test time it has a velocity close to V_s. The crack growth

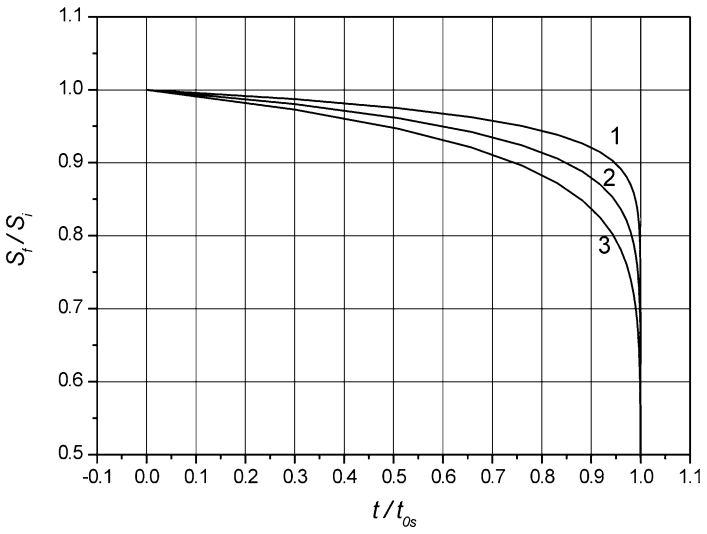


FIGURE 18B.1. Dependencies of inert strength on testing time during static fatigue test for different n-values. (1) $n = 30$; (2) $n = 20$; (3) $n = 15$.

rate changes from V_s to $10V_s$ during a period of testing time from 0 to (0.86–0.88)t_{0s} when the parameter n is in the typical 15 to 30 range. If we suppose that for $V > 10V_s$ the crack growth rate has a different dependence on the stress intensity factor, i.e., different n-value as shown in Figure 18B.3, the resulting static time-to-failure must be different. Nevertheless, the predicted change in time-to-failure due to a change in n from 15 to 30 in this region is less than 2%. For comparison, 1% variation in initial strength corresponds to ∼15–30% variation of time-to-failure for n values ranging from 15 to 30.

Thus, crack growth behavior in the range of V_s to $10V_s$ is what dictates the time to failure.

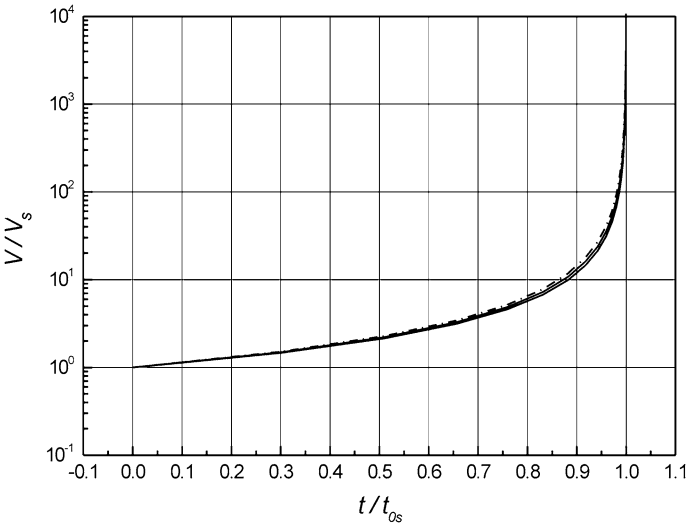


FIGURE 18B.2. Dependencies of the crack growth rate on testing time during static fatigue test for different n-values (15–30).

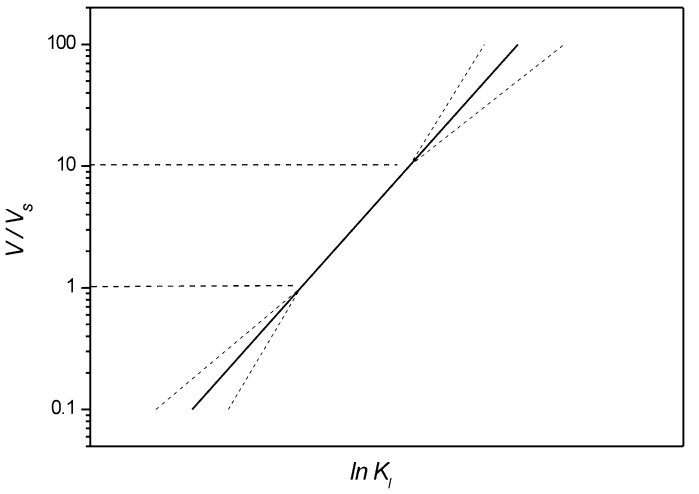


FIGURE 18B.3. Multi-region behavior for crack growth rate.

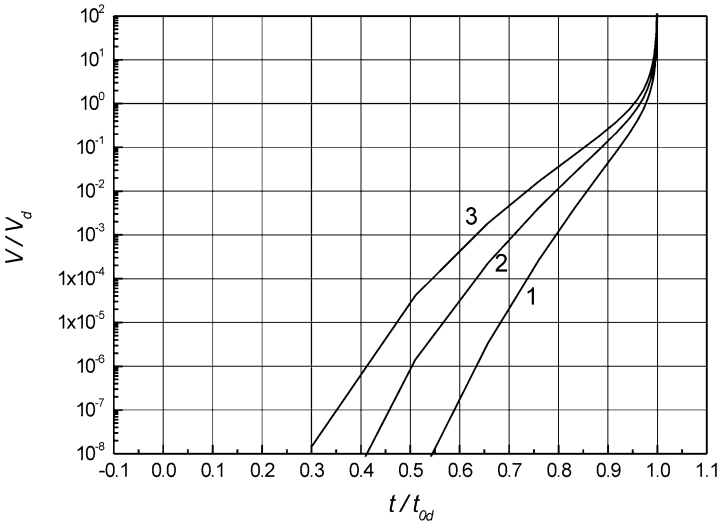


FIGURE 18B.4. Dependencies of crack growth rate on testing time during dynamic fatigue test for different n-values. (1) $n = 30$; (2) $n = 20$; (3) $n = 15$.

Figure 18B.4 describes the behavior of crack velocity as a function of time during the dynamic test. As expected, the crack growth rate increases by many orders of magnitude during the test. Nevertheless, during most of the testing time, the strength remains close to the initial strength (Figure 18B.5) because the crack growth rate is too small to noticeably affect the size of the crack. Only when the testing time approaches t_d does the strength begin to degrade significantly. It can be calculated from Equation (18B.10) that the sample strength reduces only by 1% for $t = 0.86$ ($n = 15$) or $t = 0.96$ ($n = 30$). These values correspond to the crack growth rates of (0.3 to 0.8)V_d. On the other hand, once the test time has reached 0.99 t_{0d} the strength will degrade only 1% more before failure. This corresponds to the crack growth rates $V > (2.7\text{–}6.0)V_{0d}$ for $n = 30\text{–}15$.

Thus, most of the strength degradation in dynamic fatigue testing occurs during the velocity range of $0.3V_s$ to $6.0V_s$.

Based on the discussion above, the condition for equivalent dynamic and static tests is $V_s = V_d$. Using (18B.7) and (18B.13), and supposing that the fiber samples have the same initial crack sizes a_i, dynamic fatigue data can be translated into static fatigue data by using the following;

$$\begin{cases} t_{0s} = \dfrac{t_d}{n+1} = \dfrac{\sigma_d}{\sigma'(n+1)} \\ \sigma_s = \sigma_d, \end{cases} \tag{18B.16}$$

where n is determined from stressing rates ranging from $0.3\sigma'$ to $6\sigma'$. The form of Equation (18B.16) is not new, but the point being made here is that one can translate dynamic fatigue data into static fatigue data without assuming an underlying crack growth model provided one stays within a certain range of stressing rates. Thus, with knowledge of the local "n" one can map dynamic fatigue data into static fatigue data for a wide range of stressing rates by splitting the dynamic fatigue curve into a series of segments where each one is independently mapped.

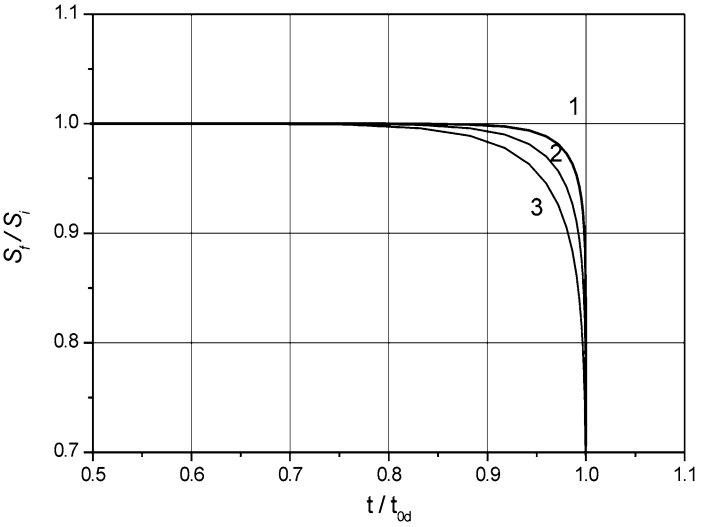


FIGURE 18B.5. Dependencies of inert strength on testing time during dynamic fatigue test for different n-values. (1) $n = 30$; (2) $n = 20$; (3) $n = 15$.

Equation (18B.16) can be reversed for the translation of dynamic fatigue data into static fatigue data;

$$\begin{cases} \sigma' = \dfrac{\sigma_s}{t_s(n+1)} \\ \sigma_d = \sigma_s, \end{cases} \tag{18B.17}$$

where n can be calculated from experimental data in a region of times-to-failure from t_s to $10t_s$.

Equations (18B.16) and (18B.17) are well known for the case of the single-region power law model of crack growth. This work has demonstrated that similar expressions work for the case of an arbitrary crack growth model, provided there are no dramatic changes of the crack growth behavior in the vicinity of V_s or V_d. With the exception of two cases, it is nearly impossible to observe any deviation from the power law in one-order of magnitude ranges of loading speed or time-to-failure.

Sharp changes in n have been observed at very fast loading rates in experiments described here, however, in this case there is really no need to translate dynamic fatigue data into static fatigue data. The other case where sharp changes in n have been observed is the static fatigue "knee" [49]. The fatigue "knee" results from the formation of small surface pits on pristine fiber exposed to hot water and differs from the fatigue phenomenon being addressed in this study.

Thus, Equation (18B.17) can be used to incorporate static fatigue results into a dynamic fatigue curve. It allows one to significantly extend the range of experimental fatigue data for further analysis.

19

The Effect of Temperature on the Microstructure Nonlinear Dynamics Behavior

Xiaoling He

University of Wisconsin, Milwaukee, USA

Abstract

The dynamics of microstructures, such as micro-electro-mechanical systems (MEMS) or thin laminated printed wiring boards (PWB) raises reliability concerns when they are subjected to mechanical loads, as well as thermal fields induced by the electric circuits on board. The microstructure dynamics behavior needs to be described in the framework of nonlinear dynamics, due to the fact that its deformation is in the same order of its critical dimension. In this chapter, studies on MEMS nonlinear dynamic characteristics—in particular, on the response of a thin laminated microstructure under both dynamic load and thermal field are presented. Equations of motion of a thin laminated MEMS structure are obtained in the form of a decoupled Duffing's equation, for the out-of-plane deformation of an isotropic thin laminate in a simply supported boundary condition. A generalized Galerkin's method is employed for the reduction of the governing equation of motion. The microstructure behaviors are studied in nonlinear resonance, bifurcation and chaos with respect to different load and temperature variation. Stress field is also obtained in a nonlinear relation with the deflection and the thermal field. Failure induced by these stresses is evaluated based on the Composite Failure Criteria for a laminate.

The significance of the temperature variation on the deformation and stress field of a PWB is demonstrated, including steady state thermal effect on the resonance behavior, and the increased deflection due to chaos induced by the transient thermal field, as well as the tendency towards failure at an elevated temperature. A stronger influence of the in-plane thermal field over that of the transverse thermal field is observed from the analytical model and numerical computation. The analytical formulation enables assessment of deflection behavior of a laminated microstructure, or a PWB, subject to the imposed thermal and mechanical fields.

19.1. INTRODUCTION

Microstructure electronic and mechanical devices (MEMS) have been developed for a wide range of applications. A MEMS structure or a component is usually processed on a silicon wafer, and then assembled with other components to make a device with a specific

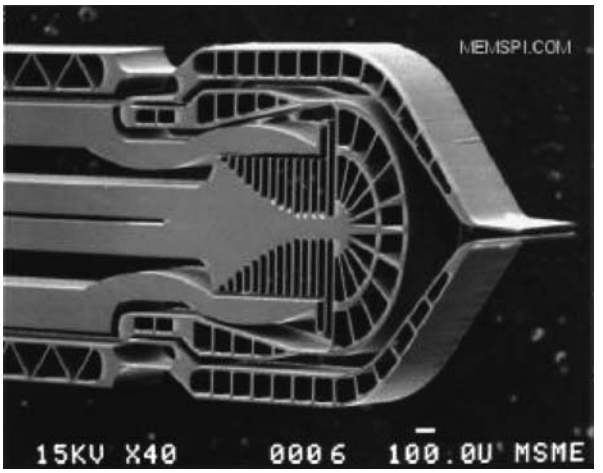


FIGURE 19.1. A MEMS single crystal silicon tweezers.

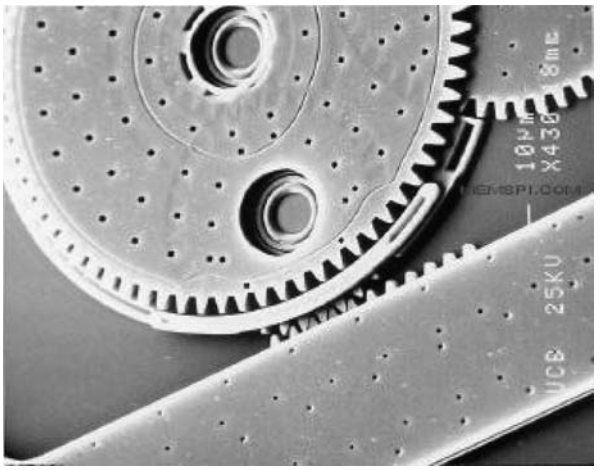


FIGURE 19.2. A MEMS gear and motor system.

functionality. Figure 19.1 shown is a pair of MEMS tweezers made of single crystal silicon. Figure 19.2 is a micro gear and motor system. A unique characteristic of a MEMS device is that it contains kinetic elements in micrometer scale, and its motion is executed in the same order of magnitude. The difference in linear and nonlinear deformation is not negligible once the deflection is over half of a thin laminate thickness such that a linear strain field can not give an accurate account of the deflection and stresses. This invokes nonlinear dynamics to describe MEMS dynamic behavior in order to control its motion.

The capability for higher precision motions makes MEMS the choice for various functional mechanisms, such as sensors and actuators for different devices. Examples of MEMS applications are tremendous, such as micro-gyroscopes for motion control and micro-satellite for position controls [1,14–16,18,31,37]; motors and gears; flow control

capacitors, fluid mixing and flow control actuators [2–6,26,35]; acoustic control actuators [10,21,50]; mass and gas sensors, radio frequency sensors; as well as optical guide devices [20,36,51].

MEMS devices are also seen in a computer hard drive, with motors and actuators for disk spinning, data writing and reading control [55]. MEMS can also be packaged in a paradigm of "system-on-a-chip" for electronic component packaging. "Lab on a chip" MEMS were developed for bio-sensors, by enabling electrochemical immunoassay-based chemical/biological detection. In such systems, functional elements are composed of chemical flow control, detection and sensor devices [11]. These types of bio-MEMS embedded into the human body can detect and monitor the functionality of human organs, and may provide self-compensation of certain functions, which would otherwise be impossible to achieve. The highly sensitive and miniaturized MEMS sensing technology, with high spatiotemporal resolution, is a blessing to medical diagnostics. E.g., it can reveal the important interplay between blood circulation and vascular cell behavior, so as to link biomechanical forces on the micro-scale with the large-scale physiology of the human system [28,38,45,51]. Powered by the nano-device and sensing technology, Nano bio-MEMS have been used for detection and healing of human diseases. In aerospace and weaponry development, miniaturized robots and sensors are advancing into more sophisticated functional mechanisms, such as invisible, unmanned mini-planes and micro robotics for reconnaissance missions. These systems contain not only sensors and actuators but also microstructures such as the printed wiring boards for device and functional controls. MEMS resonance characteristics have also been utilized in neuron-computing development [27].

Other than the nonlinear motion of the kinetic elements of MEMS, the deformation of a PWB in a thin laminated structure also falls into the category of nonlinear dynamics. Essentially, PWBs are present in all kinds of integrated microelectronic devices, from computer boards to any electronics using microelectronic circuits. In a wide range of applications, dynamic forces induced from its own motion, or imposed by the environment are exerted onto the circuit board. For example, in aerodynamics or aeronautics applications, the high acceleration alone experienced during launching can exert immense pressure load over the board to cause PWB nonlinear vibration. The high stress field generated from the large deflection can cause a breakdown of the circuits, therefore, the failure of the device.

Other than the nonlinear deformation behaviors of MEMS, a unique characteristics of a typical thin laminated MEMS structure used in micro-electronics packaging or MEMS device is that it contains both conduction lamina and insulation lamina. The embedded electrical circuits make the in-plane temperature variation non-uniform. The difference in the interfacial structures and materials causes temperature variation to be non-uniform over the laminate plane, as well as through its thickness. The combined thermal mechanical effects induce a stress field that affects the laminate deformation behavior. Other coupled field interactions also exist for MEMS devices, such as coupled thermal electric fields [32,34,56], fluid solid interactions [29,42], and thermal and fluid interactions [43]. Therefore, studies on the nonlinear response of MEMS have provided another platform for nonlinear dynamics analyses based on the classical nonlinear dynamics theory integrated with coupled field equations. The classical theories, which have matured from investigation of aerodynamic structures that experience large deformation [17,53], has been applied to MEMS dynamics analysis in numerical computation of the coupled field equations.

In analytical development, the nonlinear dynamic behavior of a PWB under a constant acceleration has been studied based on the classical plate theory [48], in which PWB's

deflection motion is formulated in the form of a Duffing equation. Suhir also investigated the PWB response with moisture and thermal field effect [49]. These analyses are based on the classical plate theory with nonlinear Von-Karman strain field in a single mode analysis. MEMS crack behavior [52], and response of a micro cantilever beam in deflection and torsion motion [33,54] have also been studied. In addition, simulations in finite element analysis have been applied to a computer hard-drive [44,46,47,55].

For coupled thermal field and structure dynamics in the context of nonlinear thermoelasticity for a thin isotropic laminate, the nonlinear multimode analysis of the laminate dynamics often led to a coupled modal form equation of motion, without approximations such as that due to Berger [9]. However, by using a generalized Galerkin's method, or the method of weighted residuals, we have found that the nonlinear governing equation of motion can be reduced to a decoupled modal form equation of motion for the laminate deflection [22]. In this chapter, the theoretical background for this development is described, and formulation obtained is applied for analysis of a thin isotropic laminate used as a PWB.

19.2. THEORETICAL DEVELOPMENT

19.2.1. Background on Nonlinear Dynamics and Nonlinear Thermo-Elasticity Theories

Nonlinear dynamics theories of a thin laminate subject to mechanical loading have been developed by Whitney and Leissa [53], and applied to the study of various laminate dynamic responses [17].

The thermal effect on the nonlinear dynamics of an orthotropic laminate, with only the stationary temperature variation, was analyzed by means of Berger's approximation [8,12,13,39–41], in order to reduce the equation of motion into a decoupled modal form nonlinear ordinary differential equation [9]. Berger's approximation neglects the second strain invariant of the middle surface in deriving the equation of motion in energy variation principles [9], which compromise the total energy of the system. However, this approximation enables solutions of many nonlinear dynamics problems. Both rectangular and circular plates have been studied for the nonlinear thermally induced steady state deformation, as well as its dynamic response [8,13,40,41]. A Duffing type equation has been obtained for the transient deflection of plate due to the in-plane and transverse thermal field, and a cubic function for the thermally induced static buckling deformation with respect to temperature.

Without Berger's approximation, a decoupled equation of motion can hardly be obtained directly from reduction of the governing equation of motion, instead, a coupled form of nonlinear ordinary equations were obtained for a thin laminate buckling and vibration, when both in-plane and transverse load are effective [17]. Alternatively, the finite element formulation is developed without modal analysis for the nonlinear thermal vibration of the laminate with temperature variation [7], which effectively eliminated the nonlinear modal coupling concern.

Taking both thermal and mechanical load into account, we studied the laminate nonlinear dynamics by using a generalized Galerkin's method, or the method of weighted residuals, without Berger's approximation [22]. The governing equation of motion of the laminate deflection was reduced to a decoupled modal form Duffing type equation for laminate in different boundary conditions by using this method. This decoupled modal form allows for modal response characterization of a laminate with respect to boundary conditions, initial conditions and loading effect [22,23,25]. Impact response of a thin laminate, i.e.,

PWB, is also studied, which yielded an analytical solution to the Duffing equation for its free response [24]. The general response of the Duffing equation subject to an arbitrary load needs numerical computation to describe its behavior, which ranges from quasi-periodic to chaotic response.

In this chapter, we demonstrate this method for analysis of an isotropic laminate in a simply supported boundary condition subject to the mechanical load, and an arbitrary thermal field with both the in-plane and transverse temperature variation. The decoupled form equation of motion enables modal analysis of both deflection and stress field behaviors for a laminated microstructure, such as a PWB, for its failure prediction.

19.2.2. Nonlinear Thermo-Elasticity Development for an Isotropic Laminate Subject to Thermal and Mechanical and Load

The equation of motion in classical laminate theory, when the in-plane inertia is neglected, takes the form of [53]:

$$\begin{aligned}
&\frac{\partial N_x}{\partial x} + \frac{\partial N_{xy}}{\partial y} = 0,\\
&\frac{\partial N_{xy}}{\partial x} + \frac{\partial N_y}{\partial y} = 0,\\
&\frac{\partial^2 M_x}{\partial x^2} + \frac{\partial^2 M_{xy}}{\partial x \partial y} + \frac{\partial^2 M_y}{\partial y^2} + \mathrm{N}(w) + q = I\frac{\partial^2 w}{\partial t^2} - I_2\nabla^2\left(\frac{\partial^2 w}{\partial t^2}\right),
\end{aligned} \tag{19.1}$$

where:

$$\mathrm{N}(w) = N_x\frac{\partial^2 w}{\partial x^2} + 2N_{xy}\frac{\partial^2 w}{\partial x \partial y} + N_y\frac{\partial^2 w}{\partial y^2}. \tag{19.2}$$

Forces N_x, N_y, N_{xy} are elements of an in-plane force vector. M_x, M_y, M_{xy} are the elements of the moment vector. The inertia of the laminate is defined as:

$$[I, I_1, I_2] = \sum_{k=1}^{N}\int_{-h_1}^{h_2}\rho_{ij}^{(k)}[1,\ z,\ z^2]dz. \tag{19.3a}$$

For homogeneous material in a symmetric laminate,

$$[I, I_1, I_2] = [I, 0, I_2]. \tag{19.3b}$$

The strain components in Equation (19.4), defined by the Von-Karman strain field, are:

$$\{\varepsilon\} = \{\varepsilon^0\} + z\{\varepsilon^1\}, \tag{19.4}$$

$$\{\varepsilon^0\} = \begin{Bmatrix} \varepsilon_{xx}^0 \\ \varepsilon_{yy}^0 \\ \gamma_{xy}^0 \end{Bmatrix}, \quad \{\varepsilon^1\} = \begin{Bmatrix} \kappa_x \\ \kappa_y \\ \kappa_{xy} \end{Bmatrix}, \tag{19.4a}$$

$$\varepsilon_{xx}^0 = \frac{\partial u_0}{\partial x} + \frac{1}{2}\left(\frac{\partial w}{\partial x}\right)^2, \quad \varepsilon_{yy}^0 = \frac{\partial v_0}{\partial y} + \frac{1}{2}\left(\frac{\partial w}{\partial y}\right)^2,$$

$$\gamma_{xy}^0 = \frac{\partial u_0}{\partial x} + \frac{\partial v_0}{\partial y} + \left(\frac{\partial w}{\partial x}\right)\left(\frac{\partial w}{\partial y}\right),$$

$$\kappa_x = -\frac{\partial^2 w}{\partial x^2}, \quad \kappa_y = -\frac{\partial^2 w}{\partial y^2}, \quad \kappa_{xy} = -2\frac{\partial^2 w}{\partial x \partial y}. \tag{19.4b}$$

This satisfies the compatibility equation:

$$\frac{\partial^2 \varepsilon_{xx}^0}{\partial y^2} + \frac{\partial^2 \varepsilon_{xx}^0}{\partial x^2} - \frac{\partial^2 \gamma_{xx}^0}{\partial x \partial y} = \left(\frac{\partial^2 w}{\partial x \partial y}\right)^2 - \frac{\partial^2 w}{\partial x^2}\frac{\partial^2 w}{\partial y^2}. \tag{19.5}$$

For a symmetric isotropic laminate, the equilibrium equation for the transverse deformation in Equation (19.1) can be expressed in terms of forces in the form of:

$$D_{11}\nabla^4 w + I\ddot{w} - I_2\nabla^2\ddot{w} = Q - \nabla^2 M^T + \mathrm{N}(w), \tag{19.6}$$

where:

$$\mathrm{N}(w) = L(F, w) + N_x^T \frac{\partial^2 w}{\partial x^2} + N_y^T \frac{\partial^2 w}{\partial y^2}, \tag{19.7a}$$

$$L(F, w) = \frac{\partial^2 F}{\partial x^2}\frac{\partial^2 w}{\partial y^2} + \frac{\partial^2 F}{\partial y^2}\frac{\partial^2 w}{\partial x^2} - 2\frac{\partial^2 F}{\partial x \partial y}\frac{\partial^2 w}{\partial x \partial y}, \tag{19.7b}$$

$$N^T = \begin{Bmatrix} N_x^T \\ N_y^T \\ 0 \end{Bmatrix} = \eta T_1(x, y) \begin{Bmatrix} 1 \\ 1 \\ 0 \end{Bmatrix},$$

$$\eta = \sum_{k=1}^{N} \alpha^{(k)} (D_{11} + D_{12})^{(k)} = \sum_{k=1}^{N} (z_{k+1} - z_k) \frac{\alpha^{(k)} E^k}{(1 - v^k)}, \tag{19.7c}$$

$$M^T = \begin{Bmatrix} M_x^T \\ M_y^T \\ 0 \end{Bmatrix} = \xi T_1(x, y) \begin{Bmatrix} 1 \\ 1 \\ 0 \end{Bmatrix},$$

$$\xi = \sum_{k=1}^{N} \alpha^{(k)} (D_{11} + D_{12})^{(k)} = \sum_{k=1}^{N} \frac{1}{3}(z_{k+1}^3 - z_k^3) \frac{\alpha^{(k)} E^k}{(1 - v^k)}. \tag{19.7d}$$

Denote E as the residual of the equation of motion, Equation (19.8), with the approximating function $w(x, y, t)$,

$$E = D_{11}\nabla^4 w + I\ddot{w} - I_2\nabla^2\ddot{w} - Q + \nabla^2 M^T - \mathrm{N}(w). \tag{19.8}$$

Using Galerkin's method,

$$\int_0^b \int_0^a E * w(x, y)dxdy = 0. \tag{19.9}$$

Then a Duffing equation, with a cubic term for the spring force, can be obtained for the out-of-plane deformation for the response with a steady-state thermal field as:

$$\ddot{W}_{mn} + \omega_{mn}^2 W_{mn} + r_{mn} W_{mn}^3 = \hat{Q}_{mn}, \tag{19.10}$$

$$\omega_{mn}^2 = \frac{1}{\tilde{I}}\left[D_{11}\left(\alpha_m^2 + \beta_n^2\right)^2 + \frac{\eta}{4} T_{mn}^0 \left(\alpha_m^2 + \beta_n^2\right)\right], \tag{19.10a}$$

$$r_{mn} = \frac{1}{16\tilde{I}}\left[4A_{12}\alpha_m^2\beta_n^2 + \frac{3A_{11}^2 - A_{12}^2}{A_{11}}\left(\alpha_m^4 + \beta_n^4\right)\right], \tag{19.10b}$$

$$q_{mn} = \frac{Q_{mn} + \xi(\beta_n^2 + \alpha_m^2)T_{mn}^1}{\tilde{I}} = q_{mn}^0 + q_{mn}^T, \tag{19.10c}$$

$$q_{mn}^0 = \frac{Q_{mn}}{\tilde{I}}, \tag{19.10d}$$

$$q_{mn}^T = \frac{\xi(\beta_n^2 + \alpha_m^2)T_{mn}^1}{\tilde{I}}. \tag{19.10e}$$

The above formulation indicates that the influence of the in-plane temperature variation affects the natural frequency ω_{mn}, and the through thickness temperature rise decreases the load q_{mn}. The stiffness of the Duffing equation $r_{mn} > 0$, determined by the fact that the laminate stiffness $A_{11} > A_{12}$, $A_{ij} > 0$. Temperature variation does not affect r_{mn}, however, the in-plane thermal field modifies the frequency of the Duffing equation. Therefore, the laminate deflection behavior can be drastically different from each other with a steady-state or a transient state thermal field. In the following analysis, both the steady-state and transient thermal field are studied for their influence on the laminate response behavior.

19.3. THIN LAMINATE DEFLECTION RESPONSE SUBJECT TO THERMAL EFFECT AND MECHANICAL LOAD

19.3.1. Steady State Temperature Effect

19.3.1.1. Thermal Field Assumption The thermal field function needs to satisfy the orthogonality conditions in Galerkin's integral evaluation. For a steady-state temperature distribution, $T(x, y, z)$is assumed in a linear function as:

$$T(x, y, z) = T_0(x, y) + z * T_1(x, y). \tag{19.11}$$

The non-uniform temperature field is assumed as the following:

$$T_0 = \sum_{n=1}^{\infty}\sum_{m=1}^{\infty} T_{mn}^c + \sum_{n=1}^{\infty}\sum_{m=1}^{\infty} T_{mn}^0 \cos(2\alpha_m x)\cos(2\beta_n y), \tag{19.12a}$$



FIGURE 19.3. PWB laminate.

$$T_1 = \sum_{n=1}^{\infty}\sum_{m=1}^{\infty} T_{mn}^1 \sin(\alpha_m x)\sin(\beta_n y). \tag{19.12b}$$

The in-plane temperature variation $T_0(x, y)$ satisfies the boundary condition of zero heat flux, i.e.,

$$\begin{aligned} &T_0(x, y),_x = 0, \quad \text{at } x = 0,\ x = a, \\ &T_0(x, y),_y = 0, \quad \text{at } y = 0,\ y = b. \end{aligned} \tag{19.13}$$

The constant series of $T_c(x, y) = \sum_{n=1}^{\infty}\sum_{m=1}^{\infty} T_{mn}^c$ in Equation (19.12a) is assumed for $T_0(x, y)$ in order to simulate the monotonic variation of the temperature field. The coefficients of the Fourier series are chosen as:

$$\begin{aligned} &T_{11}^0 = 95.42, \quad T_{12}^0 = 24.89, \quad T_{21}^0 = 23.86, \quad T_{11}^c = 35, \\ &T_{22}^0 = 6.22, \quad T_{13}^0 = 11.15, \quad T_{31}^0 = 10.60 \quad (^\circ\text{C}). \end{aligned} \tag{19.14}$$

It is obvious that the assumed T_{mn}^0 coefficient for the first mode has a more dominant influence on the temperature distribution than all the other terms.

An exemplary thin laminated PWB, composed of both conduction Cu layers and insulation FR-4 epoxy layers, is shown in Figure 19.3. The material properties of the PWB are given in Table 19.1. With the assumed thermal field, the natural frequency of the PWB is increased from $\omega_{11} = 1708.9$ Hz without thermal effect, to $\omega_{11}^T = 3589.6$ Hz with thermal effect. An additional load is introduced to the system due to the temperature variation through the laminate thickness. However, this load is negligible: $q_{11}^T = 0.051$ N/cm^2 with the maximum temperature rise $T_1(x, y) = 5^\circ$C over the thickness of 1.54 mm; $q_{11}^T = 0.051$ N/cm^2 is about 1/10th of the load that induces deflection of concern, as demonstrated from the stress analysis in the later section. Therefore, the transverse temperature effect can be ignored. The limited temperature rise chosen for $T_1(x, y)$ reflects the fact that temperature variation is restrained by the intertwined insulation and conduction layers of this laminate.

19.3.1.2. Response Behavior with a Steady State Thermal Field In the following analyses, the numerical computation of the Duffing equation in a Runge-Kutta method is employed for the deflection of a PWB subject to both thermal field and mechanical load in a harmonic excitation. With modal parameters for the laminate given in Table 19.1 for the given PWB, it is found that the deflection without thermal effect for mode_12 is 1/8th of that of mode_11, and the mode _22 has a deflection 20th of that of mode_11 [22]. Since the steady state temperature rise does not change the form of the Duffing equation, other than amplifying the magnitude of the natural frequency and the load, similar response behaviors

TABLE 19.1.
PWB material properties (130 × 180 × 1.54 mm).

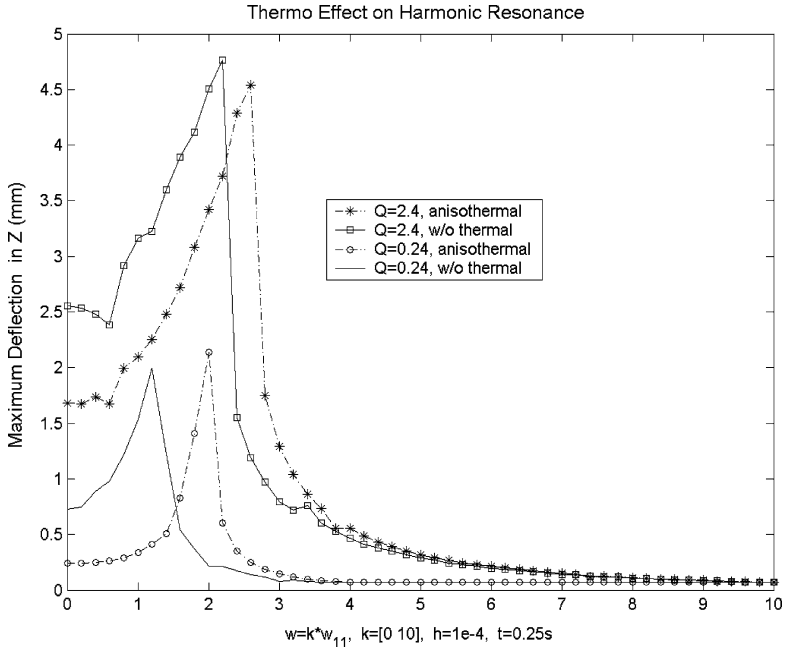
Layer	Thickness (mm)	Material	Specific weight (g/cm^3)	Young's module (MPa)	CTE (ppm/°C)	Poisson ratio
1	0.03	Cu/FR-4	1.911	2.98E+04	1.61E+01	0.211
2	0.2	FR-4	1.200	2.00E+04	1.60E+01	0.190
3	0.035	Cu	8.310	1.18E+05	1.72E+01	0.400
4	0.5	FR-4	1.200	2.00E+04	1.60E+01	0.190

can be expected to that subject to a mechanical load, without thermal source. Therefore, only the fundamental mode is considered.

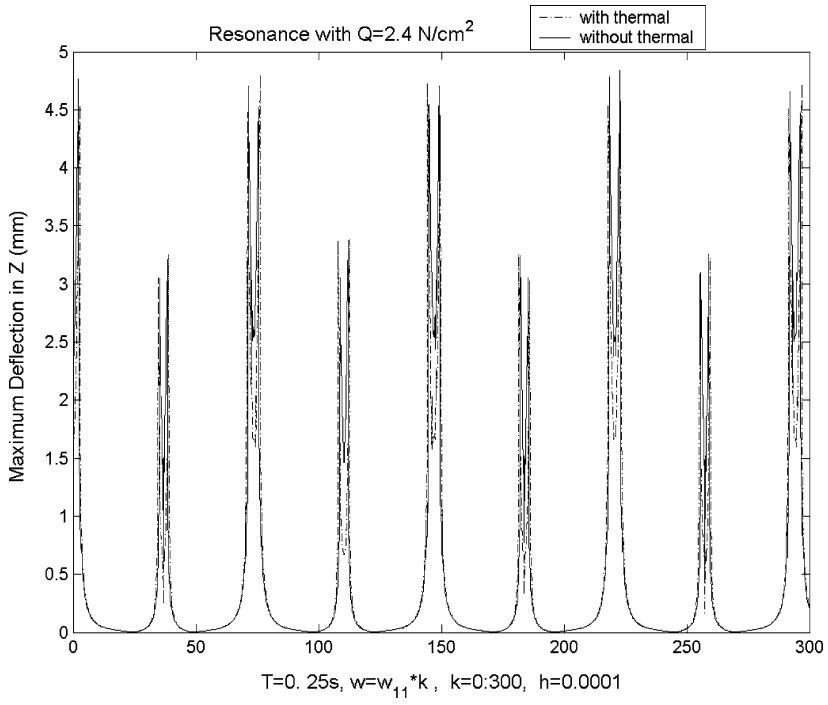
An external load effect, with the assumed thermal field, is studied for $Q = 0.24$ N/cm^2 and $Q = 2.4$ N/cm^2, respectively. The load $Q = 0.24$ N/cm^2 corresponds to the pressure load generated with $100g$ acceleration, where g is the gravitational acceleration, $g = 9.8$ m/s^2. Frequency responses, with and without the thermal source, are shown in Figure 19.4(a) for the excitation frequencies $\omega = k\omega_{11}$, $k = [0, 10]$. It is observed that the effect of the in-plane temperature rise is to up-shift the resonance frequency, while the resonance magnitude is not much affected. Under the load of $Q = 0.24$ N/cm^2, resonance frequency is shifted from $k = 1$ without the thermal source, to $k = 2$ with the thermal source. Under $Q = 2.4$ N/cm^2, the k value at resonance is shifted from $k = 2.2$ to $k = 2.6$ due to this thermal source. Higher loading leads to a higher deflection and higher resonance frequencies, within the excitation frequency range $k = [0\ 10]$. Subsequent to resonance, each response subsides and diminishes. The deflections at $k = 0$ represent the maximum deflection under a constant load, and it is reduced by the steady-state thermal effect, as observed from the figure.

With the extended excitation frequencies, resonance in the frequency domain presents quasi-periodic behavior. Figure 19.4(b) shows recurrence of nonlinear resonances, with and without a thermal source, when the excitation frequency range is in the range $k = [0\ 300]$. Recurrent resonances present alternating magnitude in resonance, which means that both primary and secondary resonances occur, and each in pairs. In addition, each pair of resonances occurs almost in symmetry with respect to a center frequency. The center is observed at about $k = 36$, 72, 108, 144, 180, 216, 252, 288, etc., which suggests that the symmetry center is in a quasi period-doubling bifurcation. This quasi-periodic resonance behavior is characteristic of the Duffing equation, with or without thermal source.

The corresponding quasi-periodic behavior in the temporal domain can be observed from a Poincare map. The Poincare map is formed by periodically taken the phase portrait trace, and the accumulation of these traces can reveal the response characteristics in periodic, quasi-periodic or chaotic behaviors. Figure 19.5(a) shows a modified Poincare map at $Q = 0.72$ N/cm^2, where $k = 2$ with thermal source. A period of $T = 175$ μs is observed for the temporal oscillation. The interval at which the traces of the points are generated in the map is taken at 1/5th of the period, i.e., 35 μs. A total of 5000 points is accumulated in the map. The corresponding map without thermal source is shown in Figure 19.5(b), with a quasi-periodic oscillation of $T = 3365$ μs. This Poincare map is also generated at a 1/5 interval of the period, i.e., each point at each 673 μs interval, for 5000 points. The five individual traces in each map may merge as more points are accumulated. Compared to the standard Poincare map, with each point in each period to form an island around one deflection-velocity locus, the above map with multiple intermediate points in each period

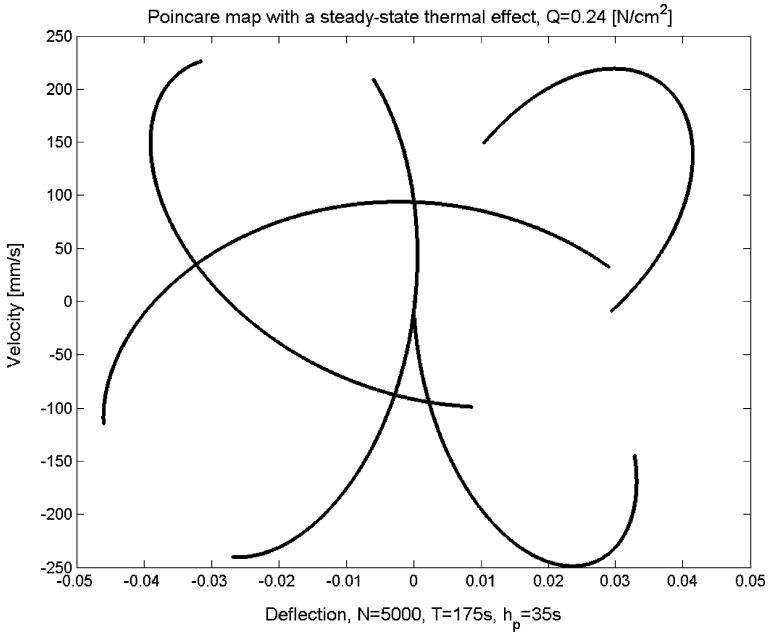


(a)

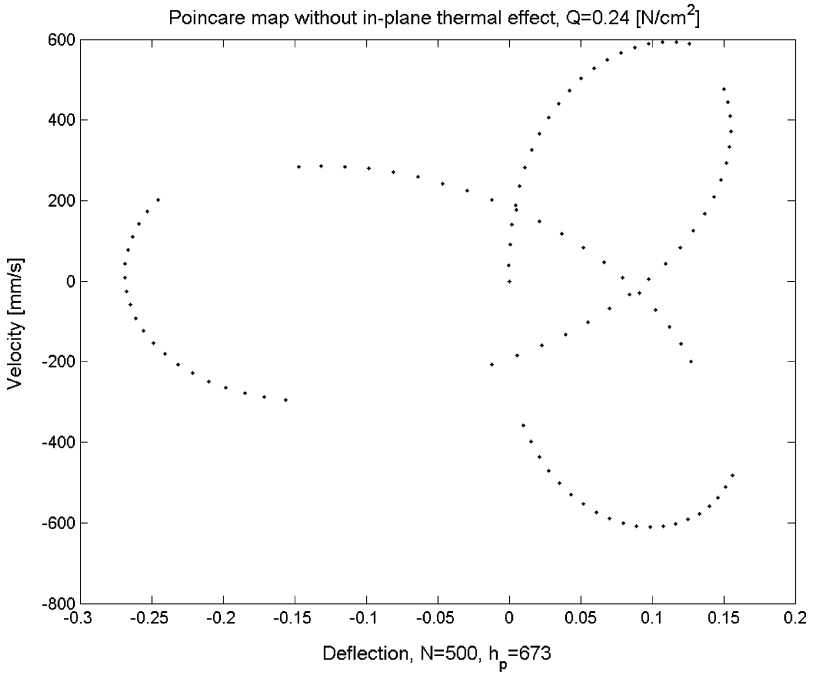


(b)

FIGURE 19.4. Thermal effect on resonance in frequency domain response. (a) Resonance for $k = [0\ 10]$, $Q = 0.24$, 2.4 N/cm^2. (b) Resonance for $k = [0\ 100]$, $Q = 2.4$ N/cm^2.



(a)



(b)

FIGURE 19.5. Poincare map for the quasi-periodic oscillation with and without thermal effect with $Q = 0.24$ N/cm^2, $k = 2$. (a) $N = 5000$ points, $T = 175$ μs. (b) $N = 500$ points, $T = 673$ μs.

provides a broader view of the deflection and velocity range in the quasi-periodic oscillation.

These results demonstrate the effect of the steady-state in-plane thermal field on the resonance frequency variation, as well as its effect on the quasi-periodic behaviors from the mechanically induced response or the thermal-mechanically induced response, although different oscillation pattern is observed with different load.

19.3.2. Transient Thermal Field Effect

19.3.2.1. Thermal Field Assumption The non-uniform transient thermal field $T(x, y, z, t)$ is assumed as:

$$T(x, y, z, t) = T_0(x, y) * f(t) + z * T_1(x, y) * g(t), \tag{19.15}$$

where its arbitrary transient variation can be specified in Fourier series as:

$$f(t) = \sum_{j=1}^{J} f_j \cos(vt), \quad v = j * \omega_{mn}, \quad j = 1, 2\ldots,$$

$$g(t) = \sum_{k=1}^{K} g_k \cos(\omega t), \quad \omega = k * \omega_{mn}, \quad k = 1, 2\ldots. \tag{19.15a}$$

Then a Duffing equation can be obtained in the form:

$$\ddot{W}_{mn} + \left[\omega_{mn}^2 + \chi_{mn} \sum_{v=j\omega_{mn}, j=1}^{j=N} f_j \cos(vt)\right] W_{mn} + r_{mn} W_{mn}^3 = q_{mn}, \tag{19.16}$$

$$\chi_{mn} = \frac{(\alpha_m^2 + \beta_n^2)\eta T_{mn}^0}{4\tilde{I}}, \quad \eta = \sum_{k=1}^{N} \frac{E_k \alpha^{(k)}}{1 - v_k}(z_{k+1} - z_k), \tag{19.16a}$$

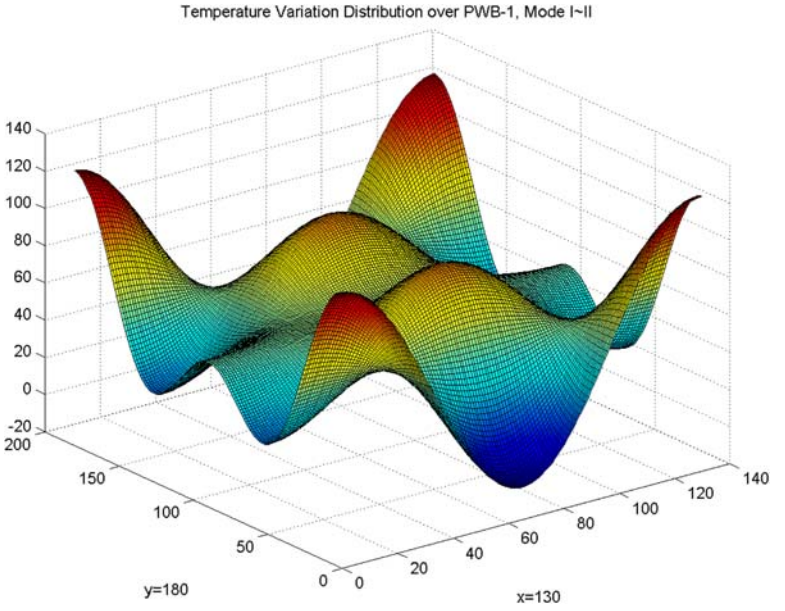
$$r_{mn} = \frac{1}{16\tilde{I}}\left[4A_{12}\alpha_m^2\beta_n^2 + \frac{3A_{11}^2 - A_{12}^2}{A_{11}}(\alpha_m^4 + \beta_n^4)\right], \tag{19.16b}$$

$$q_{mn} = q_{mn}^0 + q_{mn}^T, \tag{19.16c}$$

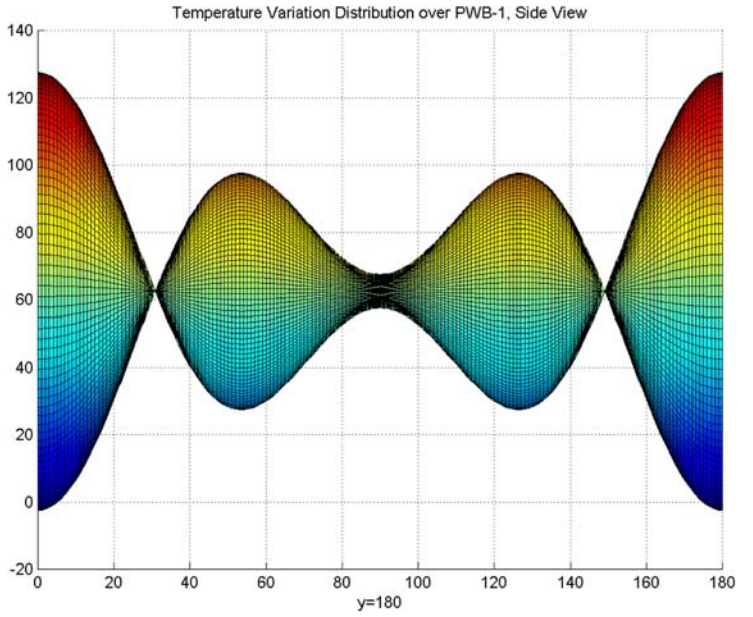
$$q_{mn}^0 = \frac{Q_{mn}}{\tilde{I}}, \quad q_{mn}^T = \frac{(\alpha_m^2 + \beta_n^2)T_{mn}^1 \sum_{k=1,2\ldots}^{\infty} g_k \cos(\omega t)}{\tilde{I}}. \tag{19.16d}$$

The time-dependent thermal field makes a system response in an oscillatory natural frequency, with an inherent pressure load q_{mn}^T due to the transverse thermal field. The deformation can be alternately viewed as being induced by both the in-plane and transverse thermal loads when the equation is re-arranged as:

$$\ddot{W}_{mn} + \omega_{mn}^2 W_{mn} + r_{mn} W_{mn}^3 = q_{mn} - \chi_{mn} W_{mn} \sum_{v=j\omega_{mn}, j=1}^{j=N} \cos(vt). \tag{19.17}$$



(a)



(b)

FIGURE 19.6. Thermal field $T_0(x, y)$. (a) First 2 terms of $T_0(x, y)$. (b) Side-view of $T_0(x, y)$.

Equation (19.17) indicates that the transverse thermal field serves as an excitation force, while the in-plane thermal field is coupled with the deflection acting as a circular forcing term for deflection.

An arbitrary thermal field $T_0(x, y)$ is shown in Figure 19.6(a) and Figure 19.6(b). The in-plane temperature variation has the following coefficients for $T_0(x, y)$:

TABLE 19.2.
System parameters for modal analysis.

Mode	ω_{mn} (s^{-1})	r_{mn} ($(mm{\cdot}s)^{-2}$)	χ_{mn} (s^{-2})	$[\omega^T_{mn}]_{\max}$	T^1_{mn} (°C)	q^T_{mn} (N/cm^2)
Mode-I $m=1, n=1$	1708.0	1,551,840	5,167,903.5	2844.0	16.2	2.52
Mode-II $m=1, n=2$	3466.0	6,490,828.3	8,983,273.3	4582.2	8.10	2.76

$$T^0_{11}=70,\quad T^c_{11}=70,\quad T^0_{12}=60,\quad T^c_{12}=55,$$
$$T^0_{21}=12,\quad T^c_{21}=7.5,\quad T^0_{22}=0.3,\quad T^c_{22}=7.5\quad (^\circ\mathrm{C}). \tag{19.18}$$

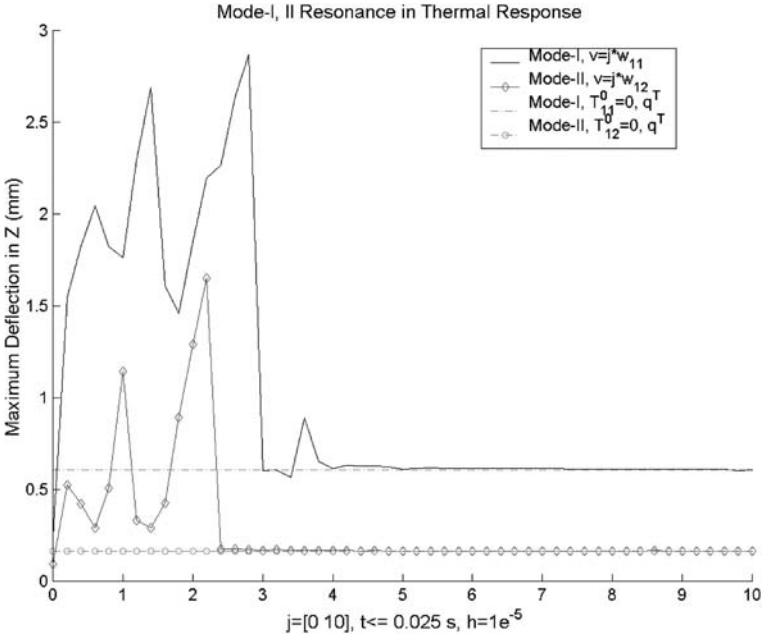
It is assumed that the first 2 modes of T^0_{mn} dominate the thermal field. The maximum in-plane temperature rise from the two-mode superposition is about 100°C at the central area. A sharp temperature rise occurs at the corners due to the assumed thermal field function in Equation (19.12a) which satisfies the Neumann boundary condition. Figure 19.6(b) indicates that the maximum value at these corner points is about 125°C, 25% higher than the maximum temperature assumed. A temperature rise of 10°C through the laminate thickness introduces a thermal load of $q^T_{11}=0.234\ \mathrm{N/cm^2}$ and $q^T_{12}=0.237\ \mathrm{N/cm^2}$. Modal parameters in Equation (19.16) for Mode-I and Mode-II with a thermal field defined by Equation (19.18) are shown in Table 19.2.

As the system's natural frequency is defined by $\omega^T_{mn}=\sqrt{\omega^2_{mn}+\chi_{mn}\cos(vt)}$ with the transient thermal effect, the relative magnitude of ω^2_{mn} and χ_{mn} determines the system behavior. For the fundamental mode with the specified thermal field, Table 19.2 indicates that $\omega^2_{11}=2{,}920{,}297.67\ \mathrm{s^{-2}}$, $\chi_{11}=5{,}167{,}903.5\ \mathrm{s^{-2}}$, i.e., $\omega^2_{11}<\chi_{11}$, the natural frequency varies between $[\sqrt{\omega^2_{mn}-\chi_{mn}},\sqrt{\omega^2_{mn}+\chi_{mn}}]$, and experiences an unstable condition when $\omega^2_{11}<|\chi_{11}\cos(vt)|$, i.e., $\omega^T_{mn}=\sqrt{\omega^2_{mn}\pm\chi_{mn}\cos(vt)}$ is imaginary. Since the stable and unstable condition interchanges in each cyclic, the response is confined within a certain range as of a stable oscillation, rather than reaching infinity as in a completely unstable system. For the second mode, $\omega^2_{12}=12{,}013{,}156.0\ \mathrm{s^{-2}}$, $\chi_{12}=8{,}983{,}273.3\ \mathrm{s^{-2}}$, $\omega^2_{12}>\chi_{12}$, $\omega_{12}>0$, the response is stable. From Equation (19.16), it can be found that the critical temperature rise for a stable oscillation is $T_{mn}=[4D_{11}(\alpha^2_m+\beta^2_n)]/\varsigma$, i.e., $T_{11}=39.5$°C and $T_{12}=80.2$°C, respectively.

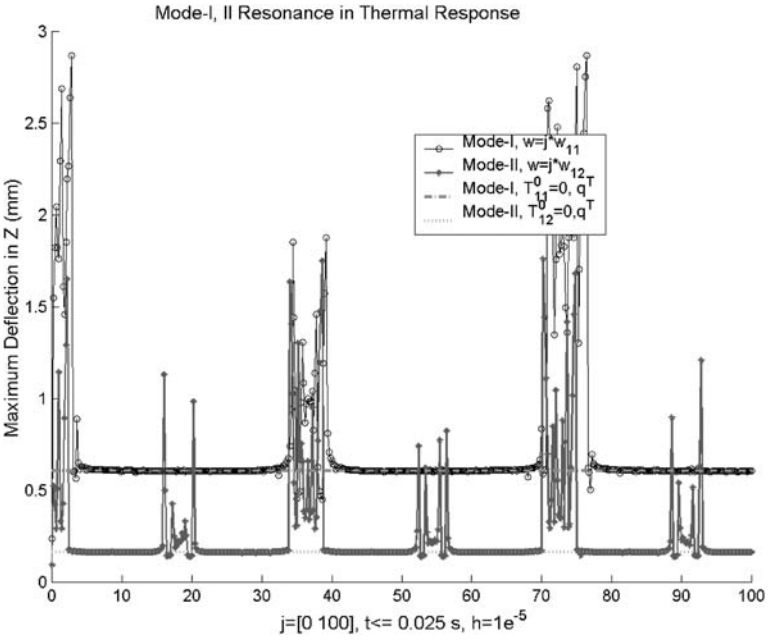
19.3.2.2. Response Characterization with Transient Thermal Field Without external excitation, the harmonic thermal sources associated with the in-plane and transverse thermal fields can make the system resonant, as indicated by Equation (19.17), which is alternately called "the thermally induced free response [40]," since the laminate is subjected to the heating source only, without the mechanical load.

- Transient in-plane thermal field effect

  The in-plane thermal field effect on the modal response is examined by assuming the transverse thermal field constant, i.e., $q^T_{11}=0.234\ \mathrm{N/cm^2}$ and $q^T_{12}=0.237\ \mathrm{N/cm^2}$, respectively. To analyze the influence of thermal field transient frequency on the response behavior, a single term in-plane temperature variation in the form of $T_0(x,y,t)=T^0_{mn}(x,y)\cos(vt)$, $v=j*\omega_{mn}$ is studied in frequency domain analysis with j variation for two modal responses with modal natural frequencies



(a)



(b)

FIGURE 19.7. Mode-I-II thermal frequency response, (a) $j = [0\ 10]$, (b) $j = [0\ 100]$.

ω_{mn} as $\omega_{11} = 1780$ Hz, $\omega_{12} = 3466$ Hz. As the modal frequency $\omega_{12} \approx 2\omega_{11}$, the thermal field frequency $v = j * \omega_{mn}$ for the second mode is almost twice that of the first mode for the same index j. Figure 19.7(a) shows the frequency responses in the

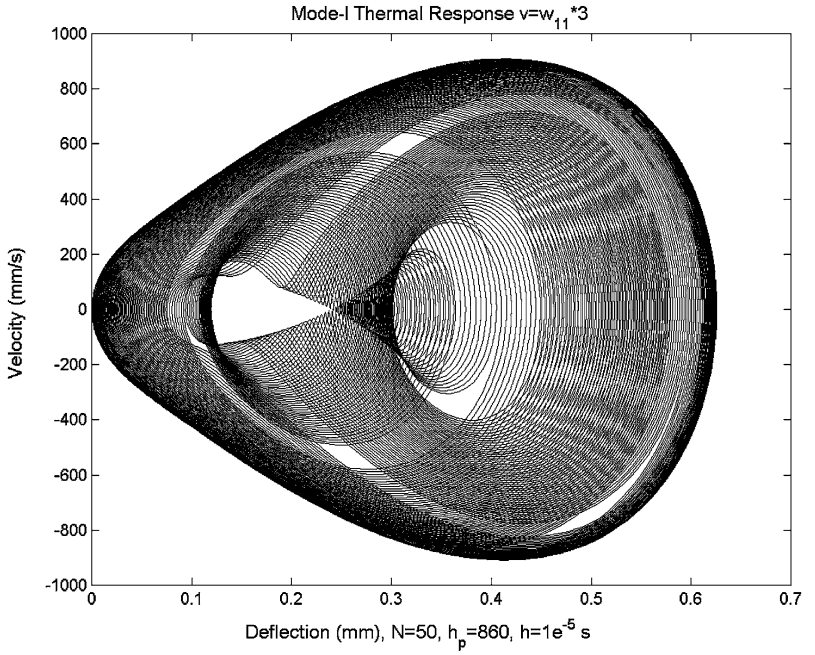
range of $j = [0, 10]$, with and without in-plane thermal sources for both modes. It is observed that resonances due to thermal field oscillation occur in the range $j = [0, 3]$ for both modes. With the thermal load $q_{11}^T \approx q_{12}^T$, the deflection range in Mode-I is about twice the laminate thickness, and twice the Mode-II response. In an extended frequency response in Figure 19.7(b) for $j = [0, 100]$, resonance recurrence is observed with overlaps for both Mode-I and Mode-II, in a quasi period-doubling bifurcation. A quasi-symmetry center for each recurrence of resonance pair is observed as in the case of a steady-state thermal effect, with an external harmonic forcing. The centers for the resonance bifurcation are at $j = 37$ and 74 for Mode-I, and at $j = 18.5$, 37, 55.5, 74, and 92 for Mode-II. Considering $\omega_{12} \approx 2\omega_{11}$, resonance frequencies for both modes are, in fact, close to each other even with different values of j. Comparatively, the response is more irregular and chaotic with a transient thermal source; i.e., irregularities in resonances occur due to transient thermal effects, as observed from a comparison of Figure 19.4(a) and Figure 19.7(a).

In the temporal domain, Mode-I resonance with an in-plane thermal source at $v = j * \omega_{mn}$, $j = 3$ is shown in a quasi-periodic oscillation in Figure 19.8(a), and in the Poincare map in Figure 19.8(b). The Poincare map is a graphical tool to illustrate periodic trace of the phase portrait, which can identify various oscillation patterns such as chaotic, quasi-periodic or periodic oscillations. The map is generated at its periodic interval of $T = 8.6$ ms for 200 iterations. For $j = 1$, quasi-periodic oscillation is found at a period of $T = 14.75$ ms. However, for $j = 2$, Figure 19.9(a) shows chaotic behavior in the Poincare map. The map is generated for 1000 cycles at a period of $T = 2\pi/v = 18$ μs, $v = 2 * \omega_{11}$ Hz. The deflection is increased from 0.6 mm without the in-plane thermal source, as indicated by the horizontal line of Mode-I in Figure 19.7(a), to 2.8 mm in chaotic oscillation in Figure 19.9(a). Chaos can also be found with other thermal frequencies in the range of $j = [2, 3]$. At $j = 1$, quasi-periodic behavior is observed for Mode-II, with a deflection of about 1.18 mm, as shown in Figure 19.9(b).

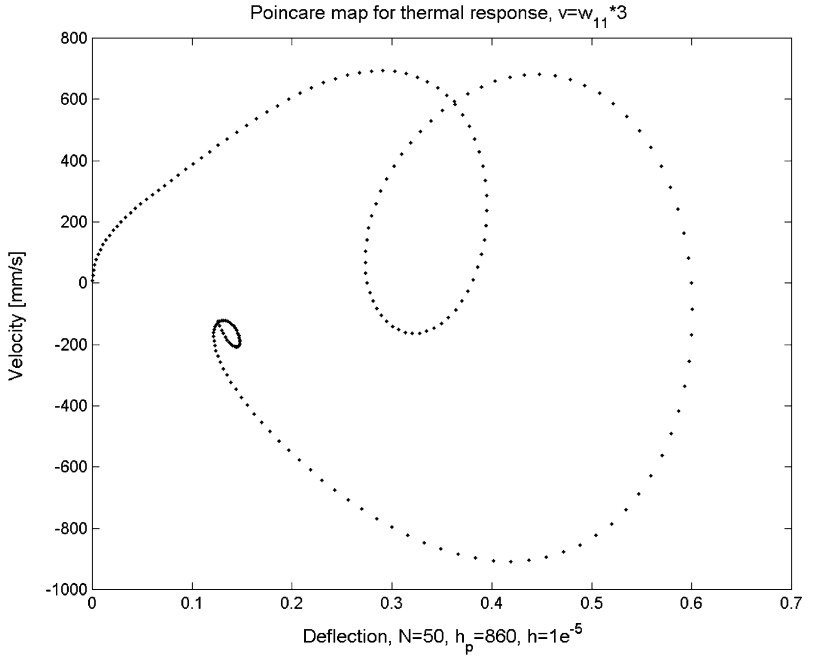
It is observed that chaotic oscillations for both modes have a deflection of over half of the laminate thickness, while the quasi-periodic oscillations are below the laminate thickness. Between half- and full-thickness of the laminate, the oscillations can be either quasi-periodic or chaotic. Both quasi-periodic and chaotic oscillations in both modes present similar patterns, i.e., stable oscillations with an attractor at the equilibrium point (0, 0) in the phase diagrams, as indicated in Figure 19.9(a) and Figure 19.9(b).

- Transient in-plane and transverse thermal fields

With a transient transverse thermal field, i.e., $Q^T = q_{mn}^T \cos(\omega t)$, $\omega = k\omega_{mn}$, $k \neq 0$, another harmonic source is introduced into the system. With frequency variations of both the in-plane and transverse thermal fields, Figure 19.10(a) and Figure 19.10(b) show Mode-I frequency responses with respect to the transverse thermal frequency $k = [0\ 10]$ with different j values. It is observed that the resonance behavior is different for each different thermal frequency j. The steady-state in-plane temperature variation represented by $j = 0$ increases the resonance frequency. With $j = 1$ and $j = 2$, a sustained higher deflection is observed with chaotic behavior with respect to the transverse thermal frequency between $k = [0\ 10]$. In Figure 19.10(b), the response with increased thermal frequency $j = 5$, 6, and 7 almost overlaps with the response without thermal source except for some minor deflection rise, or subresonance, as the response diminishes. For Mode-II, identical behavior can be found

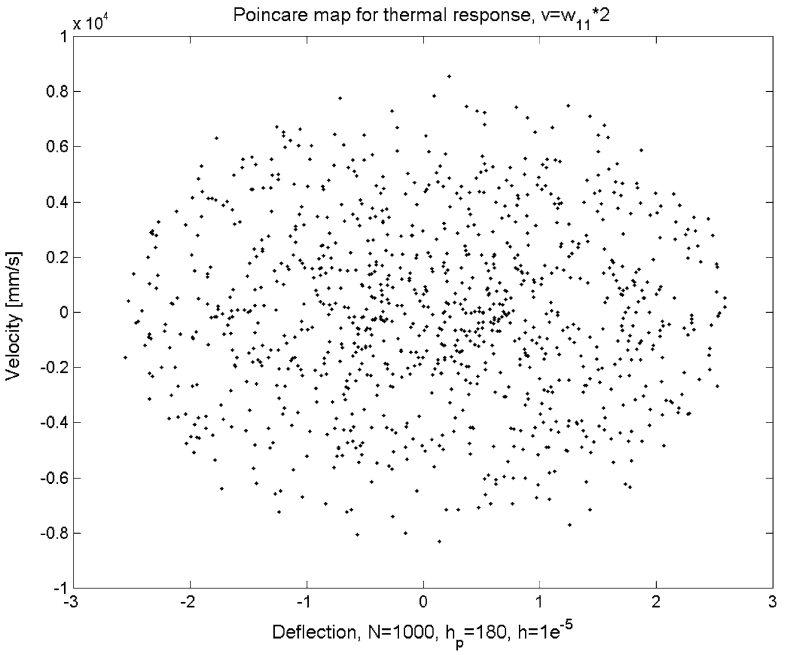


(a)

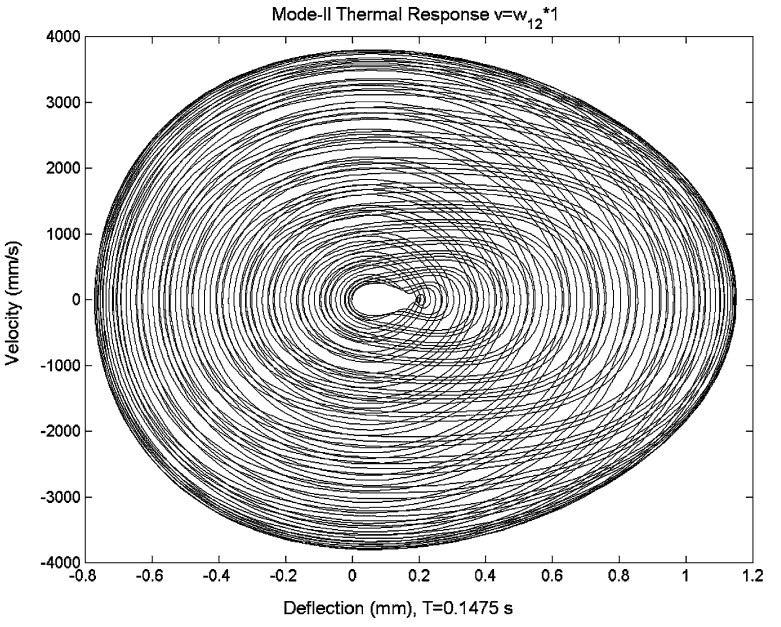


(b)

FIGURE 19.8. Mode-I thermal response at $j = 3$ with constant transverse thermal effect. (a) Velocity-deflection diagram, $N = 50$. (b) Poincare map, $N = 200$, $T = 8.6$ ms.

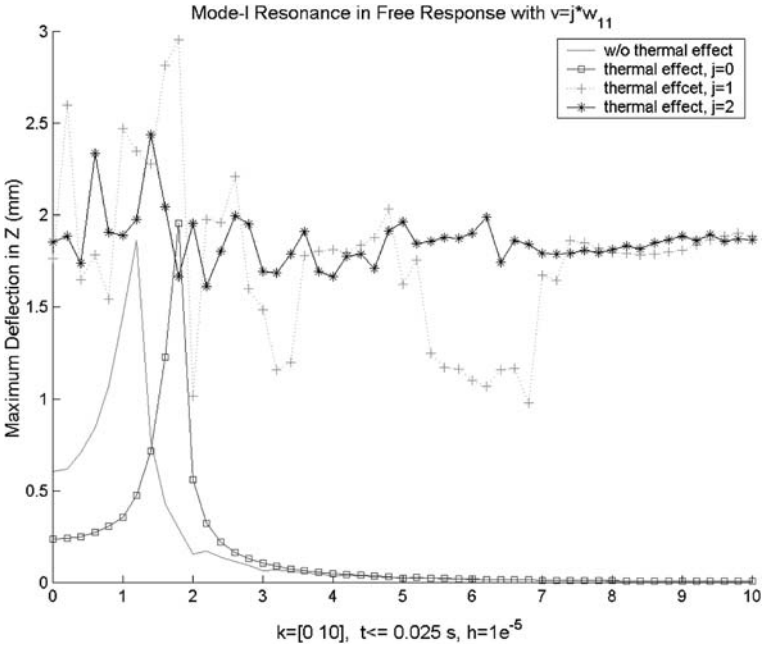


(a)

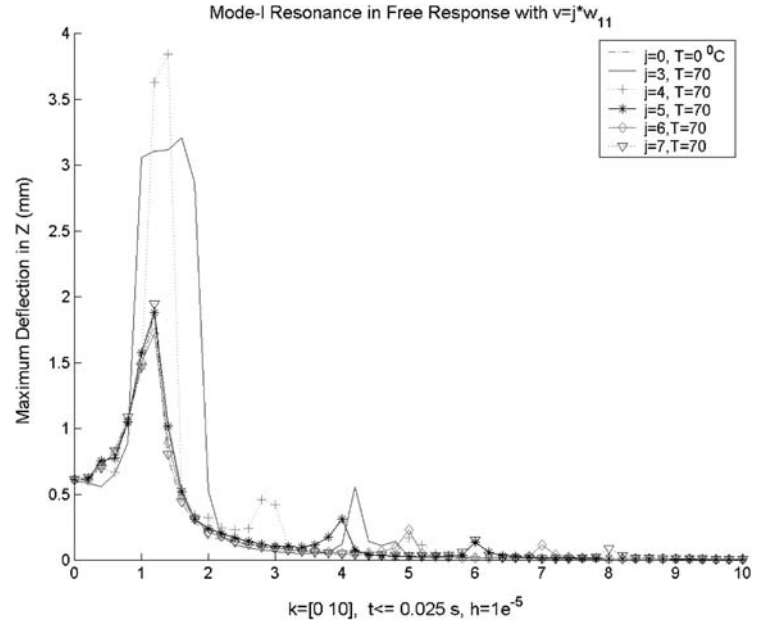


(b)

FIGURE 19.9. Thermal response with constant transverse thermal effect. (a) Poincare map for Mode-I, $j = 2$, $k = 0$, $N = 1000$, $T = 18$ μs. (b) Mode-II, $j = 1$, $k = 0$, $t = 0.1475$ s.



(a)



(b)

FIGURE 19.10. Thermal resonance vs. transverse thermal frequency for Mode-I. (a) $j = 0, 1, 2$, $k = [0\ 10]$. (b) $j = 0, 3, 4, 5, 6, 7$, $k = [0\ 10]$.

also with $j = 1$ and $j = 2$, where resonance bifurcation is replaced by a sustained deflection. Irregular deflection are also present with $j > 3$, similar to that of Mode-I, which suggests that high thermal frequencies have an insignificant influence on resonance and total responses for both modes.

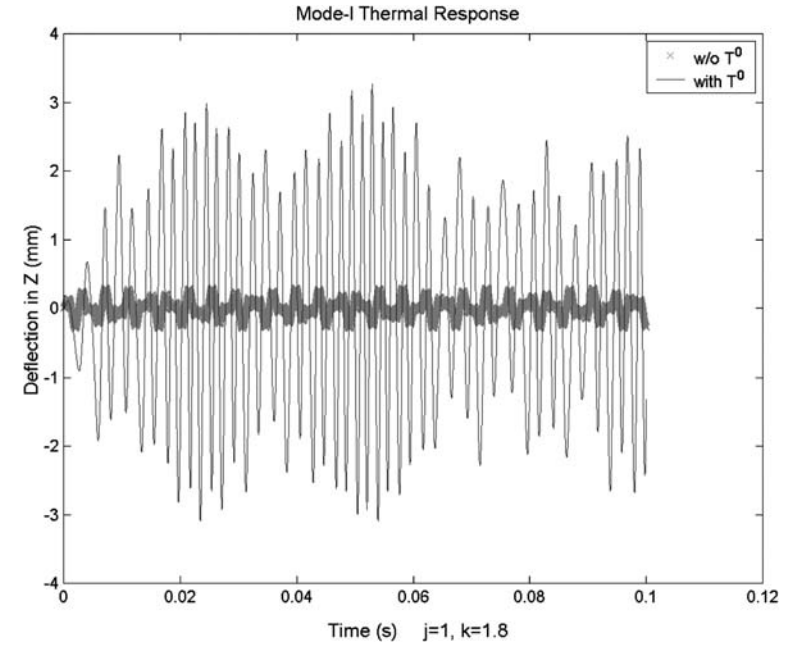
Chaotic response in the temporal domain with both transient thermal sources, i.e., the in-plane and transverse thermal fields, is shown in Figure 19.11(a), with $j = 1$, $k = 1.8$. The maximum deflection is about 3.5 mm. Without the in-plane thermal field, the deflection is about 0.3 mm in quasi-periodic oscillation, shown in Figure 19.11(b). This phenomenon associates chaos with the transient in-plane thermal field, rather than the transverse thermal field. As analyzed earlier, the circular term in the Duffing equation couples the deflection with the transient in-plane thermal source, and introduces chaos. Such influence is more evident when the deflection is higher, to cause instantaneously deflection response in chaos. Therefore, the quasi-periodic response remains for a modal response in an insignificant deflection, as shown in Figure 19.11(b) for Mode-I, with a deflection of less than half of the laminate thickness.

19.3.2.3. Thermal and Mechanical Response with Transient Thermal Field

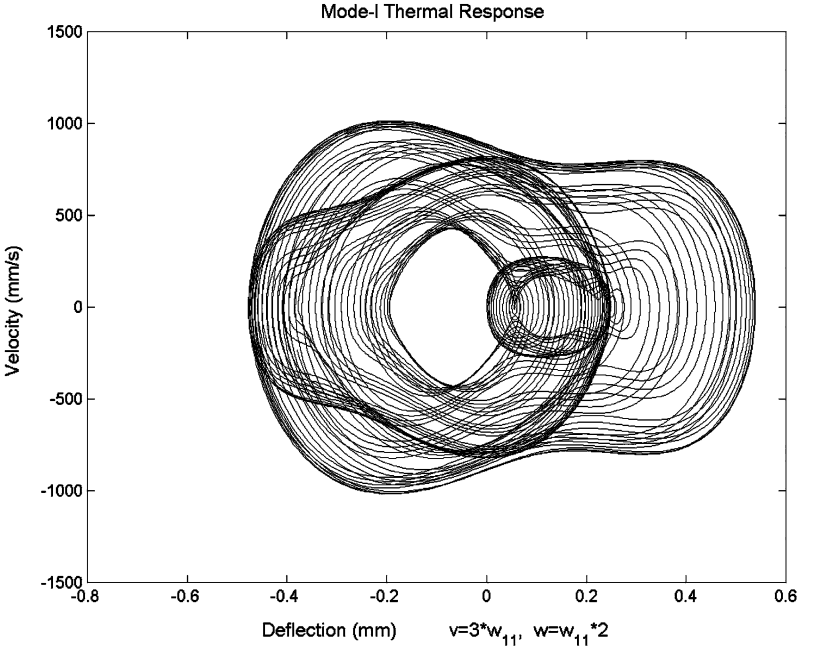
- Frequency response

  Frequency response is shown in Figure 19.12(a) for Mode-II with $Q = 2.88$ N/cm^2 and $k = [0\ 10]$, with $j = 0, 3, 4, 5, 6, 7$, with an external harmonic load in the form of $(Q + q_{mn}^T)\cos(\omega t)$. Resonance at higher deflection is observed, with this additional external load. Subsequent to resonance, deflection irregularities arise at certain frequency k with an increased deflection, similar to those observed in Figure 19.10(b) for the thermally induced responses only. However, consistency exists in these minor irregularities between j and k values. As indicated in Figure 19.12(a), it is found that such irregular phenomena occurs at $k = j \pm 1$ for each curve of $j = 3$, 4, 5, 6, 7. Since the in-plane thermal frequency is $v = j * \omega_{mn}$ and the external excitation frequency is $\omega = k * \omega_{mn}$, this means that the irregular response is caused by the $\omega - v = \pm\omega_{mn}$ terms in the harmonic function. This deduction is in agreement with the study of the same form of Equation (19.16) [19], where the equation with a small perturbation term χ_{mn} presents identical behavior. The perturbation solution for the equation yields terms in exponential function with index $\omega \pm v = \omega_{mn}$. This makes the response corresponding to the term $\omega - v = \omega_{mn}$ higher in deflection than those high modal frequency terms with index $\omega + v = (k + j)\omega_{mn}$. Alternately, it can be viewed as re-combination of the two harmonic sources, and $k = j \pm 1$ term dominates the response. Therefore, such characteristics can be observed in different systems, since it is independent of the magnitude of the system parameters.
- Constant load effect on the chaotic transition

  When the thermal mechanical load is assumed constant, i.e., $Q_total = Q_{mn} + q_{mn}^T$, the system behavior would be influenced by the external load. It is found that with a minor external load, chaotic behavior due to the transient in-plane thermal effect remains. Under an increased load Q_{mn}, chaotic response transforms into quasi-periodic behavior, since constant load dominates the forcing. Figure 19.13(a) shows Mode-II quasi-periodic response with $Q = 2.4$ N/cm^2, $q_{12}^T = 0.237$ N/cm^2, and $j = 2$, $k = 0$. Without external load, the response is chaotic as shown in Figure 19.13(b). Further increasing load Q_{mn}, the quasi-periodic behavior remains with different oscillation patterns.

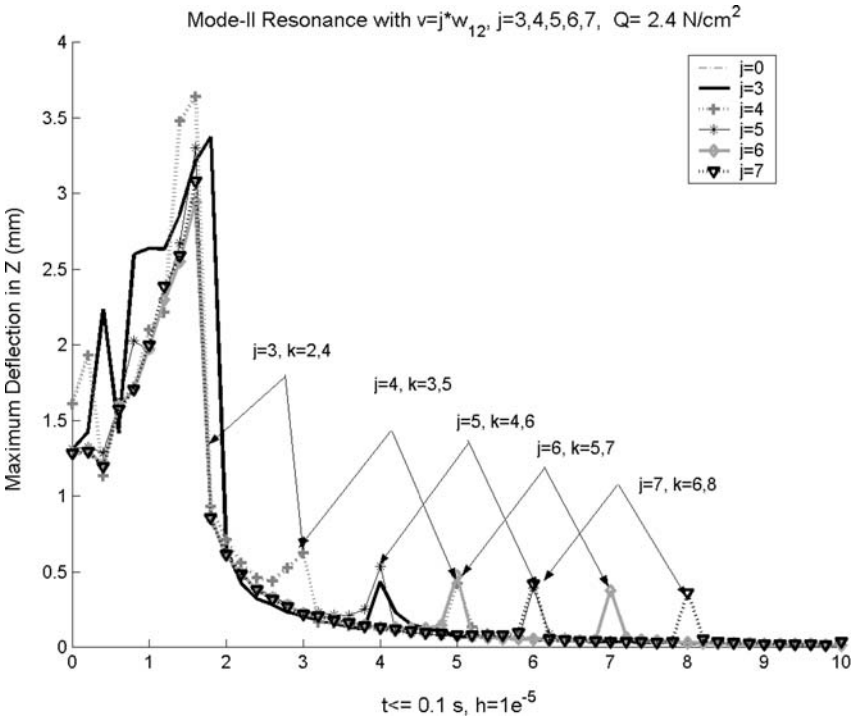


(a)

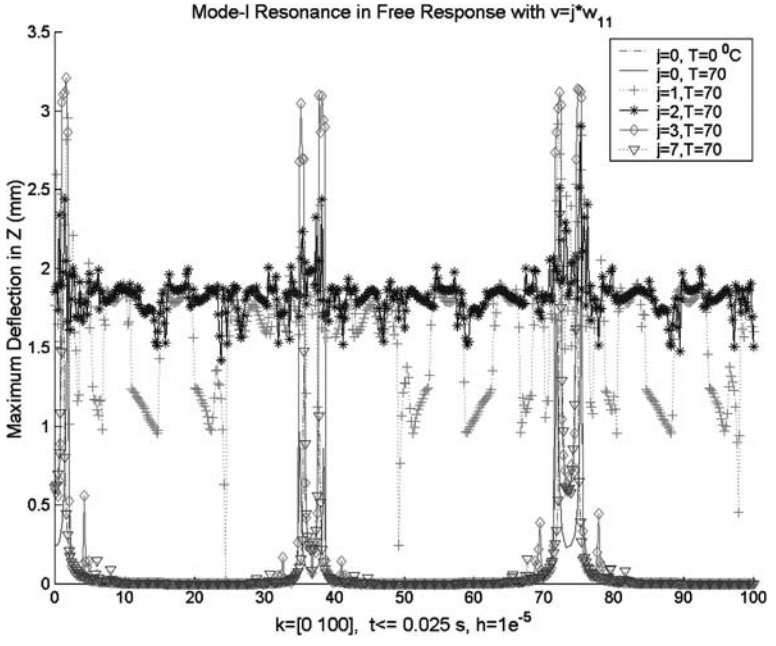


(b)

FIGURE 19.11. Response with both transient thermal sources. (a) Mode-I at $j = 1$, $k = 1.8$. (b) Mode-I, $j = 3$, $k = 2$.

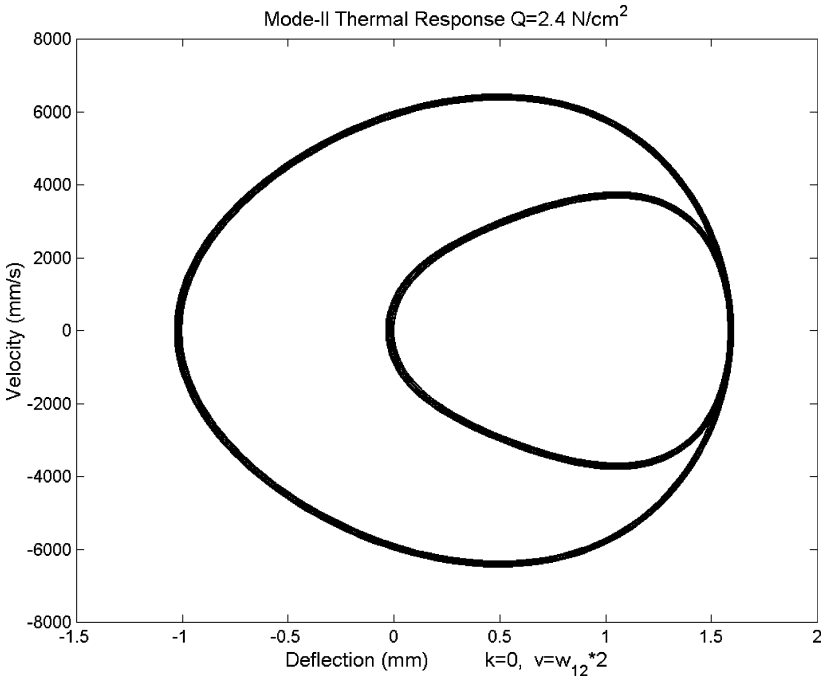


(a)

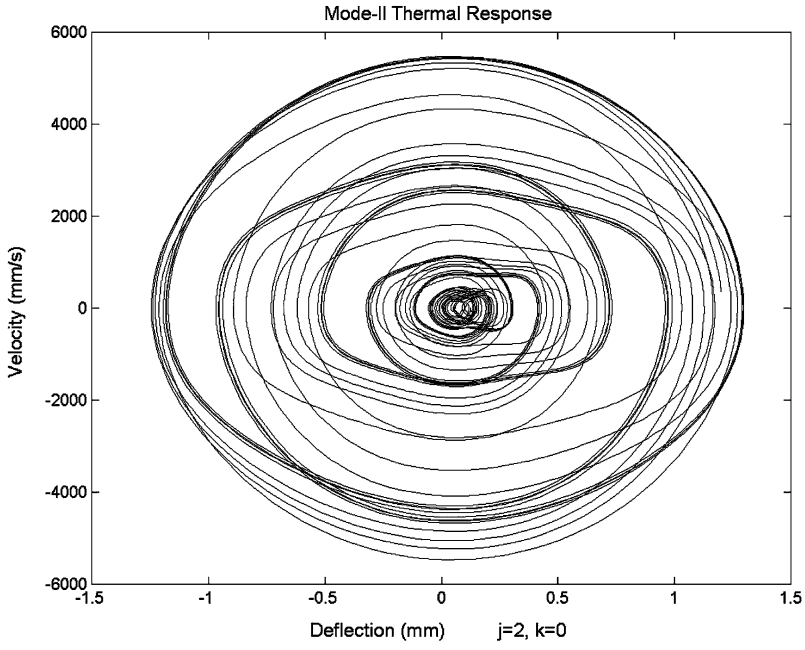


(b)

FIGURE 19.12. Frequency response. (a) Mode-II with $Q = 2.4$ N/cm^2, $k = [0\ 10]$, $j = 0, 3, 4, 5, 6, 7$. (b) Thermal response of Mode-I with $k = [0\ 100]$, $j = 0, 1, 2, 3, 7$.

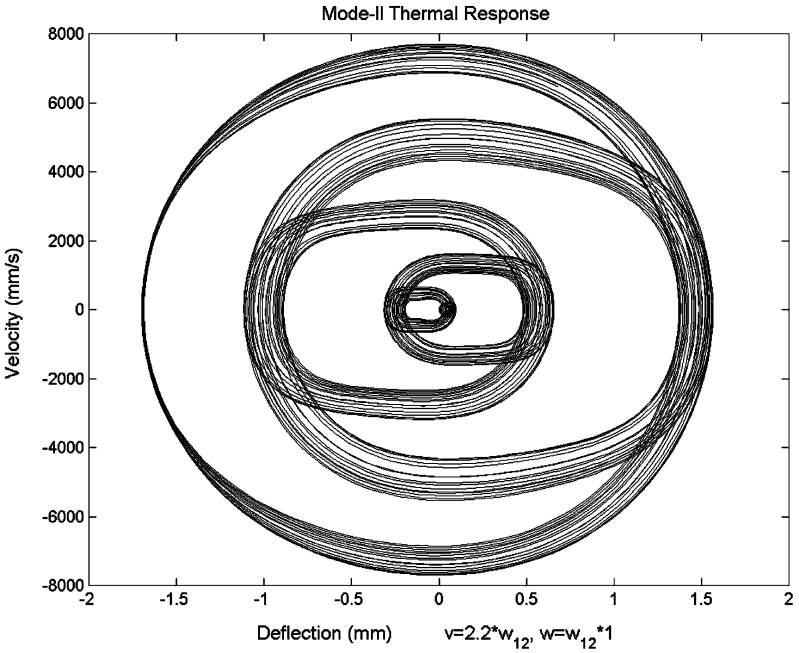


(a)

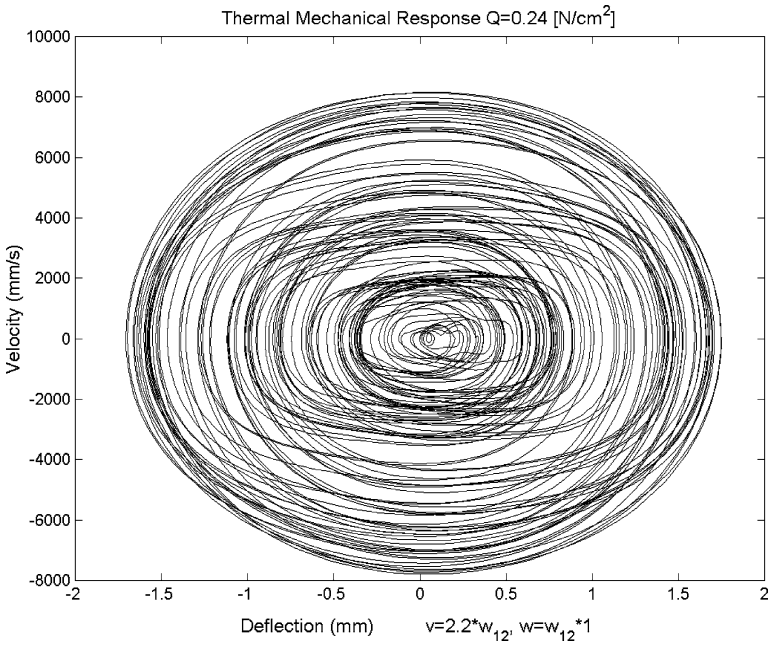


(b)

FIGURE 19.13. Transition with constant load in thermal mechanical response. (a) Mode-II with $j = 2$, $k = 0$, $Q = 2.4$ N/cm^2. (b) Mode-II with $j = 2$, $k = 0$, $Q = 0$.

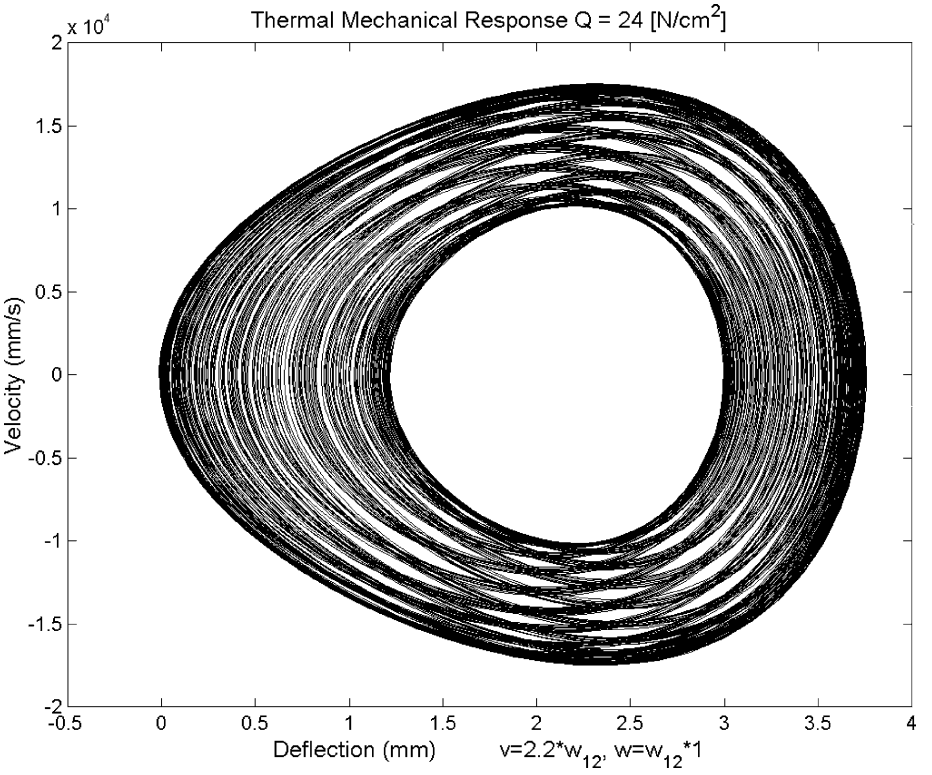


(a)



(b)

FIGURE 19.14. Transition with constant external load. (a) Mode-II with $j = 2.2$, $k = 1$, $Q = 0$. (b) Mode-II with $j = 2.2$, $k = 1$, $Q = 0.24$ N/cm^2. (c) Mode-II with $j = 2.2$, $k = 1$, $Q = 24$ N/cm^2.



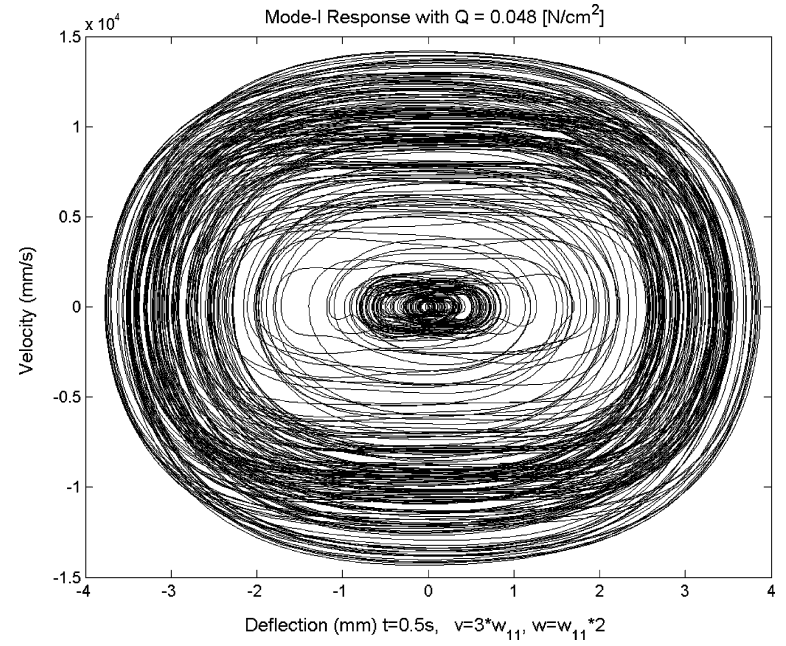
(c)

FIGURE 19.14. (Continued).

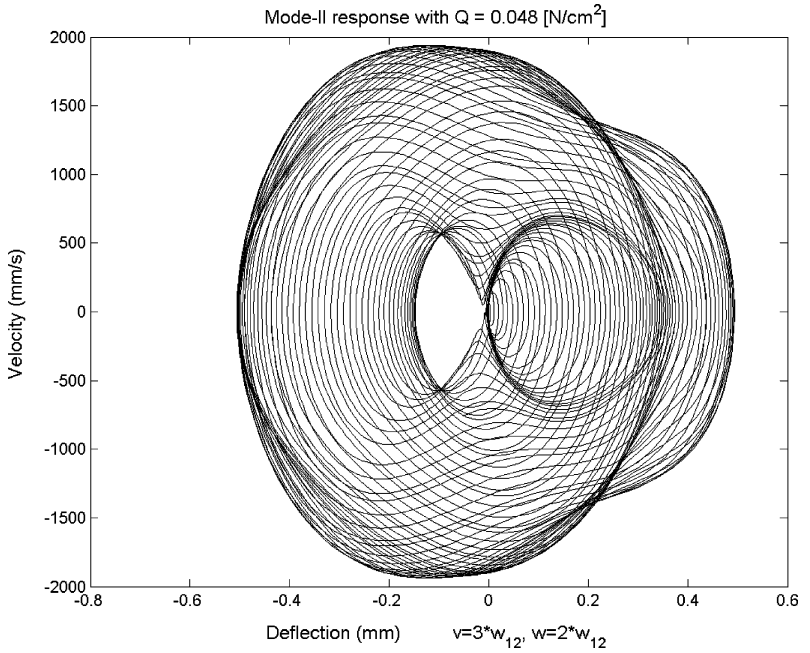
When a transient transverse thermal field is applied with a constant external load Q_{mn}, i.e., $Q = Q_{mn} + q_{mn}^T \cos(\omega t)$, it is found that quasi-periodic behavior with both transient thermal sources can be transformed into chaos with a moderate external constant load. However, the increased constant load Q_{mn} can dominate, and transform the chaotic response back into quasi-periodic behavior again. Figure 19.14(a), Figure 19.14(b) and Figure 19.14(c) show the Mode-II response with $j = 2.2$, $k = 1$ in such transition. With a constant load of $Q = 0.24$ N/cm^2, chaos is induced shown in Figure 19.10(b); while without this external load Q, it is quasi-periodic as indicated in Figure 19.14(a). When the load is further increased to $Q = 24$ N/cm^2, chaotic response returns to quasi-periodic response, shown in Figure 19.14(c).

- Harmonic forcing on the response transition

  In contrast to the transition from chaos to quasi-periodic subject to a constant external load, the harmonic forcing in the form $Q = (Q_{mn} + q_{mn}^T)\cos(\omega t)$, $\omega = k\omega_{mn}$ transforms the system from quasi-periodic to chaotic oscillation. Figure 19.15(a) shows the Mode-I chaotic response with $j = 3$, $k = 2$, $q_{11}^T = 0.234$ N/cm^2 and a small perturbation load $Q_{11} = 0.048$ N/cm^2. Without this external load, the thermal response is quasi-periodic as shown in Figure 19.11(b). Similar behavior is observed with Mode-II, as shown in Figure 19.15(b) and Figure 19.15(c) for such transition. Mode-II transition to chaos is found to occur at a higher load, i.e., $Q_{11} = 19.2$ N/cm^2

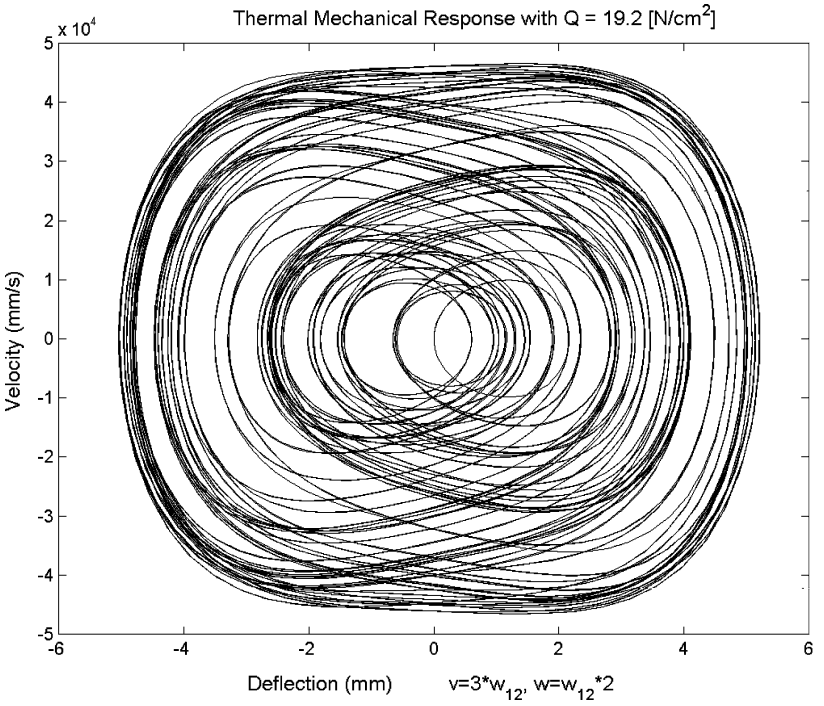


(a)



(b)

FIGURE 19.15. Transition with harmonic load in thermal mechanical response. (a) Mode-I, $j = 3$, $k = 2$, $Q = 0.048$ N/cm^2. (b) Mode-II, $j = 3$, $k = 2$, $Q = 0.048$ N/cm^2. (c) Mode-II, $j = 3$, $k = 2$, $Q = 19.2$ N/cm^2.



(c)

FIGURE 19.15. (Continued).

at $j = 3$, $k = 2$. The response remains chaotic for both modes with further increase of the harmonic external load. The transition from quasi-periodic to chaos usually sees an appreciable deflection increase.

It is noted that both modes present similar behaviors subject to identical forcing conditions, except that the loading level to cause transition from quasi-periodic oscillation to chaos, or vise versa, is higher for Mode-II than for Mode-I. This is because an unstable system can become chaotic with only a small perturbation as compared to a stable system.

19.4. STRESS FIELD IN NONLINEAR DYNAMICS RESPONSE

19.4.1. Stress Field Formulation

The lamina stresses in the plane stress condition can be expressed in modal form as a function of deflection and thermal field, in the form of:

$$\{\sigma\}^{(k)} = \sum_{n=1}^{\infty}\sum_{m=1}^{\infty}\left\{f_1[W_{mn}(t)]^2 - f_2 z W_{mn} - f_3 \varsigma T^c_{mn} + f_4 \varsigma T^0_{mn}\right\}, \tag{19.19}$$

where f_1, f_2, f_3, f_4 are functions defined by the material properties and geometries [25].

The expression of the stress field indicates that the temperature variation introduces stress redistribution by both the in-plane thermal function T^0_{mn} and T^c_{mn}. The transverse thermal field T^1_{mn} does not affect the stress distribution because the in-plane stress condition requires that the transverse coefficient of thermal expansion be assumed zero. However, the transverse thermal effect is embedded in the deflection, since the deflection $|W_{mn}(t)|$ is affected by thermal fields, mechanical load, as well as initial conditions. Therefore, although the stress field formulation identifies the contribution of deflection and the in-plane temperature variation only, the nonlinear coupling of both thermal fields can be evaluated.

19.4.2. Stress Distribution

Figure 19.16(a) is the stress distribution at the top of the Cu lamina for its fundamental mode response, with a deflection $|W_{11}(t)| = 3.3$ mm only. The thermal mechanical stress field σ_x is shown in Figure 19.16(b) with the in-plane thermal field as assumed above, i.e., $T^0_{11} = 94.3°$C. The high stresses at the corner due to temperature function $T_0(x, y)$ are noticeable in Figure 19.16(b). The stress distribution of σ_y is identical to that of σ_x by exchanging the x and y axis. Figure 19.16(c) is the shear stress τ_{xy}, which is much lower in magnitude than those of σ_x and σ_y. The thermal stress σ_x with $T^0_{11} = 42.75$ and $T^c_{11} = 35°$C is shown in Figure 19.16(d). It is noted that compression is induced from the temperature rise due to boundary constraints to the thermal expansion.

19.4.3. Failure Analysis

Failure analysis for the laminate can be made based on the stress field associated with the specified thermal field and deflection. Since the laminate is made of composite structures in epoxy and Cu, the Hsai-Wu Failure Criteria can be applied for failure analysis of composite materials in a plane stress condition [30]. This criterion has been verified empirically adaptable for the composites composed of epoxy and other laminates. It identifies failure conditions due to the interaction of the stresses σ_x, σ_y and τ_{xy}, even the laminate may not fail according to the Maximum Stress Criteria in certain cases. The Hsai-Wu failure criterion states that failure of a composite material will occur if:

$$R = F_1\sigma_1 + F_2\sigma_2 + F_{11}\sigma_1^2 + F_{22}\sigma_2^2 - \sqrt{F_{11}F_{22}}\sigma_1\sigma_2 + F_{66}\tau_{12}^2 \geq 1, \tag{19.20}$$

where F_{ij} and F_k $(i, j, k = 1, 2, 6)$ are the coefficients pertaining to the material properties and structures. For isotropic materials, assuming the same yield strength in tension and compression, R is reduced to:

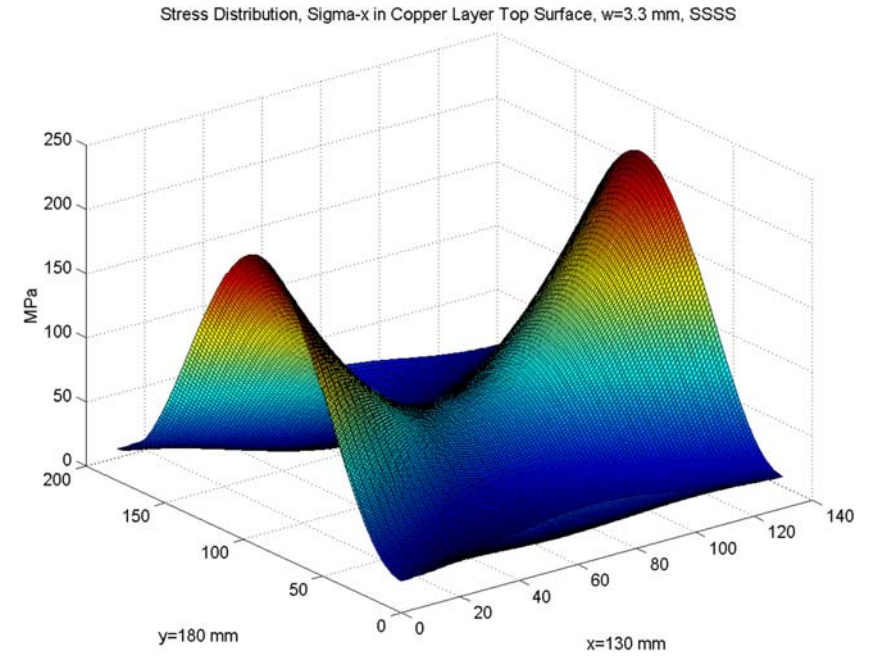
$$R = F_{11}(\sigma_1^2 + \sigma_2^2 - \sigma_1\sigma_2) + F_{66}\tau_{12}^2 \geq 1, \tag{19.20a}$$

or

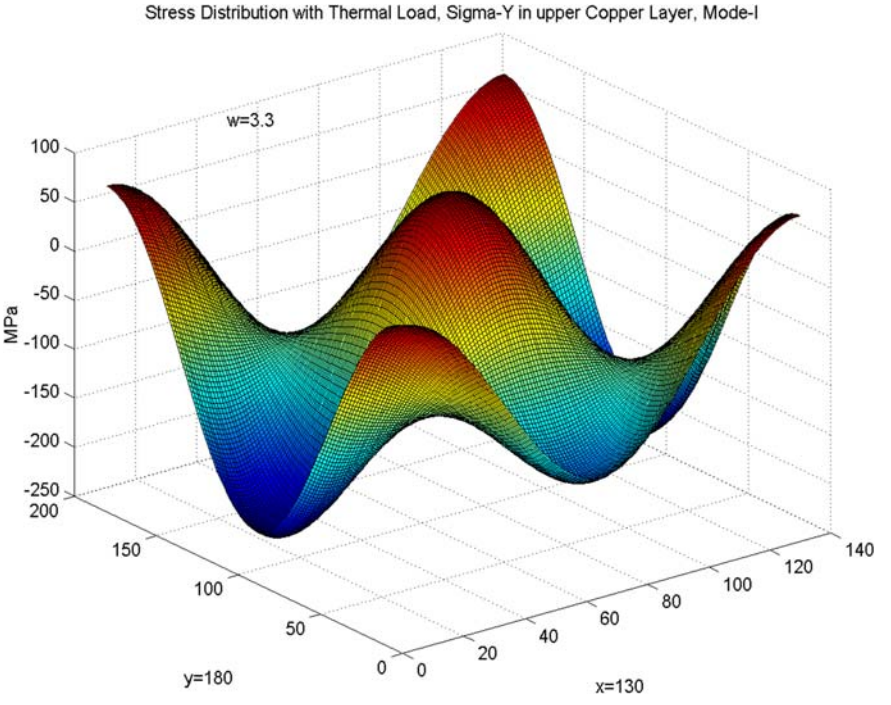
$$R = \frac{\sigma_1^2 + \sigma_2^2 - \sigma_1\sigma_2}{(\sigma^Y)^2} + \frac{\tau_{12}^2}{(\tau_{12}^F)^2} \geq 1. \tag{19.20b}$$

The material properties are chosen as:

$$\text{Cu layer: } \sigma^Y = \sigma_1^Y = \sigma_2^Y = 188.32 \text{ MPa}, \quad \tau_{12}^F = 150.00 \text{ MPa}, \quad v = 0.4, \tag{19.21a}$$

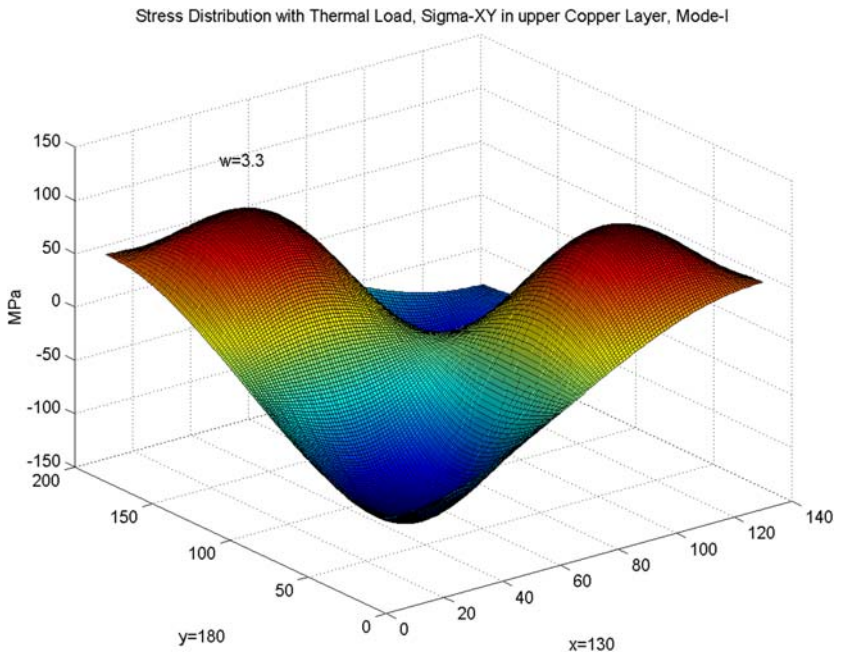


(a)

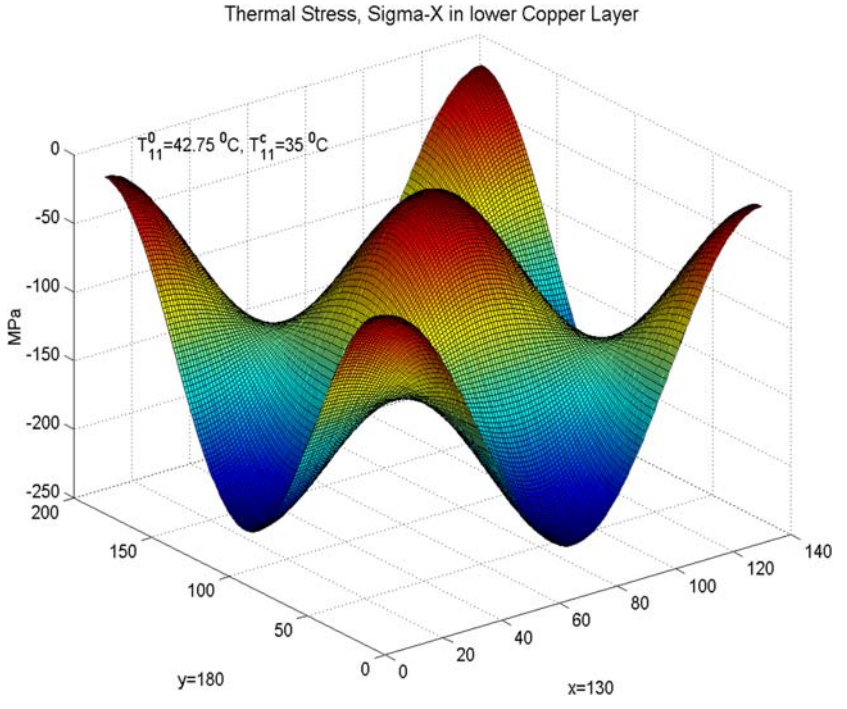


(b)

FIGURE 19.16. Mode-I stress distribution at the top of the Cu lamina. (a) σ_x without thermal effect $|W(t)| = 4.2$ mm. (b) σ_x with in-plane $T_0(x, y)$ source. (c) τ_{xy} with $T_0(x, y)$ effect. (d) σ_x with $T_{11}^0 = 42.75°\text{C}$ and $T_{11}^c = 35°\text{C}$.



(c)



(d)

FIGURE 19.16. (Continued).

TABLE 19.3.
Case study of stress and failure analysis.

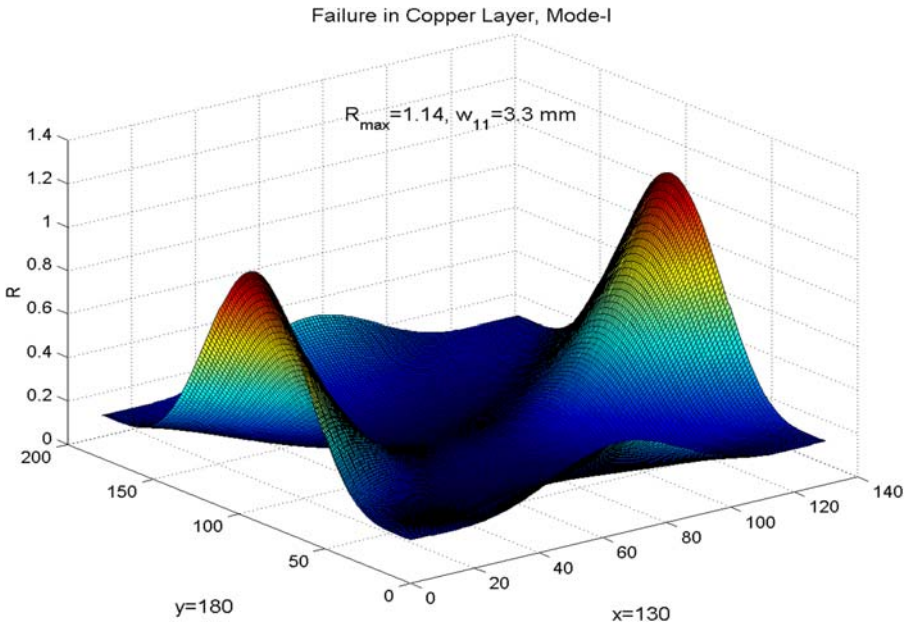
Cases	Deflection (mm)	Thermal T^0 (°C)	Thermal T^C (°C)	Max. stress (MPa) s_x	Max. stress (MPa) s_y	Failure R
I Figure 19.17(a)	3.3	0	0	+226.63	+137.85	1.14
				+5.35	+40.19	
II Figure 19.17(b)	3.3	95.42	35	+168.94	+89.65	1.69
				−282.97	−44.83	
III Figure 19.17(c)	4.2	47.71	35	+169.92	+81.94	1.19
				−163.74	+24.05	
IV Figure 19.17(d)	3.3	47.71	17.5	+119.48	+64.70	0.45
				−132.35	+18.84	
V Figure 19.17(e)	0	42.75	35	−14.99	−36.96	1.02
				−199.91	−109.75	
VI	4.2	0	0	367.10	223.29	3.0
				32.19	72.52	

$$\text{FR-4 layer: } \sigma^Y = \sigma_1^Y = \sigma_2^Y = 250.00 \text{ MPa}, \quad \tau_{12}^F = 230.00 \text{ MPa}, \quad v = 0.19. \tag{19.21b}$$

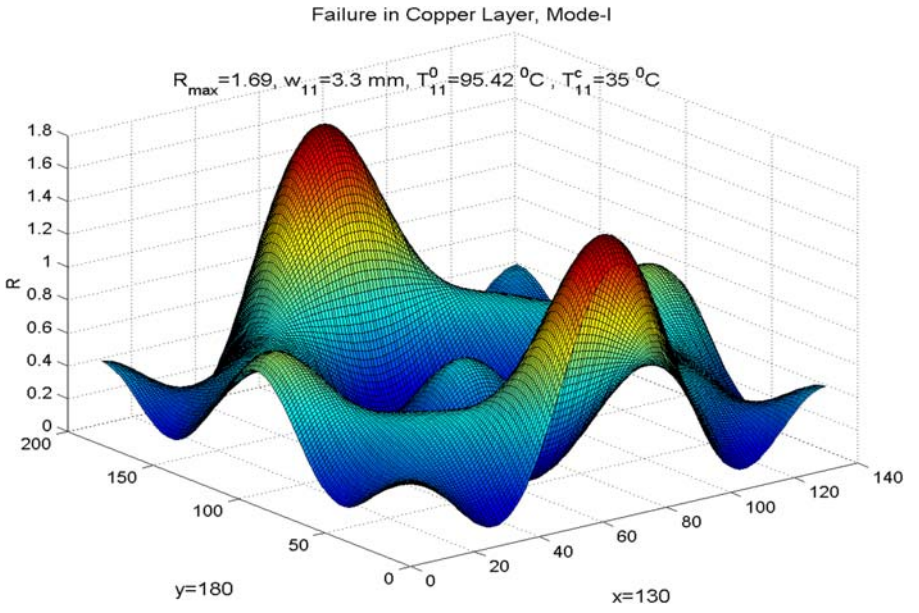
For the given material properties and the structures of the laminate, the scalar R is a function of temperature coefficients T_{mn}^0 and the deflection magnitude $|W_{mn}(t)|$. The locus with magnitude of $|R(x, y)| > 1$ corresponds to the zone of failure. For a specified deflection, the critical temperature can be found from failure analysis, or vise versa. It can also be deduced that the most vulnerable lamina for the PWB is Cu, due to its higher stress field associated with the higher stiffness, as compared to the FR-4 lamina.

The effect of temperature variation on the stress condition is studied for Cu lamina for its fundamental mode response. Different combinations of thermal field and mechanical load that lead to failure are analyzed, as listed in Table 19.3. Finite difference computation generates the stress field and failure diagram associated with the cases listed in Table 19.3. It is noted from Table 19.3 that failure would not occur, without a temperature rise $T_0(x, y)$, until the maximum deflection reaches about twice the lamina's thickness, i.e., $|W_{11}| =$ 3.3 mm, as indicated in Figure 19.17(a) with R_max = 1.14. The corresponding maximum stress $\sigma_{x_\max}$ is 226.63 MPa, exceeding the tensile strength of 189.27 MPa. Therefore, failure would occur based on the Maximum Stress Failure Criteria, which agrees with the Hsai-Wu Failure Criterion since R_max > 1.

$R(x, y)$ function with $|W_{11}| = 3.3$ mm, $T_{11}^0 = 95.42°\text{C}$ and $T_{11}^c = 35°\text{C}$, as shown in Figure 19.17(b), increases significantly to R_max = 1.69 with this thermal effect. The joint thermal and mechanical stress field has the maximum tensile stress $\sigma_x = 168.94$ MPa, and the maximum compression in σ_x of 282.97 MPa. Failure zones with $|R| > 1$ are caused by the intense compression on the edges of the lamina, from the observations shown in Figure 19.16(a) and Figure 19.16(c). Reducing the temperature field by 50% to $T_{11}^0 = 47.71°\text{C}$ and $T_{11}^c = 35°\text{C}$, failure is initiated with a higher deflection of $|W_{11}| = 4.2$ mm as shown in Figure 19.17(c), with R_max = 1.19. Correspondingly, the maximum tension in σ_x is $\sigma_{x_\max} = 367.10$ MPa with deflection $|W_{11}| = 4.2$ mm alone, while this tension is reduced to 169.92 MPa with the above thermal filed indicated in Figure 19.17(c).

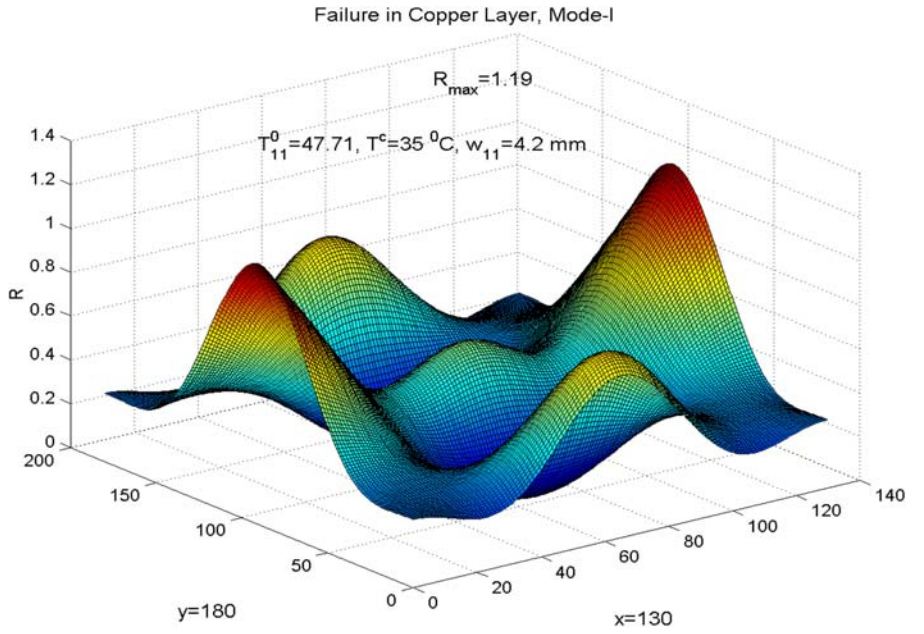


(a)

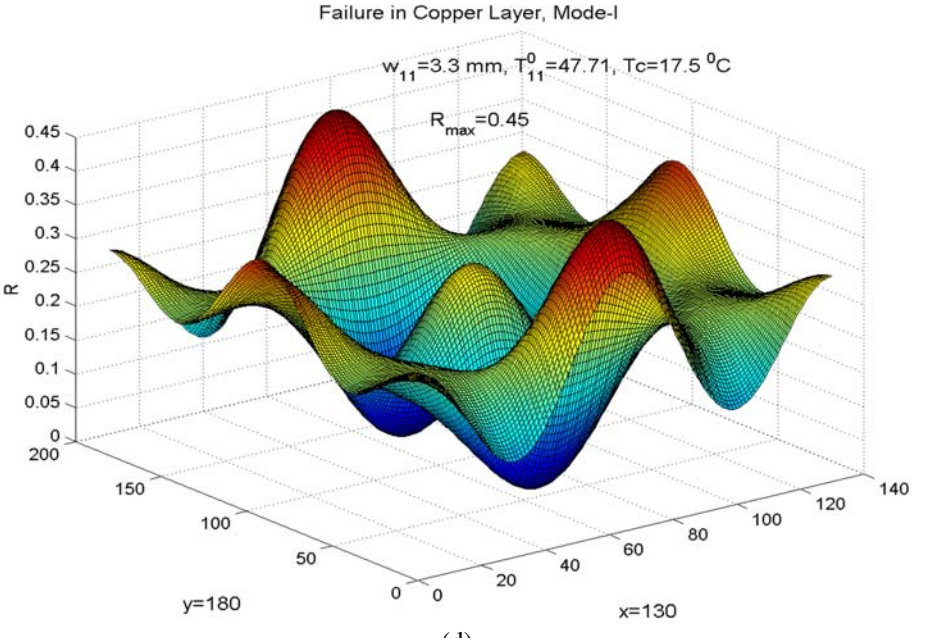


(b)

FIGURE 19.17. Failure diagrams of Cu lamina with different thermal load and deflections. (a) $|W(t)| = 3.3$ mm. (b) $|W(t)| = 3.3$ mm, $T_{11}^0 = 95°\text{C}$, $T_{11}^c = 35°\text{C}$. (c) $|W(t)| = 4.2$ mm, $T_{11}^0 = 47.71°\text{C}$, $T_{11}^c = 35°\text{C}$. (d) $|W(t)| = 3.3$ mm, $T_{11}^0 = 47.71°\text{C}$, $T_{11}^c = 17.5°\text{C}$. (e) $T_{11}^0 = 42.75°\text{C}$, $T_{11}^c = 35°\text{C}$.

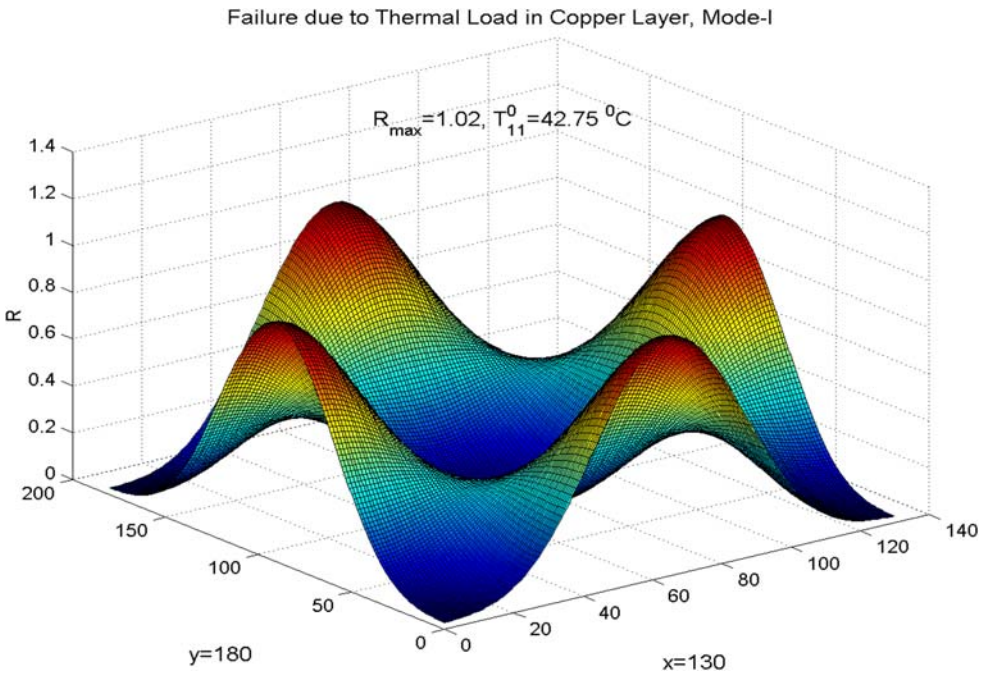


(c)



(d)

FIGURE 19.17. (Continued).



(e)

FIGURE 19.17. (Continued).

For $|W_{11}| = 3.3$ mm, $T_{11}^0 = 47.71°$C, and $T_{11}^c = 17.5°$C, as shown in Figure 19.17(d), it is found that R_max = 0.45. The maximum tension without thermal source, $\sigma_{x_\max} =$ 226.63 MPa, is reduced to 119.48 MPa, and the maximum compression changes to 132.35 MPa due to thermal effect. A comparison between Figure 19.17(a) and Figure 19.17(d) means that the thermal field at a certain level could lessen the tension due to the thermally induced compression. The tensile stress field is either being transformed into compression, or being reduced to a lower magnitude.

A comparison between Figure 19.17(c) and Figure 19.17(d) indicates that a higher in-plane temperature rise T_{11}^c and the higher deflection cause more damage. The thermal field itself could induce failure. In Figure 19.17(e), with a temperature rise of $T_{11}^0 = 42.75°$C and $T_{11}^c = 35°$C, failure is initiated, since R_max = 1.02. Both R values in Figure 19.17(a) for failure due to deflection only, and in Figure 19.17(e) for failure due to thermal load only, reach the threshold value. However, a combination of both fields, results in a reduced R, as can be deduced from Figure 19.17(c). This suggests a non-monotonic relation between the critical stress to failure and the thermal field and deflection to induce such a stress field, although either increased deflection, or intensified thermal field will certainly makes the laminate prone to failure.

19.5. DISCUSSIONS

The response analyses indicate that both the steady-state and the transient thermal fields have strong influences on the response behavior of a thin laminated microstructure.

As far as the in-plane thermal field is concerned, a steady state temperature variation affects quasi-periodic resonance frequencies and deflection magnitudes, while a transient thermal source is likely to transform the response from quasi-periodic to chaotic oscillation, with a deflection of the laminate thickness or higher. It is also demonstrated that the non-uniform in-plane temperature variation is more influential than the transverse thermal field on the dynamic behavior of a thin laminate. This is not only because of the difference in the magnitude of the two thermal sources, but also because of the direct influence of the in-plane thermal field on the system's natural frequency. The natural frequency oscillation due to the transient in-plane thermal field manifests itself into the chaotic behavior of the system, with increased deflection.

The influence of the transient thermal field on the response is also observed in its unique resonance bifurcation behavior in the frequency domain, when harmonic oscillation of both thermal fields is assumed. It is observed that sub-resonance occurs at the frequencies $\omega - v = \pm\omega_{mn}$ for systems with higher in-plane thermal frequencies $j \geq 3$, and a sustained deflection without resonance bifurcation occurs at $j = 1$, $j = 2$. The dependence on load for transition between quasi-periodic and chaotic response is another characteristic of the system.

The stress field formulation shows that the in-plane temperature rise contributes significantly to the stress field, while the transverse temperature variation has its effect on the deflection. It is found that, to a limited level of temperature rise, tensile stress intensity can be reduced due to the thermally induced compression. High deflection and high-rise temperatures will cause failure of lamina.

The thermal mechanical stress and failure analyses identify the critical conditions of deflection and thermal field with respect to various thermal mechanical loading conditions and initial conditions. This is important for a nonlinear system analysis, due to the fact that a direct relationship between excitation load and stress field cannot be established in an explicit function for such a nonlinear system. The stress field and failure criterion as a function of instantaneous deflection and thermal field make the formulation applicable for the transient thermal effect study. Meanwhile, the decoupled modal form equation and stress formulation can be applied to multi-mode analysis when high modes need to be taken into account.

19.6. SUMMARY

Theoretical development is made for the nonlinear dynamic response of a laminate under both mechanical and thermal load, induced by either steady-state or transient state temperature variations. The equation of motion for the laminate deflection is reduced to the Duffing equation by a Galerkin-type method. This makes it possible to characterize the nonlinear dynamic behavior in a modal analysis for a thin laminate subject to mechanical loading, and an arbitrary thermal field in either steady state or transient state.

It is demonstrated that an in-plane thermal source has a more significant effect on both deflection and stress field of the laminate than that of the transverse thermal source. Although the transverse thermal effect is almost negligible in a steady-state thermal field study, it serves as a perturbation in the transient thermal field, and could cause transition of the response behavior. Thermally induced chaos is mainly attributed to the transient in-plane temperature variation, or to both in-plane and transverse transient thermal fields.

The study indicates that increased deflection at chaos increases the higher mode contribution to the total response, and makes multi-mode analysis necessary. This is in con-

trast to the response without thermal effect, where the dominant response results from the fundamental mode. It is also observed that an external loading can effectively control the response behavior, i.e., the transition between quasi-periodic and chaotic response can be defined by load.

Although most of the analyses are based on a single term of the Fourier series of both thermal and mechanical loading, the response characteristics, i.e., resonance, deflection, stress and failure condition, pertain to an arbitrary thermal field and loading conditions that may be represented by multi-mode functions, subject to different oscillation frequencies and loading level. The analytical formulation can also be applied to analyze the response with respect to the properties of the laminated materials and geometry of the microstructures for its behavior analysis and functional control.

NOMENCLATURE

A	Stiffness matrix of the laminate
A^*	Inverse matrix of A
A^*_{ij}	Elements of A^*
B	Coupling rigidity matrix of the laminate
C	Stiffness matrix of the laminate for constitutive equation of stress and strain
C_{ij}	Coefficient of the stiffness matrix
$(E)^{(k)}$	k-th lamina Young's modulus
$F(x, y, t)$	Airy's stress function
$\widetilde{F}(x, y, t)$	Airy's stress function with complimentary function
F_1, F_2, F_{11}, F_{22}, F_{66}	Coefficients of Failure function
N	In-plane force vector
N^T	Thermal force vector
N^F	Airy's stress function vector
M	Applied moment vector
M^T	Thermal moment vector
$\mathrm{N}(w)$	Function of in-plane force and deflection
$R(x, y)$	Failure Criteria function
$T_0(x, y)$	Temperature variation over the board
$T_1(x, y)$	Temperature variation through the thickness
T_c	Constant terms of $T_0(x, y)$
T^0_{mn}, T^1_{mn}, T^c_{mn}	Fourier series coefficients for $T_0(x, y)$, $T_1(x, y)$ and $T_c(x, y)$ respectively
r^*	Stiffness coefficient
p^*	Stiffness coefficient
s^*	Stiffness coefficient
$u(x, y)$	In-plane deformation of laminate in x direction
$v(x, y)$	In-plane deformation of laminate in y direction
$w(x, y, t)$	Transverse deflection
$\lvert W(t)\rvert$	Transverse deflection magnitude
$(\sigma)^{(k)}$	k-th lamina plane stress vector
σ^Y	Yield strength of the lamina

τ^F	Shear failure strength of the lamina
σ_i	Stress in i direction, $i = x, y, xy$
σ_i^2	Square of σ_i
ς, ξ	Coefficient for thermal force N^T and M^T, respectively
ω	Excitation frequency
$\omega_{mn}, \omega_{mn}^T$	Natural frequency without and with thermal effect, respectively.

ACKNOWLEDGMENT

This chapter is dedicated to the late Professor Robert Fulton for his support to this research.

REFERENCES

1. C. Acar and A.M. Shkel, Structural design and experimental characterization of torsional micromachined gyroscopes with non-resonant drive mode, Journal of Micromechanics and Microengineering, 14(1), pp. 15–25 (2004).
2. O. Aktas, N.R. Aluru, and U. Ravaioli, Application of a parallel DSMC technique to predict flow characteristics in microfluidic filters, Journal of Microelectromechanical Systems, 10(4), pp. 538–549 (2001).
3. A. Allison and D. Abbott, A MEMS Brownian ratchet, Microelectronics Journal, 33(3), pp. 235–243 (2002).
4. H. Aref, The Development of chaotic advection, Physics of Fluids, 14(4), pp. 1315–1325 (2002).
5. S. Baik, J.P. Blanchard, and M.L. Corradini, Development of micro-diesel injector nozzles via MEMS technology and effects on spray characteristics, Atomization and Sprays, 13(5-6), pp. 443–474 (2003).
6. A. Balogh, W.J. Liu, and M. Krstic, Stability Enhancement by Boundary Control in 2-D Channel Flow, IEEE Transactions on Automatic Control, 46(11), pp. 1696–1711 (2001).
7. A. Barut, E. Madenci, and A. Tessler, Nonlinear thermoelastic analysis of composite panels under non-uniform temperature distribution, International J. Solids and Structures, 37, pp. 3861–3713 (2000).
8. S. Basuli, Large deflection of plate problems subjected to normal pressure and heating, Indian J. Mech. Math., 6, pp. 1–14 (1968).
9. H.M. Berger, Large amplitude free vibration of rectangular plates subjected to aerodynamic heating, Journal of Applied Mechanics, 4(1), pp. 39–49 (1955).
10. L. Bergman, Special issue on MEMS modeling in dynamics and acoustics—Foreword, Journal of Vibration and Acoustics, Transactions of the ASME, 126(3), pp. 325–325 (2004).
11. S. Bhansali, H. Benjamin, V. Upadhyay, N. Okulan, K.W. Oh, H.T. Henderson, and C.H. Ahn, Modeling multilayered MEMS-based micro-fluidic systems, JOM, 56(3), pp. 57–61 (2004).
12. P. Biswas, Large deflection of heated orthotropic plates, Indian Pure and Applied Mathematics, 9, pp. 1027–1032 (1978).
13. P. Biswas, Nonlinear analysis of heated orthotropic plates, Indian Pure and Applied Mathematics, 12, pp. 1380–1389 (1981).
14. J.S. Burdess, A.J. Harris, D. Wood, and J.L. Cruickshank, The dynamics of a vibrating silicon diaphragm micro gyroscope, Proceedings of the Institution of Mechanical Engineers, Part C, Journal of Mechanical Engineering Science, 214(11), pp. 1379–1388 (2000).
15. L.F. Che, B. Xiong, X.Z. Huang, and Y.L. Wang, Simulation of vibratory micromachined gyroscope with fence structure, International Journal of Nonlinear Sciences and Numerical Simulation, 3(3-4), pp. 333–336 (2002).
16. L.F. Che, B. Xiong, and Y.L. Wang, System-level simulation of vibratory micromachined gyroscope with fence structure, Acta Mechanica Sinica, 20(2), pp. 172–177 (2004).
17. C.Y. Chia, Nonlinear Analysis of Plates, McGraw Hill, New York, 1980.
18. W.O. Davis, O.M. O'Reilly, and A.P. Pisano, On the nonlinear dynamics of tether suspensions for MEMS, Journal of Vibration and Acoustics, Transactions of the ASME, 126(3), pp. 326–331 (2004).
19. M. Eissa, Dynamics and resonance of a nonlinear mechanical oscillator subjected to parametric and external excitation, Integral Methods in Science and Engineering, Chapman & Hall/CRC, 2000.

20. I. Glesk, R.J. Runser, and P.R. Prucnal, New generation of devices for all-optical communications, Acta Physica Slovaca, 51(2), pp. 151–162 (2001).
21. K. Grosh, Special Issue on MEMS modeling in dynamics and acoustics—Foreword, Journal of Vibration and Acoustics, Transactions of the ASME, 126(3), pp. 325–325 (2004).
22. X. He, A theoretical framework for the dynamics response of electronic packaging under mechanical and thermal loading, Ph.D. thesis, Georgia Institute of Technology, Atlanta, 2000.
23. X. He and R. Fulton, Nonlinear dynamic analysis of a printed wiring board, Journal of Electronic Packaging, 124(2), pp. 77–84 (2002).
24. X. He and M. Stallybrass, Impact response of a printed wiring board, International J. Solids and Structures, 39(24), pp. 5979–5990 (2002).
25. X. He, Nonlinear Dynamic response of a thin laminate subject to non-uniform thermal field, Int. J. Non-linear Mechanics, 41(1), pp. 43–56 (2006).
26. C.M. Ho and Y.C. Tai, Review: MEMS and its applications for flow control, Journal of Fluids Engineering, Transactions of the ASME, 118(3), pp. 437–447 (1996).
27. F.C. Hoppensteadt and E.M. Izhikevich, Synchronization of MEMS resonators and mechanical neuron computing, IEEE Transactions on Circuits and Systems, I Fundamental Theory and Applications, 48(2), pp. 133–138 (2001).
28. T.K. Hsiai, S.K. Cho, P.K. Wong, M.H. Ing, A. Salazar, S. Hama, M. Navab, L.L. Demer, and H. Chih-Ming, Micro sensors: linking real-time oscillatory shear stress with vascular inflammatory responses, Annals of Biomedical Engineering, 32(2), pp. 189–201 (2004).
29. J. Hyvarinen and J. Soderkvist, Aeroelasticity—the interplay between fluids and solids, Journal of Micromechanics and Microengineering, 11(4), pp. 416–422 (2001).
30. R. Jones, Mechanics of Composite Materials, 2nd edition, Taylor & Francis, PA, 1999.
31. S.K. Koh, J.P. Ostrowski, and G.K. Ananthasuresh, Control of micro-satellite orientation using bounded-input, fully-reversed MEMS actuators, International Journal of Robotics Research, 21(5-6), pp. 591–605 (2002).
32. W.Z. Lin, K.H. Lee, S.P. Lim, and P. Lu, A model reduction method for the dynamic analysis of micro-electromechanical systems, International Journal of Nonlinear Sciences and Numerical Simulation, 2(2), pp. 89–100 (2001).
33. S. Liu, A. Davidson, and Q. Lin, Simulation studies on nonlinear dynamics and chaos in a MEMS cantilever control system, Journal of Micromechanics and Microengineering, 14(7), pp. 1064–1073 (2004).
34. A.C.J. Luo and F.Y. Wang, Nonlinear dynamics of a micro-electro-mechanical system with time-varying capacitors, Journal of Vibration and Acoustics, Transactions of the ASME, 126(1), pp. 77–83 (2004).
35. S.G. Mallinson, C.Y. Kwok, and J.A. Reizes, Numerical simulation of micro-fabricated zero mass-flux jet actuators, Sensors and Actuators, A Physical, 105(3), pp. 229–236 (2003).
36. A.W. McFarland and J.S. Colton, Production and analysis of injection molded micro-optic components, Polymer Engineering and Science, 44(3), pp. 564–579 (2004).
37. R.T. M'Closkey, S. Gibson, and J. Hui, System identification of a MEMS gyroscope, Journal of Dynamic Systems Measurement and Control, Transactions of the ASME, 123(2), pp. 201–210 (2001).
38. T. Mengden, B. Binswanger, T. Spuhler, B. Weisser, and W. Vetter, The use of self-Measured blood-pressure determinations in assessing dynamics of drug compliance in a study with amlodipine once-a-day, morning versus evening, Journal of Hypertension, 11(12), pp. 1403–1411 (1993).
39. M.C. Pal, Large deflection of heated circular plates, Acta Mechanica, 8, pp. 82–103 (1969).
40. M.C. Pal, Large amplitude free vibration of rectangular plates subjected to aerodynamic heating, Journal of Engineering Mathematics, 4(1), pp. 39–49 (1970).
41. M.C. Pal, Static and dynamic nonlinear behavior of heated orthotropic circular plates, International J. Non-linear Mechanics, 8, pp. 489–504 (1973).
42. F.X. Pan, J. Kubby, and J.K. Chen, Numerical simulation of fluid-structure interaction in a MEMS diaphragm drop ejector, Journal of Micromechanics and Microengineering, 12(1), pp. 70–76 (2002).
43. J.H. Park and S.W. Baek, Investigation of influence of thermal accommodation on oscillating micro-flow, International Journal of Heat and Mass Transfer, 47(6-7), pp. 1313–1323 (2004).
44. R.E. Rudd, The atomic limit of finite element modeling in MEMS: coupling of length scales, Analog Integrated Circuits and Signal Processing, 29(1-2), pp. 17–26 (2001).
45. S. Santoli, A "lab-on-chip" design for miniature autonomous bio/chemoprospecting planetary rovers, JBIS, Journal of the British Interplanetary Society, 55(3-4), pp. 115–130 (2002).
46. G. Sheng, B. Liu, W. Hua, Y.B. Miao, B. Xu, L. Yan, and U. Sridhar, Design and analysis of MEMS-based slider suspensions for a high-performance magnetic recording system, Journal of Micromechanics and Microengineering, 10(1), pp. 64–71 (2000).

47. F. Shi, P. Ramesh, and S. Mukherjee, Dynamic analysis of micro-electro-mechanical systems, International Journal for Numerical Methods in Engineering, 39(24), pp. 4119–4139 (1996).
48. E. Suhir, Response of a flexible printed circuit board to periodic shock loads applied to its support contour, J. of Applied Mechanics, 59(2) (1992).
49. E. Suhir, Failure criterion for moisture-sensitive plastic packages of integrated circuits (IC) devices: application of Von Karman's equation with consideration of thermoelastic strains, International Journal of Solids and Structures, 114, pp. 368–377 (1996).
50. Q.H. Sun and I.D. Boyd, A direct simulation method for subsonic, microscale gas flows, Journal of Computational Physics, 179(2), pp. 400–425 (2002).
51. K. Turner, R. Baskaran, and W.H. Zhang, Micro-scale sensors and filters utilizing non-linear dynamic response of single and coupled oscillators, Multidisciplinary Research in Control, pp. 101–112 (2003).
52. W.W. Van Arsdell and S.B. Brown, Subcritical crack growth in silicon MEMS, Journal of Microelectromechanical Systems, 8(3), pp. 319–327 (1999).
53. J. Whitney and A. Leissa, Analysis of heterogeneous anisotropic plates, Journal of Applied Mechanics, (June), pp. 261–266 (1969).
54. W.C. Xie, H.P. Lee, and S.P. Lim, Nonlinear dynamic analysis of MEMS switches by nonlinear modal analysis, Nonlinear Dynamics, 31(3), pp. 243–256 (2003).
55. J.P. Yang, J. Chai, and Y. Lu, Dynamics simulation of MEMS device embedded hard disk drive systems, Microsystem Technologies, Micro- and Nanosystems-Information Storage and Processing Systems, 10(2), pp. 115–120 (2004).
56. Y.J. Yang, S.Y. Cheng, and Shen, K.Y., Macromodeling of coupled-domain MEMS devices with electrostatic and electrothermal effects, Journal of Micromechanics and Microengineering, 14(8), pp. 1190–1196 (2004).

20

Effect of Material's Nonlinearity on the Mechanical Response of some Piezoelectric and Photonic Systems

Victor Birman[a] and Ephraim Suhir[b]

[a] *University of Missouri-Rolla, St. Louis, Missouri 63121, USA*
[b] *University of California, Santa Cruz, CA, University of Maryland, College Park, MD, and ERS/Siloptix Co., Los Altos, CA, USA*

Abstract Nonlinear stress–strain relationship (physical or material's nonlinearity) may have a substantial effect on the structural response of piezo-electric, as well as on micro- and opto-electronic assemblies, packages and systems, including stresses, deformations, stability, and vibrations. The chapter addresses, as illustrations, two reprehensive examples, where material's nonlinearity significantly affects the behavior of a piezoelectric or a photonic (fiber-optic) structure. The first example has to do with vibrations of piezoelectric rods driven by an alternating electric field. The second example deals with silica optical fibers, in which mechanical test data, strength, buckling and vibration phenomena are affected, to a greater or lesser extent, by physical nonlinearity of the material (silica). In both cases, it is shown that the non-linear stress–strain relationship of the material cannot be neglected without causing a significant error in the predicted mechanical behavior of the material.

20.1. INTRODUCTION

Linear mathematical models are extensively employed in engineering and applied science to design structures and model their response to static or dynamic loading. This is due to a number of factors. In particular, the analysis of nonlinear equations describing the behavior of a typical structure requires substantial computational power that was not available until the last decades. This problem was dealt with using simplified models that often overlooked essential nonlinear effects and sacrificed the accuracy of the prediction. A noticeable exception is the famous "elastica" problem solved by Leonard Euler in 1744 ("Methodus inveniendi lineas curves maximi minimive proprietate gaudentes"). In spite of the fact that ideal linear systems do not exist in nature, solutions neglecting nonlinear effects have often been successfully applied: practically all systems designed until the introduction of computers and the majority of systems designed today rely on linear models. This is because while nonlinearity is always present, is often relatively small and

its effect on the solution is either negligible or can be effectively incorporated in the design by introducing some margins of safety. In addition, linear solutions are often conservative and designers like them for this "hidden" supplementary safety margin. Nowadays, various nonlinear problems can be accurately solved using modern computational techniques, but the ability to obtain analytical (closed form) solutions is still very important, because of their compactness and a clear indication on "what affects what and what is responsible for what."

There are two major classes of nonlinearity that are often treated separately. The first is nonlinearity related to large deformations of the system, although the material remains within its linearly elastic state of stress. This is a "geometrical nonlinearity" that manifests itself as a nonlinear relationship between the applied loading and the displacement. The second type of nonlinearity is associated with the properties of the material and is reflected in nonlinear constitutive relations (stress–strain law). This is "physical nonlinearity" (or "material's nonlinearity"). A typical example is plastic response of the material when the stresses exceed the yield limit, or the nonlinearly elastic behavior of the material under a relatively high level of loading (deformation).

In the present chapter, we address the effects of the physical nonlinearity in the behavior of piezo-electric and fiber-optic structures. The examples discussed show the effect of the physical nonlinearity on the response of piezoelectric rods driven by an alternating electric field and on the static and dynamic response of optical fibers. In both examples, neglecting physically nonlinear effects can result in a significant and unacceptable error.

20.2. EFFECT OF PHYSICAL NONLINEARITY ON VIBRATIONS OF PIEZOELECTRIC RODS DRIVEN BY ALTERNATING ELECTRIC FIELD

This problem is important in connection with the design of piezoelectric transducers. Piezoelectricity, discovered by Jacques and Pierre Curie in 1880, is a phenomenon reflecting coupling between mechanical and electrical phenomena. When the piezoelectric material is mechanically stressed, electric field is generated (direct piezoelectric effect). On the other hand, if such material is subject to an electric field it experiences deformations (inverse piezoelectric effect). These properties of piezoelectric materials made their application useful as sensors (electric measurements of mechanical deformations) and as actuators (electrically induced static or dynamics deformations of structures). Notably, piezoelectric properties of the material are realized if the material is "poled," i.e., is subject to a short-time high electric field resulting in its polarization in the appropriate direction.

The well known physically linear relationships between the tensors of stress and strain and the vectors of electric field and electric displacement for piezoelectric materials are applicable to a particular case where nonlinear effects are negligible. While geometric nonlinearity can be partially incorporated into these equations by an appropriate choice of the strain-displacement relationships, neither this nonlinearity nor physical nonlinearity of the material are fully reflected in the linear version of the constitutive equations. A general form of constitutive equations accounting for nonlinear products of the components of the strain tensor (geometrically nonlinear effect) and physically nonlinear terms has been derived and published (see, for example, [1]). However, the complexity of these equations as well as the lack of experimental data on the coefficients at the nonlinear terms and the difficulty involved in their evaluation prevented a wide acceptance of nonlinear constitutive equations in design and practical applications. This makes it important to elucidate a

relative contribution of geometrically and physically nonlinear terms and to assess the necessity of their incorporation in the analysis and their relative qualitative and quantitative effects on the solution.

Early studies of the effects of stress and electric fields on the response of piezoelectric materials were presented in a number of papers [2–10]. A nonlinear nature of the problem is evident in these investigations. The theoretical formulation was also developed by Tiersten [11] who later applied it to the problems of thin and membrane piezoelectric plates subjected to high electric fields [12,13]. Other derivations of physically nonlinear equations were published by Nakagawa et al. [14] and Cho and Yamanouchi [15].

A physically nonlinear behavior of piezoelectric materials is also reflected in the hysteresis or butterfly loops in the electric field–strain or strain–stress planes. Such loops and relevant nonlinear phenomena were described and discussed by a number of authors [16–20].

Beige and Schmidt [21] and Beige [22] included higher-order electric and elastic terms in their studies of longitudinal vibrations of a plate with a 31 piezoelectric effect. This effect has also been studied by von Wagner and Hagedorn [23] for piezoelectric beams. Other studies of von Wagner [24,25] and von Wagner and Hagedorn [26] were concerned with physical nonlinearity in 33-piezoelectrics. The work of Chattopadhyay et al. [27] employed nonlinear constitutive equations incorporating a cubic nonlinearity for the electric field to analyze helicopter blades (modeled by composite box beams). The subsequent work accounted for transverse shear deformability of monocoque and sandwich plates combined with the nonlinear piezoelectric effect of embedded or mounted sensors and actuators [28].

Among recent studies that attempt to implicitly account for nonlinearity by using variable coefficients in linear constitutive relationships one can mention the paper by Sherrit et al. [29] where the piezoelectric coefficient d_{33} of lead zirconate titanate ceramics was shown dependent on stress, temperature and frequency. Experimental data elucidating nonlinear phenomena was presented in the papers of Wiederick et al. [30] and Sherrit et al. [29,31] that showed that both the piezoelectric constants as well as the permittivity are nonlinear functions of the applied electric field. It was also observed that piezoelectric coefficients increase almost linearly with the stress.

As was shown in [32–34], piezoelectric coefficients are nonlinear functions of the applied compressive stress. While piezoelectric coefficients d_{31}, d_{33}, d_{15} typically increased with the electric field, particularly in soft piezoceramics, an increase in the frequency resulted in a very small decrease of these coefficients.

Wang and Carman [35] showed in their experiments that both the coefficient of thermal expansion as well as the piezoelectric coefficient d_{31} are affected by a cryogenic temperature. Moreover, it was shown that d_{31} is affected by the magnitude of strain. In addition, the electric field was recently shown to affect the elastic modulus [36]. Barrett [37] presented a nonlinear relationship between the strain in the piezoelectric PZT-5H actuator and the applied electric field.

Bert and Birman [38] proved that both the stress and temperature affect the coefficient d_{31} in a one-dimensional problem. In addition, they characterized the variations of the coefficient of thermal expansion with the stress and electric field. This work was further expanded to two-dimensional and three-dimensional cases [39].

Joshi derived physically nonlinear constitutive equations for piezoceramics by the assumption that material constants are independent of the magnitude of stress or electric field [40]. This solution was obtained using the thermodynamic Gibbs potential and retaining the second-order terms in the total differentials of dependent variables (strains, electric

flux density, and entropy). The constitutive equations presented in [40] are similar to those published by Maugin et al. [1].

A nonlinear relationship between the deflection of the tip of a bimorph working in the 31 mode and the applied electric field was observed for various piezoelectric actuators by Wang et al. [41] who attributed this nonlinearity to an increase of d_{31} with the applied electric field. The physical nonlinearity effect was reported in this paper, even though the electric field was relatively low (150 V/mm). The effect of physical nonlinearity on shape control of composite laminated beams was considered by Achuthan et al. [42] who illustrated that the voltage required to control the shape of the beam can be significantly reduced when physical nonlinearity of piezoelectric patches on the beam surface is accounted for.

Notably, the electrostrictive response can include higher-order nonlinear terms, in addition to the quadratic relationship between the stress and the electric field [43]. However, higher-order nonlinear contributions become essential only in the case of very high applied fields. Explicit expressions for piezoelectric coefficients d_{31}, d_{33} as functions of the peak-to-peak voltage amplitude were obtained for a class of materials by Williams [44]. These relationships involve quadratic power of the peak-to-peak voltage.

Two examples of representative simplified empirical equations incorporating physical nonlinearity are shown below. For example, according to Barrett [45], the response of G-1195 piezoceramics under the electric field reaching 600 V/mm can be represented by the following relationship between the microstrain and the electric field (31 effect):

$$\varepsilon = \left(0.227E_z + 0.000243E_z^2\right) \times 10^{-6}. \tag{20.1}$$

Priya et al. [46] found that the relationship between the coefficient d_{31} and the squared applied elastic strain is linear. In particular, for soft PZT-5A, this relationship was obtained in the form

$$d_{31} = 2.03 \times 10^{-10} + 1.12 \times 10^{-4}\varepsilon^2, \tag{20.2}$$

where the squared strain varied from zero to 2.8×10^{-7}.

The present chapter concentrates on the investigation of physically nonlinear effects on the behavior of piezoelectric rods polarized in the axial direction. Such problems are important in piezoelectric transducers used in underwater hydrophones, acoustic imaging and medical applications. Typically, the behavior of piezoelectric rods and so-called 1–3 piezocomposites consisting of rods embedded in the matrix are studied using linear constitutive equations [47–50]. Recent papers by Tan and Tong [51,52] extended the study to the physically nonlinear static formulation. Physically nonlinear dynamic problems were also considered by Wagner [24,25] and Wagner and Hagedorn [26]. A related problem of linear vibrations of thin piezoceramic disks accounting for coupled axial, tangential and radial modes was analyzed by Huang et al. [53]. A unified formulation presented below enables us to identify essential terms in physically nonlinear constitutive equations and conduct a comprehensive analysis of the problem.

20.2.1. Physically Nonlinear Constitutive Relationships for an Orthotropic Cylindrical Piezoelectric Rod Subject to an Electric Field in the Axial Direction

Consider a cylindrical rod shown in Figure 20.1 subject to an electric field in the z-direction. In the following analysis, the cylindrical coordinate axes z, r, θ identified in

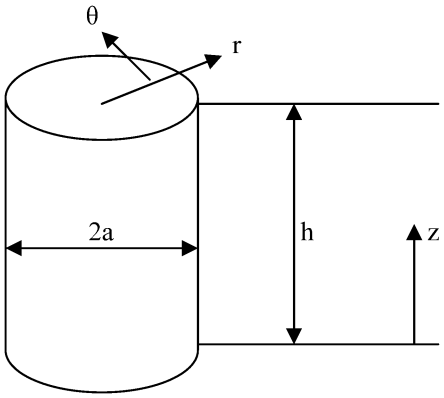


FIGURE 20.1. Cylindrical piezoelectric rod and the coordinate system adopted in the analysis.

Figure 20.1 are also denoted as 3, 1 and 2, respectively. The problem being axisymmetric, the circumferential displacement is equal to zero, while the axial and radial displacements are denoted by w and u, respectively.

The rod subject to an alternating electric field $E_z = E_3$ experiences vibrations in the axial and radial directions. Shearing stresses and strains are equal to zero since the motion is axisymmetric. In the subsequent analysis the rod is assumed "anchored" at the plane $z = 0$, preventing axial displacements at this location.

The comprehensive formulation of the problem, including field equations, boundary conditions and constitutive equations could be derived using the Hamilton principle and the electric enthalpy density. An example of such approach is found in [25] where the electric enthalpy density for a one-dimensional problem was represented in terms of strain and electric field, including nonlinear terms of the third and fourth order. In the case where the third-order strain dependent terms and the fourth-order nonlinear terms are negligible, the electric enthalpy density is reduced to the energy density formulated by Maugin et al. [1]. The following equation presents this energy density for the general three-dimensional problem:

$$\Psi = \frac{1}{2}C_{\alpha\beta}\varepsilon_\alpha\varepsilon_\beta - e_{m\alpha}E_m\varepsilon_\alpha - \frac{1}{2}e_{m\alpha\beta}E_m\varepsilon_\alpha\varepsilon_\beta - \frac{1}{2}\varepsilon_{mn}E_mE_n - \frac{1}{6}\varepsilon_{mnp}E_mE_nE_p - \frac{1}{2}l_{mn\alpha}E_mE_n\varepsilon_\alpha, \tag{20.3}$$

where $C_{\alpha\beta}$ are elastic stiffness constants, ε_α are strains, $e_{m\alpha}$ are piezoelectric constants, E_m are components of the electric field in the corresponding direction, $e_{m\alpha\beta}$ are electroelastic constants, ε_{mn} and ε_{mnp} are dielectric coefficients (permittivity and third-order dielectric coefficients, respectively), and $l_{mn\alpha}$ are electrostrictive coefficients.

It is instructive to consider the energy density for a one-dimensional case where both the electric field and strain are considered in the axial (z or 3 directions):

$$\Psi' = \frac{1}{2}C_{33}\varepsilon_z^2 - e_{33}E_3\varepsilon_z - \frac{1}{2}e_{333}E_3\varepsilon_z^2 - \frac{1}{2}\varepsilon_{33}E_3^2 - \frac{1}{6}\varepsilon_{333}E_3^3 - \frac{1}{2}l_{333}E_3^2\varepsilon_z + f\left(\varepsilon_z^3, \varepsilon_z^4, E_3^3\varepsilon_z, E_3^2\varepsilon_z^2, E_3\varepsilon_z^3, E_3^4\right). \tag{20.4}$$

The last term in the right side of (20.4) refers to higher-order nonlinear contributions incorporated in [25]. A relative contribution of these terms is expected to become prominent in a highly nonlinear formulation involving large values of the strain and electric field. Note that geometrically nonlinear effects can be noticeable in piezoelectric materials at stresses exceeding 8 MPa (soft PIC 255) and 20 MPa (hard PC4D) as was reported in a recent paper of Guillon et al. [54]. However, such level of stress corresponds to significant strains exceeding 2.5%. Accordingly, all higher-order nonlinear contributions in the last term in (20.4) and the corresponding nonlinear contributions in the three-dimensional case that could affect the Equation (20.3) are omitted in the following analysis. This simplification limits the present solution to the case of small strains and moderate electric fields.

The constitutive relations yielding stresses and electric displacements are obtained from

$$\sigma_\alpha = \frac{\partial \Psi}{\partial \varepsilon_\alpha}, \quad D_m = -\frac{\partial \Psi}{\partial E_m}. \tag{20.5}$$

This results in the following expression for the components of the stress tensor:

$$\sigma_\alpha = \underline{C_{\alpha\beta}\varepsilon_\beta - e_{m\alpha}E_m} - e_{m\alpha\beta}E_m\varepsilon_\beta - \frac{1}{2}l_{mn\alpha}E_mE_n, \tag{20.6}$$

$$D_m = \underline{e_{m\alpha}\varepsilon_\alpha} + \frac{1}{2}e_{m\alpha\beta}\varepsilon_\alpha\varepsilon_\beta + \underline{\varepsilon_{mn}E_n} + \frac{1}{2}\varepsilon_{mnp}E_nE_p + l_{mn\alpha}E_n\varepsilon_\alpha. \tag{20.7}$$

The terms underlined in the right side of (20.6) and (20.7) are retained in the linear formulation.

In the problem considered in this chapter, we are concerned with the dynamic response of the rod to the applied field in the axial direction, i.e., $E_3(t) = E_z(t)$. Accordingly, the nonzero elements of the tensor of stress given by (20.6) are

$$\begin{Bmatrix} \sigma_r \\ \sigma_\theta \\ \sigma_z \end{Bmatrix} = \begin{bmatrix} C_{11} - e_{311}E_3 & C_{12} - e_{312}E_3 & C_{13} - e_{313}E_3 \\ C_{12} - e_{321}E_3 & C_{22} - e_{322}E_3 & C_{23} - e_{323}E_3 \\ C_{13} - e_{331}E_3 & C_{23} - e_{332}E_3 & C_{33} - e_{333}E_3 \end{bmatrix} \begin{Bmatrix} \varepsilon_r \\ \varepsilon_\theta \\ \varepsilon_z \end{Bmatrix} - \begin{Bmatrix} e_{31} \\ e_{32} \\ e_{33} \end{Bmatrix} E_3 - \frac{1}{2}\begin{Bmatrix} l_{331} \\ l_{332} \\ l_{333} \end{Bmatrix} E_3^2. \tag{20.8}$$

Equations (20.8) are in agreement with the expressions for the stress published by Joshi neglecting nonlinear stiffness coefficients [40]. Tiersten [13] included the same terms but omitted the products of strain and electric field.

In conclusion of this paragraph it is useful to account for the effect of temperature on the constitutive relations of piezoelectric rods experiencing forced vibrations. The corresponding extension of Equations (20.3), (20.6) and (20.7) is shown below by assumption that temperature is constant. Accordingly, the terms that depend on temperature only are excluded from the expression for the electric enthalpy. The third-order strain term and its product with temperature are omitted following the considerations used to define the energy density in (20.3). The fourth-order nonlinear terms dependent on strain and temperature are also excluded following the approach employed to generate (20.3). The effect of temperature is reflected by incorporating terms linearly dependent on it. In addition, two terms

dependent on the second power of temperature are added to the energy density, these terms are proportional to the elements of the tensors of strain and electric field. In this case the corresponding equations are:

$$\begin{aligned}\Psi = {} & \frac{1}{2}C_{\alpha\beta}\varepsilon_\alpha\varepsilon_\beta - e_{m\alpha}E_m\varepsilon_\alpha - \frac{1}{2}e_{m\alpha\beta}E_m\varepsilon_\alpha\varepsilon_\beta - \frac{1}{2}\varepsilon_{mn}E_mE_n - \frac{1}{6}\varepsilon_{mnp}E_mE_nE_p \\ & - \frac{1}{2}l_{mn\alpha}E_mE_n\varepsilon_\alpha - \lambda_\alpha\varepsilon_\alpha T - \delta_\alpha\varepsilon_\alpha T^2 - p_mE_mT - \frac{1}{2}\rho_{\alpha\beta}\varepsilon_\alpha\varepsilon_\beta T - \eta_mE_mT^2 \\ & - \frac{1}{2}\chi_{mn}E_mE_nT - \kappa_{ma}E_m\varepsilon_aT - \frac{1}{2}\mu_{m\alpha\beta}E_m\varepsilon_\alpha\varepsilon_\beta T - \frac{1}{2}\phi_{mn\alpha}E_mE_n\varepsilon_\alpha T \\ & - \frac{1}{6}\varsigma_{mnp}E_mE_nE_pT, \end{aligned} \tag{20.9}$$

$$\begin{aligned}\sigma_\alpha = {} & \underline{C_{\alpha\beta}\varepsilon_\beta - e_{m\alpha}E_m} - e_{m\alpha\beta}E_m\varepsilon_\beta - \frac{1}{2}l_{mn\alpha}E_mE_n \\ & - \underline{\underline{\lambda_\alpha T - \delta_\alpha T^2 - \rho_{\alpha\beta}\varepsilon_\beta T - \kappa_{m\alpha}E_mT}} - \underline{\underline{\mu_{m\alpha\beta}E_m\varepsilon_\beta T - \frac{1}{2}\phi_{mn\alpha}E_mE_nT}}, \end{aligned} \tag{20.10}$$

$$\begin{aligned}D_m = {} & \underline{e_{m\alpha}\varepsilon_\alpha} + \frac{1}{2}e_{m\alpha\beta}\varepsilon_\alpha\varepsilon_\beta + \underline{\varepsilon_{mn}E_n} + \frac{1}{2}\varepsilon_{mnp}E_nE_p + l_{mn\alpha}E_n\varepsilon_\alpha \\ & + \underline{\underline{p_mT + \eta_mT^2 + \chi_{mn}E_nT + \kappa_{m\alpha}\varepsilon_\alpha T}} \\ & + \underline{\underline{\frac{1}{2}\mu_{m\alpha\beta}\varepsilon_\alpha\varepsilon_\beta T + \phi_{mn\alpha}E_n\varepsilon_\alpha T + \frac{1}{2}\varsigma_{mnp}E_nE_pT}}. \end{aligned} \tag{20.11}$$

In these equations, T is temperature in excess of the reference value; λ_α are thermoelastic coefficients; δ_α are second order thermoelastic coefficients; p_m are pyroelectric coefficients; $\rho_{\alpha\beta}$, η_m, χ_{mn}, κ_{ma}, $\mu_{m\alpha\beta}$, $\phi_{mn\alpha}$, ς_{mnp} are other coefficients accounting for the interaction of the tensor of strains, vector of electric field and temperature. The terms that are double-underlined in (20.10) and (20.11) illustrate the linear thermally-dependent contributions. Some of these terms are static and do not affect the solution of the dynamic problem. As in the formulation without thermal effects, cubic terms accounting for the nonlinear strain contribution are omitted in the formulation presented by (20.9)–(20.11).

20.2.2. Analysis of Uncoupled Axial Vibrations

The problem of uncoupled axial vibrations is considered by assumption that the effect of radial vibrations on the axial motion can be disregarded. The Generalized Galerkin procedure employed in the analysis enables us to satisfy both kinematic and static boundary conditions. Although this procedure is relatively little known, contrary to the standard Galerkin procedure, the solutions utilizing it have been published [55,56]. However, the application of the Generalized Galerkin procedure to the present problem considered below is original.

The boundary conditions employed in the chapter are that the cross section $z = 0$ is prevented from axial motion, i.e., the axial displacement $w(z = 0) = 0$. The axial stresses at the free end of the rod should be equal to zero, i.e., $\sigma_z(z = h) = 0$.

In the case where it is impossible to satisfy the stress boundary conditions, following the Generalized Galerkin procedure, the displacements can be sought in the form

$$w(z, t) = \sum_i Z_i(z) W_i(t), \tag{20.12}$$

and the system of equations of motion can be obtained in the form:

$$\int_0^h (\sigma_{z,z} - m\ddot{w}) Z_i(z) dz - \sigma_z(z = h) Z_i(h) = 0, \tag{20.13}$$

where m is the mass density of the rod material.

The motion considered in this problem is represented in terms of normal modes of axial vibrations of the rod without the piezoelectric effect. The axial free vibrations of such rod with the boundary conditions specified above are represented in the form

$$w(x, t) = \sum_i (A_{1i} \sin \lambda_i t + A_{2i} \cos \lambda_i t) \sin \frac{i\pi z}{2h}, \tag{20.14}$$

where i is a natural odd number, A_{ki} are constants of integration specified from the initial conditions, and

$$\lambda_i = \frac{i\pi}{2h} \sqrt{\frac{C_{33}}{m}}$$

is the natural frequency.

Following the Galerkin procedure, the axial displacements should be chosen in the form of series satisfying all boundary conditions. However, it appears impossible to specify such series satisfying the conditions both at the "anchored" cross section $z = 0$ as well as at the free end of the rod $z = h$. Therefore, the solution is sought in the form (i are odd numbers)

$$w = \sum_i W_i(t) h \sin \frac{i\pi z}{2h}. \tag{20.15}$$

The substitution of (20.15) into (20.13) where $Z_i = \sin(i\pi z/2h)$ and using $\varepsilon_z = w_{,z}$ yields the system of uncoupled equations:

$$\frac{mh^2}{2} \ddot{W}_i + (C_{33} - e_{333} E_z) \frac{i^2\pi^2}{8} W_i = \left(e_{33} E_z + \frac{1}{2} l_{333} E_z^2 \right) \sin \frac{i\pi}{2}. \tag{20.16}$$

Note that the same system of equation of motions could be derived using the Lagrange equation, i.e.,

$$\frac{\partial}{\partial t} \left(\frac{\partial T}{\partial \dot{W}_i} \right) - \frac{\partial T}{\partial W_i} + \frac{\partial V}{\partial W_i} = 0, \tag{20.17}$$

where the energy contributions are defined by

$$T = \frac{1}{2}\int_0^{2\pi}\int_0^h\int_0^a m\dot{w}^2 r dr dz d\theta, \quad V = \int_0^{2\pi}\int_0^h\int_0^a \Psi' r dr dz d\theta. \tag{20.18}$$

Equations (20.16) can be integrated if the electric field is a known function of time. In the case of an alternating and harmonic electric field, i.e., $E_z(t) = E\cos\omega t$, the equation for the i-th mode becomes

$$\frac{mh^2}{2}\ddot{W}_i + (C_{33} - e_{333}E\cos\omega t)\frac{i^2\pi^2}{8}W_i = \left(\frac{1}{4}l_{333}E^2 + e_{33}E\cos\omega t + \frac{1}{4}l_{333}E^2\cos 2\omega t\right)\sin\frac{i\pi}{2}. \tag{20.19}$$

Equation (20.19) is a nonhomogeneous Mathieu-Hill equation. Similar equations were first employed in the first part of the last century when studies were often limited to homogeneous equations characterizing dynamic or parametric stability of structures [57]. A typical problem where the analysis is reduced to a nonhomogeneous Mathieu equation is the motion of a rod with initial imperfections subject to a periodic in time axial force. A related problem is forced vibrations of a rod subject to an eccentrically applied driving force. In addition to the classical monograph of Bolotin [57], a number of investigations have been concerned with the problem of interaction between forced and parametric vibrations [58–64].

The analytical solution of Equation (20.19) can be sought in the form of trigonometric time series:

$$W_i = A_{0i} + \sum_r A_{ri}\cos\frac{r\omega t}{2} + \sum_p B_{pi}\sin\frac{p\omega t}{2}, \tag{20.20}$$

where A_{0i}, A_{ri} and B_{pi} are unknown coefficients. These coefficients can be determined by substituting the series (20.20) into (20.19), equating the coefficients at the same trigonometric functions of time and solving the system of resulting linear algebraic equations with respect to A_{0i}, A_{ri} and B_{pi}.

The equation representing the motion in terms of the fundamental (first) normal mode is

$$\frac{mh^2}{2}\ddot{W}_1 + (C_{33} - e_{333}E\cos\omega t)\frac{\pi^2}{8}W_1 = \frac{1}{4}l_{333}E^2 + e_{33}E\cos\omega t + \frac{1}{4}l_{333}E^2\cos 2\omega t. \tag{20.21}$$

As is shown below, single-term solutions that result in Equation (20.21) are accurate if the driving frequency ω is smaller than the fundamental frequency of the rod.

Equations (20.19) or (20.21) could be simplified for numerous piezoelectric materials. Consider for example, Equation (20.21) where it is instructive to compare the magnitude of the following terms:

$$k_1 = e_{333}EW_1, \quad k_2 = \frac{1}{4}l_{333}E^2, \quad k_3 = e_{33}E. \tag{20.22}$$

TABLE 20.1.
Material constant of representative piezoelectrics. Reproduced with permission of ASME.

Material	e_{33} (C/m^2)	e_{333} (C/m^2)	C_{33} (GPa)	l_{333} (F/m)
PZT-5H	24.54 (or 31.02)	5.7×10^4	108.0	-21.21×10^{-6}
PZN-4.5%PT	12.1 (or 13.02)	4.2×10^4	89.0	-1.61×10^{-6}
$LiNbO_3$	1.3	-17.3	24.5	-2.76×10^{-9}

TABLE 20.2.
Coefficients of equations of motion for representative piezoelectrics ($E = 2.0$ MV/m). Reproduced with permission of ASME.

Material	k_1	k_2	k_3
PZT-5H	$11.4\times10^{10}W_1$	-2.12×10^7	5.09×10^7
PZN-4.5%PT	$8.9\times10^{10}W_1$	-1.61×10^6	2.42×10^7
$LiNbO_3$	-34.6×10^6W_1	-2.76×10^3	2.6×10^6

Note that a physically linear formulation can be obtained by setting $k_1 = k_2 = 0$.

Two materials chosen for the following comparison and considered in numerical examples are PZT-5H and PZN-4.5%PT [51,52]. In addition, $LiNbO_3$ [1] is used in the analysis. The amplitude of typical electric fields is usually limited to 2.0 MV/m [1]. The values of material constants needed to evaluate the coefficients in (20.22) are listed in Table 20.1. Using a high electric field, i.e., 2.0 MV/m, since it results in a stronger nonlinear effect, one obtains the values of the coefficients in (20.22) listed in Table 20.2.

Note that W_1 in Table 20.2 is nondimensional and the units of all terms are C*V/m^3. The comparison of the coefficients in Table 20.2 yields the conclusion that the terms proportional to the squared electric field, i.e., k_2, can be neglected if the electric field remains within certain limits. In particular, these coefficients are negligible for PZN-4.5%PT and $LiNbO_3$ even at 2.0 MV/m, while the term proportional to k_2 can be neglected for PZT-5H if the field remains smaller than 0.5 MV/m since in this case k_2 is an order of magnitude higher than k_3. Accordingly, if the electric field is below the limits specified above, Equations (20.19) and (20.21) can be simplified. In particular, the latter equation is reduced to a nonhomogeneous Mathieu equation:

$$\ddot{W}_1 + \lambda^2 W_1 - 2\mu\lambda^2 W_1 \cos\omega t = p\cos\omega t, \tag{20.23}$$

where the fundamental frequency, the parametric loading coefficient and the amplitude of the forcing function are

$$\lambda = \frac{\pi}{2h}\sqrt{\frac{C_{33}}{m}}, \quad \mu = \frac{e_{333}E}{2C_{33}}, \quad p = \frac{2e_{33}E}{mh^2}. \tag{20.24}$$

As will be shown below, viscous damping is essential in the case where the driving frequency is close to the fundamental frequency of the rod. The corresponding expansion of (20.23) is

$$\ddot{W}_1 + \beta\dot{W}_1 + \lambda^2 W_1 - 2\mu\lambda^2 W_1 \cos\omega t = p\cos\omega t, \tag{20.25}$$

where $\beta = 2c/(mh^2)$, c being a damping coefficient. The value of the damping coefficient can be evaluated from published data for a quality factor Q as $c = c_{cr}/(2Q)$ where c_{cr} is a critical damping coefficient of the rod experiencing axial vibrations.

Equations (20.23) or (20.25) are similar to the equation for lateral vibrations of an imperfect rod subjected to a periodic in time axial force [57,65]. As was shown in [57], the solution for the steady state vibrations in the vicinity of the secondary region of parametric instability, i.e., in the case where the frequency of the electric field is close to the fundamental frequency, can be adequately predicted retaining only three terms in series (20.20):

$$W_1 = A_0 + A_2 \cos \omega t + B_2 \sin \omega t. \tag{20.26}$$

Using (20.26), the ratio of the amplitude of vibrations neglecting physical nonlinearity, i.e., using $\mu = 0$ in (20.23) or (20.25), to the corresponding amplitude accounting for the physically nonlinear effect is obtained in the form

$$R = 1 - \frac{2\mu^2}{1 - (\omega/\lambda)^2}, \tag{20.27}$$

or

$$R = 1 - \frac{2\mu^2}{1 - (\omega/\lambda)^2 - \dfrac{\beta^2(\omega/\lambda)^2}{\lambda^2[(\omega/\lambda)^2 - 1]}} \tag{20.28}$$

for the cases without and with damping, respectively.

20.2.3. *Solution for Coupled Axial-Radial Axisymmetric Vibrations by the Generalized Galerkin Procedure*

Equations of the coupled axial-radial axisymmetric motion of the rod are

$$\frac{\partial \sigma_r}{\partial r} + \frac{\sigma_r - \sigma_\theta}{r} = m\ddot{u},$$
$$\frac{\partial \sigma_z}{\partial z} = m\ddot{w}. \tag{20.29}$$

The solution must satisfy the following conditions:

$$r = 0: \quad u = 0,$$
$$z = 0: \quad w = 0,$$
$$r = a: \quad \sigma_r = 0,$$
$$z = h: \quad \sigma_z = 0. \tag{20.30}$$

The solution is sought using normal modes of motion of the rod without the piezoelectric effect as generalized coordinates, i.e.,

$$u = \sum_{n=1} U_n(t) \sin \frac{n\pi r}{2a}, \quad w = \sum_{n=1} W_n(t) \sin \frac{n\pi z}{2h}, \tag{20.31}$$

where n are odd numbers. The Generalized Galerkin procedure implies

$$\int_0^{2\pi}\int_0^h\int_0^a\left(\frac{\partial\sigma_r}{\partial r}+\frac{\sigma_r-\sigma_\theta}{r}-m\ddot{u}\right)r\sin\frac{n\pi r}{2a}drdzd\theta$$
$$-\int_0^{2\pi}\int_0^h\sigma_r(r=a)a\sin\frac{n\pi}{2}dzd\theta=0, \tag{20.32}$$

and

$$\int_0^{2\pi}\int_0^h\int_0^a\left(\frac{\partial\sigma_z}{\partial z}-m\ddot{w}\right)r\sin\frac{n\pi z}{2h}drdzd\theta-\int_0^{2\pi}\int_0^a\sigma_z(z=h)r\sin\frac{n\pi}{2}drd\theta=0. \tag{20.33}$$

If the driving electric field is a periodic and harmonic function of time, the solution of the system of time-dependent equations of motion is available from (20.32) and (20.33) in the form of series

$$\left\{\begin{array}{c}U_n\\W_n\end{array}\right\}=\left\{\begin{array}{c}U_{n0}\\W_{n0}\end{array}\right\}+\sum_{k=1,2,3,\ldots}\left\{\begin{array}{c}U'_{nk}\sin\dfrac{k\omega t}{2}+U''_{nk}\cos\dfrac{k\omega t}{2}\\W'_{nk}\sin\dfrac{k\omega t}{2}+W''_{nk}\cos\dfrac{k\omega t}{2}\end{array}\right\}. \tag{20.34}$$

20.2.4. Numerical Results and Discussion

The following results were obtained for two representative materials, i.e., PZT-5H and PZN-4.5%PT, using data from Table 20.1 (the values of e_{33} shown without brackets were employed in calculations). In the examples presented below the amplitude of the electric field was chosen equal to 2.0 MV/m and 0.5 MV/m for PZN-4.5%PT and PZT-5H, respectively (unless indicated otherwise). Notably, even in the case of a relatively low electric field for PZT-5H, the effect of physical nonlinearity was significant as is shown in the following examples.

The effect of physical nonlinearity on the accuracy of the analysis of vibrations of a piezoelectric rod subject to electric fields with driving frequencies close to the fundamental frequency of the rod is illustrated in Figures 20.2 and 20.3. The horizontal axes in these figures represent $F=\omega/\lambda$, so that in the absence of damping the amplitude ratio becomes infinite at the resonant frequency. As follows from Figures 20.2 and 20.3, neglecting physical nonlinearity may cause a significant numerical error in the vicinity of the fundamental frequency, particularly at moderate and high electric fields. Even outside the resonant region, neglecting physical nonlinearity may result in a noticeable error.

The effect of damping on the ratio of the amplitude of linear vibrations to its nonlinear counterpart is elucidated in Figures 20.4 and 20.5. The quality factor employed to generate these results was chosen based on published data for PZT-5H and also applied to PZN-4.5%PT since damping data for this material is not available. Nevertheless, independent of the exact damping coefficient value, it is evident from the comparison with Figures 20.2 and 20.3 that the qualitative effect of damping is present only in the immediate vicinity of the fundamental frequency (the same conclusion is valid for other resonant frequencies, though the effect of damping becomes weaker at higher modes of motion). The same amplitude of the electric field was chosen in Figures 20.4 and 20.5 to illustrate

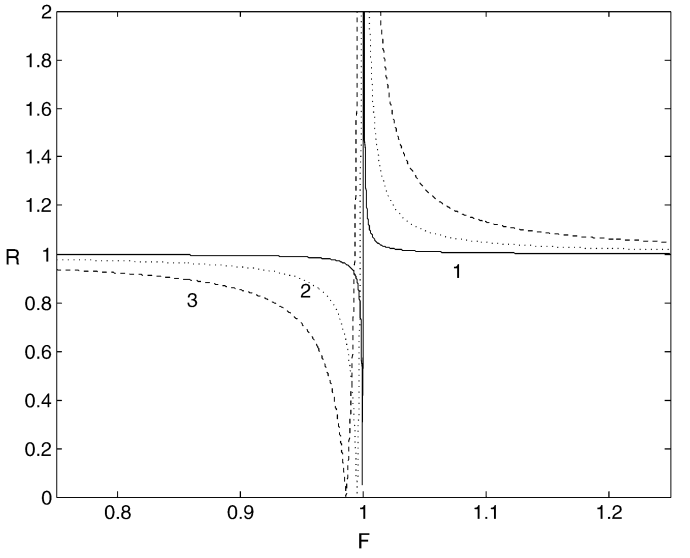


FIGURE 20.2. The ratio of the amplitude obtained without accounting for physical nonlinearity to the physically nonlinear counterpart as a function of the nondimensional frequency for a PZT-5H rod. Damping is neglected. The electric field corresponds to 0.1 MV/m, 0.3 MV/m and 0.5 MV/m for cases 1, 2 and 3, respectively. Reproduced with permission of ASME.

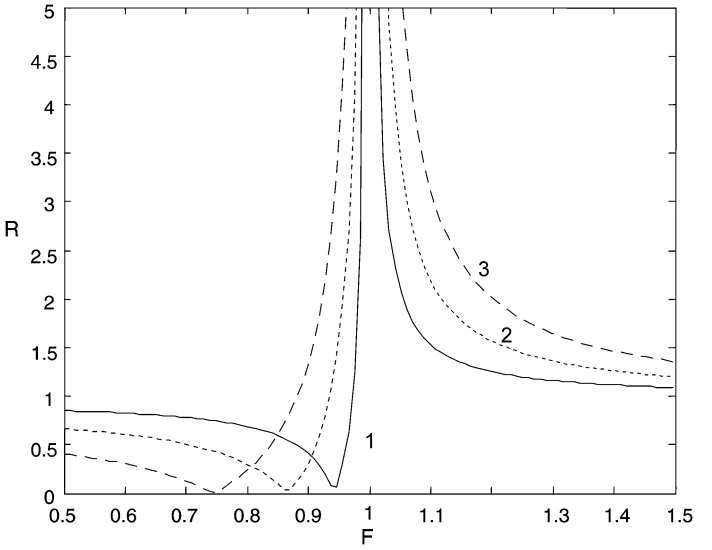


FIGURE 20.3. The ratio of the amplitude obtained without accounting for physical nonlinearity to the physically nonlinear counterpart as a function of the nondimensional frequency for a PZN-4.5%PT rod. Damping is neglected. The electric field corresponds to 1.0 MV/m, 1.5 MV/m and 2.0 MV/m for cases 1, 2 and 3, respectively. Reproduced with permission of ASME.

that the effect of physical nonlinearity is virtually identical for both materials considered in examples. The effect of damping on vibrations of physically nonlinear piezoelectric rods is further illustrated in Figures 20.6 and 20.7. As follows from these figures, the amplitudes

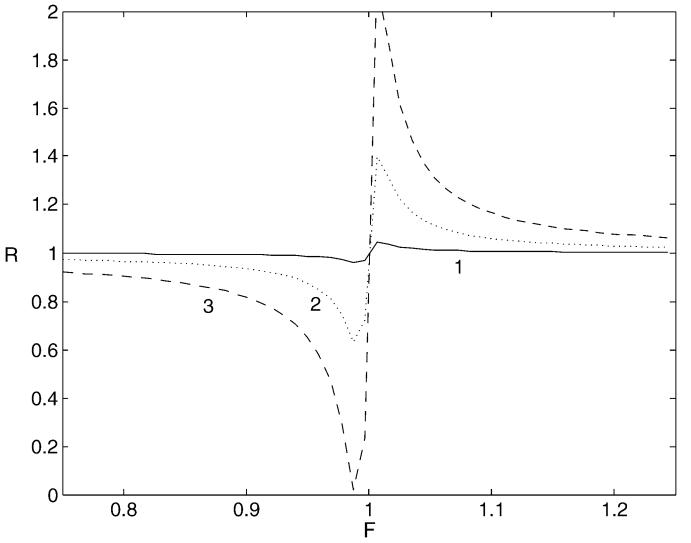


FIGURE 20.4. The ratio of the amplitude of a PZT-5H rod with a physically linear material behavior to the amplitude of the rod accounting for the physically nonlinear effect in the presence of damping. The quality factor is $Q = 65$. The height of the rod is $h = 0.1$ m. The electric field corresponds to 0.1 MV/m, 0.3 MV/m and 0.5 MV/m for cases 1, 2 and 3, respectively. Reproduced with permission of ASME.

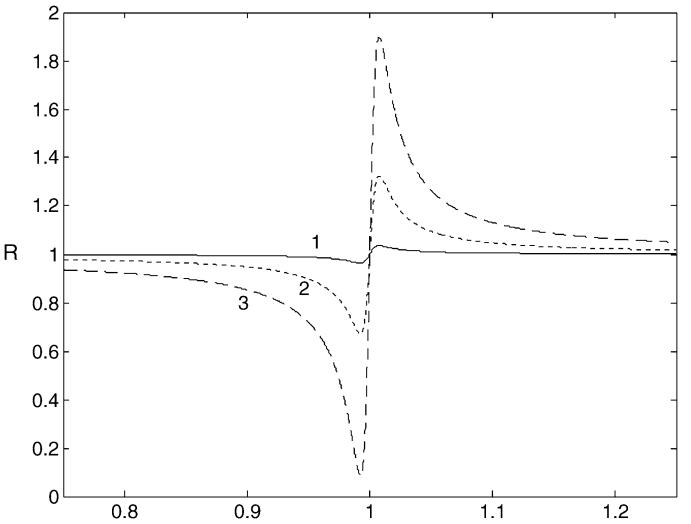


FIGURE 20.5. The ratio of the amplitude of a PZN-4.5%PT rod with a physically linear material behavior to the amplitude of the rod accounting for the physically nonlinear effect in the presence of damping. The quality factor is $Q = 65$. The height of the rod is $h = 0.1$ m. The electric field corresponds to 0.1 MV/m, 0.3 MV/m and 0.5 MV/m for cases 1, 2 and 3, respectively. Reproduced with permission of ASME.

of vibration are almost unaffected by damping outside a very narrow spectrum of driving frequencies encompassing the fundamental frequency. However, at the resonant frequency damping reduces the ratio of the amplitude of nonlinear vibrations with damping to that

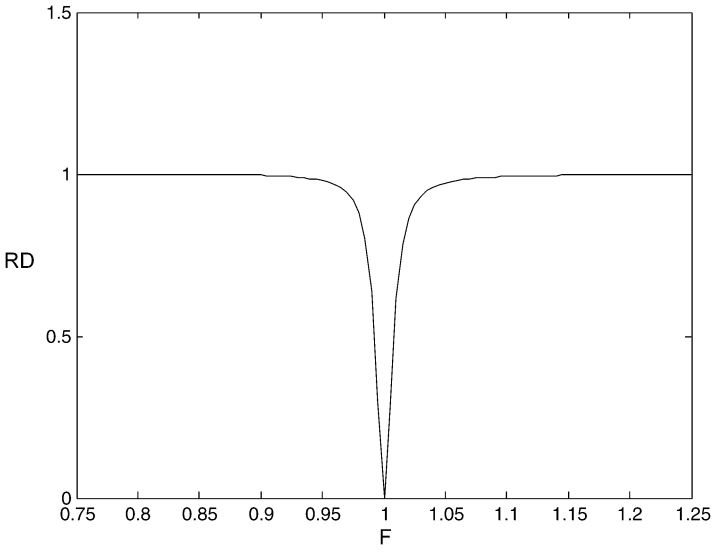


FIGURE 20.6. The ratio of the amplitude of vibrations of a physically nonlinear PZT-5H rod with damping to the amplitude without damping as a function of the nondimensional frequency. The electric field is equal to 0.5 MV/m. Reproduced with permission of ASME.

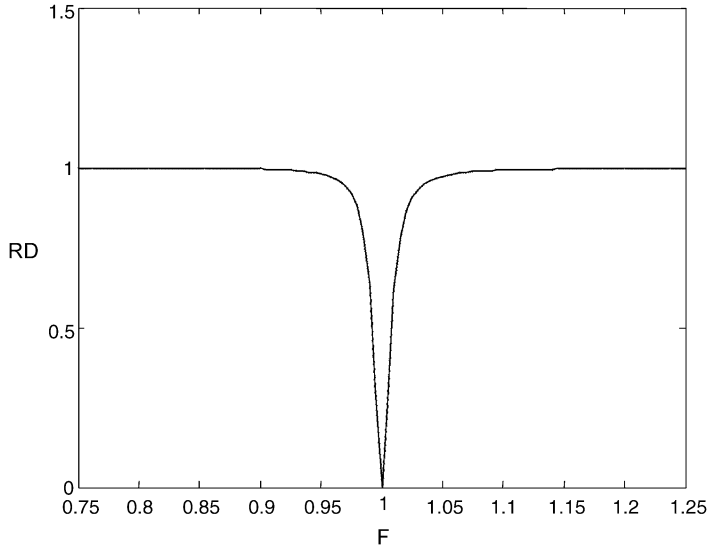


FIGURE 20.7. The ratio of the amplitude of vibrations of a physically nonlinear PZN-4.5%PT rod with damping to the amplitude without damping as a function of the nondimensional frequency. The electric filed is equal to 2.0 MV/m. Reproduced with permission of ASME.

without damping (denoted RD in Figures 20.6 and 20.7) to zero. This is anticipated since the amplitude of undamped motion is infinite at the resonance.

The limits of the accuracy of a one-term solution are elucidated in Figures 20.8 and 20.9. As follows from these figures, the contribution of higher modes is negligible

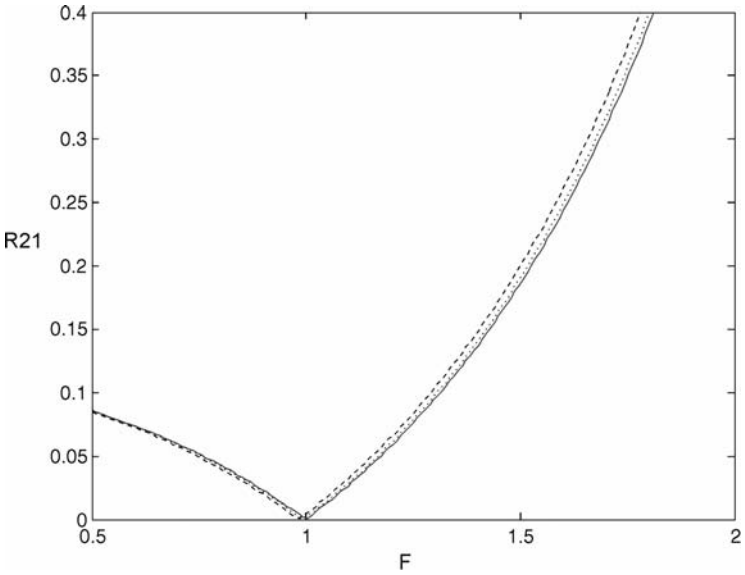


FIGURE 20.8. The ratio of the amplitude of motion corresponding to the second normal mode to that for the first normal mode as a function of the nondimensional frequency for a PZT-5H rod. The curves for the electric fields equal to 0.1 MV/m, 0.3 MV/m and 0.5 MV/m shown by solid, dotted and dashed lines, respectively, practically coincide.

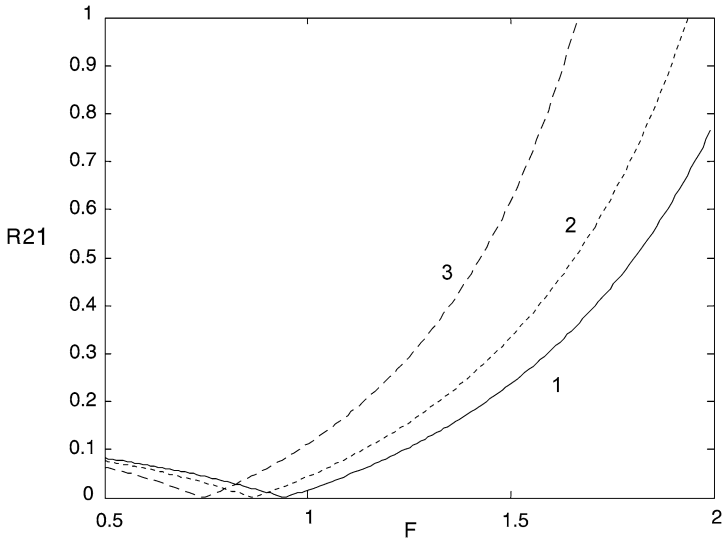


FIGURE 20.9. The ratio of the amplitude of the motion corresponding to the second normal mode to that for the first normal mode as a function of the nondimensional frequency for a PZN-4.5%PT rod. The electric field corresponds to 1.0 MV/m, 1.5 MV/m and 2.0 MV/m for cases 1, 2 and 3, respectively.

if the driving frequency is close to or smaller than the fundamental frequency of the rod. As the driving frequency increases, the contribution of the second normal mode becomes essential (the third and higher modes have a negligible effect in the range of frequencies

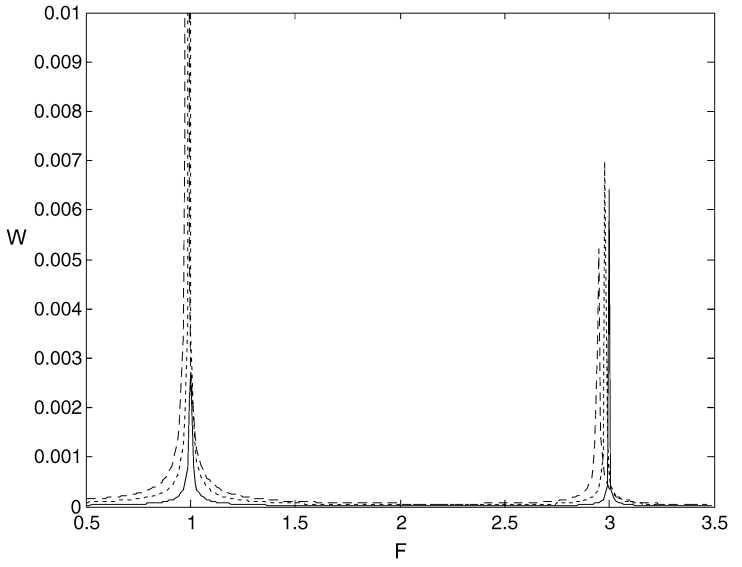


FIGURE 20.10. The nondimensional amplitude of axial vibrations of a PZT-5H rod ($h = 25$ mm) as a function of the nondimensional frequency. Electric field: — 0.1 MV/m; - - - - 0.3 MV/m; – – – – 0.5 MV/m. Reproduced with permission of ASME.

considered in Figures 20.8 and 20.9). The increase of the contribution of the second mode is particularly pronounced in case of a high electric field.

Finally, the multi-mode solutions for the nondimensional amplitude of physically nonlinear axial vibrations of piezoelectric rods within a broad range of driving frequencies are illustrated in Figures 20.10 and 20.11. As is shown in these figures, the amplitudes of motion remain quite small, except for very narrow regions encompassing the resonant frequencies. These regions of high-amplitude vibrations are very narrow, so that in the scale used in Figure 20.10, the peak corresponding to the second resonance ($F = 3$) could not even be shown approaching infinity (damping was not considered). Hence, the assumption employed in this solution, i.e., neglecting third-order strain terms in (20.3) and (20.9) was justified, except for the resonant frequencies. However, though geometric nonlinearity can be neglected (with the exception of the above-mentioned resonant frequencies of excitation), the physical nonlinearity should be always accounted for as follows from the previous discussion.

20.3. THE EFFECT OF THE NONLINEAR STRESS–STRAIN RELATIONSHIP ON THE RESPONSE OF OPTICAL FIBERS

The nonlinearity of the stress–strain relationship in glass optical fibers was reported in [66–68]. As was shown in these papers, the stresses in a fiber experiencing tensile strains limited to 5% can be represented by

$$\sigma = E_0 \varepsilon \left(1 + \frac{1}{2}\alpha\varepsilon\right), \tag{20.35}$$

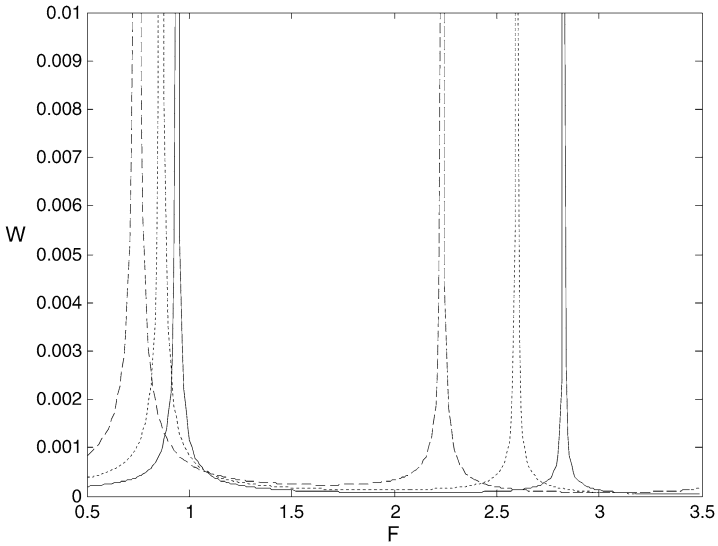


FIGURE 20.11. The nondimensional amplitude of axial vibrations of a PZN-4.5%PT rod ($h = 25$ mm) as a function of the nondimensional frequency. Electric field: — 1.0 MV/m; - - - - 1.5 MV/m; – – – – 2.0 MV/m. Reproduced with permission of ASME.

where σ and ε are the stress and strain, respectively, E_0 is the elasticity modulus corresponding to small strains, i.e., a physically linear range of the response, and α is a nonlinearity factor. Typical silica optical fibers are characterized by $E_0 = 72$ GPa and $\alpha = 6$. The analysis of stability, vibrations and bending of optical fibers and lightweight couplers conducted by Suhir [69–71] was based on the extrapolation of Equation (20.35) by assumption that in the case of compression the stress–strain relationship can be represented as

$$\sigma = E_0\varepsilon\left(1 - \frac{1}{2}\alpha\varepsilon\right). \tag{20.36}$$

20.3.1. Stability of Optical Fibers

Instability of optical fibers is detrimental to their reliability and causes transmission losses. Two types of instability are considered below include microbuckling of long dual-coated and very short bare fibers.

20.3.1.1. Microbending of Long Dual-Coated Fibers The critical stress that results in microbending (microbuckling) of dual-coated fibers with compliant thick low-modulus coatings is given by

$$\sigma_c = \sqrt{\frac{KE}{\pi}}, \tag{20.37}$$

where E is the fiber modulus and K is the spring constant of the coating system. Note that the stress given by (20.37) is applied to the fiber proper, i.e., it is assumed that coating does not carry any compressive loads.

In the case of a physically nonlinear fiber material, the modulus in (20.37) is affected by the stresses. As follows from (20.36),

$$E = \frac{d\sigma}{d\varepsilon} = E_0(1 - \alpha\varepsilon). \tag{20.38}$$

The substitution of (20.38) into (20.37) and transformations yield the following equation for the ratio of the critical stress for the nonlinear material to its counterpart obtained neglecting the nonlinearity

$$\eta_1^4 + 2\bar{\alpha}\eta_1 - 1 = 0, \tag{20.39}$$

where

$$\eta_1 = \frac{\sigma_c}{\sqrt{\dfrac{K E_0}{\pi}}}, \quad \bar{\alpha} = \alpha\sqrt{\frac{K}{\pi E_0}}. \tag{20.40}$$

The numerical solution of (20.39) is shown for $\alpha = 6$ in Figure 20.12. The analysis of the corresponding curve in this figure was conducted accounting for material constants of fibers with a dual acrylate coating and a silicone/nylon coating systems where $K = 3046$ MPa and $K = 90$ MPa, respectively. Using the modulus of elasticity of the fiber $E_0 = 72$ GPa yields $\sigma_0 = 8339$ MPa, $\eta_1 = 0.595$ for the fiber with a dual acrylate coating and $\sigma_0 = 1433$ MPa, $\eta_1 = 0.932$ for the fiber with a silicone/nylon coating. Evidently, the effect of physical nonlinearity on the stability of fibers with a dual acrylate coating cannot be disregarded. In reality, this effect may be smaller than predicted since the strains corresponding to the critical stress found above are too high (7.9%), exceeding the 5% limit of validity of Equation (20.36). In the case of the fiber with a silicone/nylon coating, the error due to neglecting physical nonlinearity is still significant being about 7%. The strains in the fiber experiencing microbuckling are only 1.8%, i.e., within the limits of validity of Equation (20.36).

20.3.1.2. Buckling of Short Bare Fiber Short bare fibers clamped at the ends are found in termination fixtures. The critical stress of such fibers is given by the Euler's formula as

$$\sigma_c = \left(\frac{\pi r_0}{2\mu l}\right)^2 E, \tag{20.41}$$

where l is the length of the fiber and $\mu = 0.5$ is the factor identifying the boundary conditions, i.e., clamped ends.

Using the expression for the modulus of elasticity (20.38) one obtains the following formula for the critical stress of a physically nonlinear fiber:

$$\sigma_c = \eta_2\left(\frac{\pi r_0}{2\mu l}\right)^2 E_0, \tag{20.42}$$

where

$$\eta_2 = 1 - \alpha\varepsilon. \tag{20.43}$$

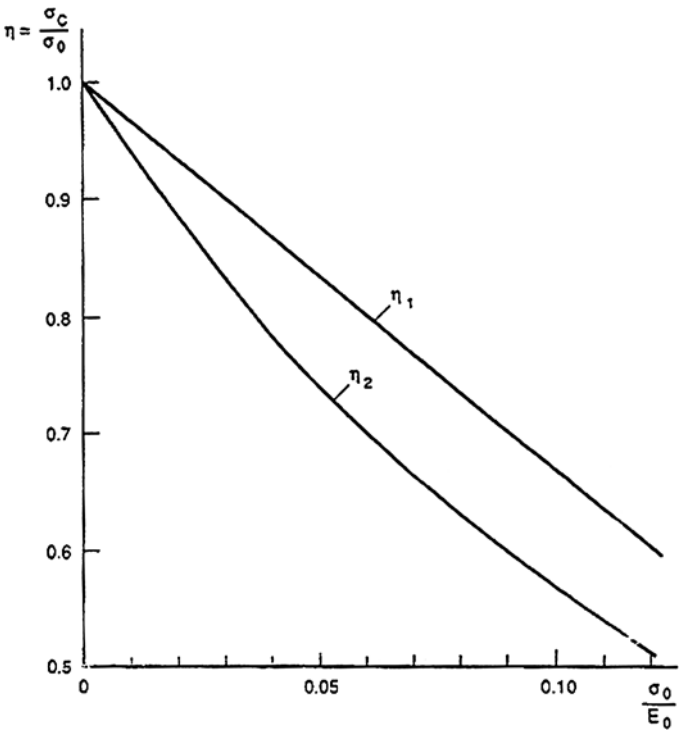


FIGURE 20.12. Effect of the nonlinear stress–strain relationship on the critical stress in coated (η_1) and bare (η_2) fibers. Reproduced with permission of ASME and Optical Society of America.

The combination of (20.36) and (20.43) yields the equation for the ratio of the critical stress in a physically nonlinear short bare fiber to its linear counterpart:

$$\eta_2 = \sqrt{1 - \bar{\alpha}^2} - \bar{\alpha}. \tag{20.44}$$

The corresponding curve is shown as a function of the ratio σ_0/E_0 in Figure 20.12. The applicability of the analysis can be assessed if we compare the buckling stress of a 2 mm long fiber ($\sigma_0 = 696$ MPa, $\sigma_0/E_0 = 0.0096$, $\bar{\alpha} = 0.0578$) to that of a 1 mm long fiber. The ratio η_2 is equal to 0.9439 and 0.7951, respectively. Therefore, the nonlinearity of the stress–strain curve becomes essential for very short fibers.

20.3.2. Stresses and Strains in a Lightwave Coupler Subjected to Tension

The cores of the fibers in fused biconical taper couplers (FBT) are positioned very close to each other to ensure coupling of two fundamental modes through their evanescent fields [72–75]. In order to bring the cores of the fibers in close proximity, the cladding of the midsection of the coupler has to be made very thin (Figure 20.13). At the same time, the coupler must possess sufficient strength both on a short and long-time scale to withstand appreciable axial deformations caused by its thermal mismatch with the substrate. Additionally, axial deformations may be applied deliberately to improve dynamic stability of the coupler increasing its natural frequencies.

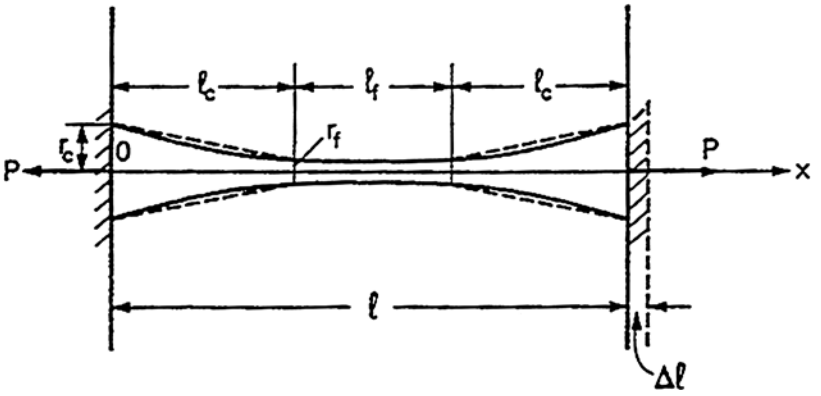


FIGURE 20.13. Geometry of a fused biconical taper (FBT) coupler. Reproduced with permission of ASME.

The stresses and strains in a FBT coupler subject to tensile deformations are considered here accounting for its nonprismacity and a nonlinear material behavior. It is assumed that the coupler geometry can be approximated by two circular conical parts connected by a circular cylindrical midsection, as shown in Figure 20.13 in broken lines. Such an approximation is adequate as long as the larger radius of the conical part r_c and the radius of the fused midsection r_f are chosen so that the areas of the corresponding circles are equal to the actual cross sectional areas.

The axial stress in an arbitrary cross section of the coupler is given by

$$\sigma(x) = \sigma_f \frac{r_f^2}{r^2(x)}, \tag{20.45}$$

where $r(x)$ is the radius of the cross section where the stress is evaluated, and $\sigma_f = P/(\pi r_f^2)$ is the stress at the midsection of the coupler subjected to a tensile force P. The origin of the coordinate x is chosen at the left end of the coupler.

According to the assumption regarding the geometry of the coupler introduced above, the radius of the cross section is given as

$$r = r(x) = \left\{ \begin{array}{ll} r_c - (r_c - r_f)\dfrac{x}{l_c}, & 0 \le x \le l_c \\ r_f, & l_c \le x \le l_c + l_f \\ r_c - (r_c - r_f)\dfrac{l-x}{l_c}, & l_c + l_f \le x \le l \end{array} \right\}. \tag{20.46}$$

In (20.46), l, l_c and l_f are the total length of the coupler, the length of one of its conical sections, and the length of the fused midsection, respectively.

The relationship between the strain and applied stress in an arbitrary cross section follows from (20.35) and (20.45):

$$\varepsilon = \frac{1}{\alpha}\left[\sqrt{1 + \beta^2 \frac{r_f^2}{r^2}} - 1\right], \tag{20.47}$$

where

$$\beta = 2\alpha \frac{\sigma_f}{E_0} = 2\alpha \frac{P}{\pi r_f^2 E_0}. \tag{20.48}$$

The total elongation of the coupler is available in the form [69]

$$\Delta l = \int_0^l \varepsilon(x)dx = \frac{1}{\alpha}[2f_c l_c + f_f l_f - l], \tag{20.49}$$

where the factors accounting for the effects of the magnitude of the applied force and the material nonlinearity on the elongations of the conical and fused sections are

$$\begin{aligned} f_c &= \frac{1}{l_c}\int_0^{l_c} \sqrt{1+\beta^2 \frac{r_f^2}{r^2}}dx \\ &= \frac{1}{1-\bar{\rho}}\left[\sqrt{1+\beta^2\bar{\rho}^2} - \bar{\rho}\sqrt{1+\beta^2} + \beta\bar{\rho}\ln\frac{\sqrt{1+\beta^2}+\beta}{\sqrt{1+\beta^2\bar{\rho}^2}+\beta\bar{\rho}}\right], \end{aligned} \tag{20.50}$$

$$f_f = \sqrt{1+\beta^2}.$$

In (20.50), $\bar{\rho} = r_f/r_c$.

As follows from the calculations presented in [69], in typical couplers the factors f_f reflecting the effect of the fused midsection on the total elongation of typical couplers are substantially larger than the factors f_c accounting for the effect of the conical sections (Table 20.3). This is due to a relatively high compliance of the fused section. The results shown in Table 20.3 for $\bar{\rho} = 0.08$ also illustrate that for sufficiently large deformations, the actual nonlinear strain at the fused midsection, ε_f, calculated for the applied stress σ_f by (20.45) is noticeably smaller than the "nominal" strain $\varepsilon_0 = \sigma_f/E_0 = \beta^2/2\alpha$ predicted by the linear theory.

The values of the applied forces P and the total strains $\Delta l/l$ shown in Table 20.3 were calculated using (20.49) to evaluate β for a prescribed total elongation. The fiber considered to generate the results in Table 20.3 had the following parameters: $l = 38$ mm, $l_f = 11.5$ mm, $r_f = 0.010$ mm, $r_c = 0.125$ mm. Obviously, rather low total strains lead to significantly higher strains in the fused midsection. Physically, this is explained by a higher stiffness of the conical sections of the coupler compared to the fused midsection.

If the nonlinearity of the stress–strain relationship was disregarded, Equation (20.49) would result in the following expression for the applied force:

$$P = \frac{\pi E_0 r_f^2 \Delta l}{2\rho l_c + l_f}. \tag{20.51}$$

The values of the force calculated by (20.51) are shown in Table 20.3 in parentheses. As follows from the comparison with the force found accounting for nonlinearity of the stress–strain relationship of the fibers, the linear approach results in a significant underestimation of the applied axial force, and hence, of the tensile stresses in the coupler subject to a prescribed deformation.

TABLE 20.3.
Strains and applied forces in a lightwave coupler subjected to tension. Reproduced with permission of ASME.

β	0	0.02	0.04	0.06	0.1	0.2	0.4	0.5	0.6	0.7	0.8
f_f	1	1.0002	1.0008	1.0018	1.0050	1.0198	1.0770	1.1180	1.1662	1.2207	1.2806
f_e	1	1.00000	1.00006	1.00013	1.00036	1.00296	1.00631	1.00979	1.01397	1.01886	1.02429
$\varepsilon_0 = \frac{\sigma_t}{E_0}$, %	0	0.0033	0.0133	0.0300	0.0833	0.3333	1.3333	2.0833	3.0000	4.0833	53333
$\varepsilon_t = \frac{f_f - 1}{\alpha}$, %,	0	0.0033	0.0133	0.0300	0.0833	0.3300	1.2833	1.9667	2.7700	3.6783	4.6767
$\frac{\Delta l}{l}$, %	0	0.0010	0.00466	0.00985	0.0291	0.1332	0.4571	0.7019	0.9908	1.3193	1.6810
P, gf	0	0.0773	0.3093	0.6959	1.933	0.732	30.93	48.33	69.59	94.72	123.72
		(0.0654)	(0.3047)	(0.6440)	(1.903)	(8.709)	(29.89)	(45.89)	(64.78)	(86.26)	(109.91)

20.3.3. Free Vibrations

20.3.3.1. Vibrations of a Long Fiber Subject to Tension The equation of free vibrations of a long and slender fiber with a negligible bending stiffness subjected to tension is

$$T\frac{\partial^2 w}{\partial x^2} - m\frac{\partial^2 w}{\partial t^2} = 0, \tag{20.52}$$

where $w = w(x,t)$ is a deflection, T is a tensile force, m is the mass per unit length, x is the axial coordinate, and t is time.

Following the classical approach, the deflection is represented by the series

$$w = \sum_{i=1}^{\infty} \theta_i(t) \sin\frac{i\pi x}{l}, \tag{20.53}$$

where l is the length of the fiber and $\theta_i(t)$ is the principal coordinate of the i-th mode of vibration.

The substitution of (20.53) into (20.52) yields

$$\ddot{\theta}_i + \omega_i^2 \theta_i = 0, \tag{20.54}$$

where the natural frequency corresponding to the i-th mode is given by

$$\omega_i = \frac{i\pi}{l}\sqrt{\frac{T}{m}} = \frac{i\pi}{l}\sqrt{\frac{\sigma}{\rho}}. \tag{20.55}$$

In (20.55), σ is the tensile stress in the fiber and ρ is the mass density of the material that is typically equal to 2.2 g/cm^3.

The substitution of (20.35) into (20.55) yields

$$\omega_i = \eta_3\omega_0, \quad \omega_0 = \frac{i\pi}{l}\sqrt{\frac{E_0}{\rho}},$$

$$\eta_3 = \sqrt{1 + \frac{\alpha\varepsilon}{2}}. \tag{20.56}$$

Obviously, ω_0 is the natural frequency of the physically linear fiber, while η_3 is the ratio of the natural frequency of the physically nonlinear fiber to that of its linear counterpart. For example, if the actual strain in the fiber is equal to $\varepsilon = 0.05$, the third Equation (20.56) yields $\eta_3 = 1.072$ reflecting a 7.2% error due to the linear approximation of the fiber stress–strain relationship.

20.3.3.2. Vibrations of a Lightwave Coupler The kinetic energy of a lightwave coupler considered above (Figure 20.13) is

$$T' = \frac{1}{2}\int_0^l m(x)\left(\frac{\partial w}{\partial t}\right)^2 dx. \tag{20.57}$$

The strain energy of the coupler experiencing small-amplitude vibrations is found as that of a nonuniform beam subject to tensile force P:

$$V = \frac{P}{2}\int_0^l \left(\frac{\partial w}{\partial x}\right)^2 dx + \frac{1}{2}\int_0^l EI(x)\left(\frac{\partial^2 w}{\partial x^2}\right)^2 dx, \tag{20.58}$$

where $I(x) = \pi r^4(x)/4$ represents the moment of inertia of the circular cross section of the coupler.

Couplers found in applications are quite long compared to their cross sectional dimensions so that the effect of the actual boundary conditions is negligible. As was shown in [69], the difference between the solutions for clamped and simply supported couplers is negligible, even if tension is relatively small. Moreover, the analysis in [69] illustrated that the effect of the bending term in the expression for the strain energy can be neglected compared to the contribution of tensile force, i.e.,

$$V = \frac{P}{2}\int_0^l \left(\frac{\partial w}{\partial x}\right)^2 dx. \tag{20.59}$$

The simplifications outlined above, i.e., simple support of the ends of the coupler and the adoption of expression (20.59) for the strain energy are applicable if the tensile force exceeds the value

$$P_s = \frac{\pi^3 i^2 (1+\sqrt{2})}{2} E \frac{r_0^4}{l^2},$$
$$r_0 = \frac{r_c + r_f}{2}. \tag{20.60}$$

For example, if $r_0 = 0.0675$ mm, $E = E_0$, $i = 1$, the limiting value of the force $P_s = 0.00387$ g. The fact that this force is so low and the actual tensile force is almost always much higher justifies the validity of the above-mentioned simplifications.

The analysis can now be conducted using the normal modes for a simply supported beam, i.e.,

$$w = \sum_{i=1} A_i \sin\frac{i\pi x}{l} \sin\omega_i t, \tag{20.61}$$

where A_i are the amplitudes of motion and ω_i are the natural frequencies. For example, limiting the analysis to the fundamental vibration mode ($i = 1$), the kinetic and strain energy given by (20.57) and (20.59) become (the index $i = 1$ is omitted):

$$T' = \rho A^2 \omega^2 l^3 C \cos^2\omega t,$$
$$V = \frac{\pi^2 P}{4l} A^2 \sin^2\omega t, \tag{20.62}$$

where the constant C is given by

$$C = \left(\frac{r_c - r_f}{l_c}\right)^2 \left\{ \frac{1}{6}\left(\frac{l_c}{l}\right)^3 + \left[\frac{1}{8\pi^3} - \frac{1}{4\pi}\left(\frac{l_c}{l}\right)^2\right] \sin\frac{2\pi l_c}{l} - \frac{l_c}{4\pi^2 l}\cos\frac{2\pi l_c}{l} \right\}$$
$$- 2\frac{r_c}{l}\frac{r_c - r_f}{l_c}\left[\frac{1}{8\pi^2} + \frac{1}{4}\left(\frac{l_c}{l}\right)^2 - \frac{l_c}{4\pi l}\sin\frac{2\pi l_c}{l} - \frac{1}{8\pi^2}\cos\frac{2\pi l_c}{l}\right]$$
$$+ \frac{1}{2}\left(\frac{r_c}{l}\right)^2\left(\frac{l_c}{l} - \frac{1}{2\pi}\sin\frac{2\pi l_c}{l}\right)$$
$$+ \frac{1}{4}\left(\frac{r_f}{l}\right)^2\left[\frac{l_f}{l} + \frac{1}{2\pi}\sin\frac{2\pi l_c}{l} - \frac{1}{2\pi}\sin\frac{2\pi(l_c + l_f)}{l}\right]. \tag{20.63}$$

The Rayleigh method, i.e., the requirement that the maximum values of the kinetic and strain energy should be equal yields the following formula for the fundamental (lowest) vibration frequency:

$$\omega = \frac{\pi}{2l^2}\sqrt{\frac{\rho P}{C}}. \tag{20.64}$$

The initial strain that results in the prescribed value of the fundamental frequency determined by (20.64) causes the following stress in the fused midsection of the coupler

$$\sigma_f = \frac{2\rho C}{\pi^3}\left(\frac{\omega l^2}{r_f}\right)^2. \tag{20.65}$$

Let the highest expected frequency be equal to 2000 Hz. Using the factor of safety equal to two, the required natural frequency of the coupler vibration is 4000 Hz. Then, considering the coupler analyzed above, $C = 16.866 \times 10^{-8}$, $\sigma_f = 67.8$ kgf/mm^2, $\varepsilon_f = 0.894\%$. This implies that the silica material should be able to withstand long-term strains exceeding 0.9%. Equation (20.65) could be used for the prediction of the initial tensile force, stress and strain from the measured vibration frequency.

The tensile force resulting in the stress in the fiber equal to 67.8 kgf/mm^2 is equal to $P = 21.3$ g. Neglecting the nonlinearity in the stress–strain relationship the tensile force would be about $P = 20.7$ g. The comparison of the natural frequencies corresponding to these values of the tensile force and given by (20.64) indicates that the difference between the values obtained for physically nonlinear and linear material models is less than 2%, i.e., the nonlinearity is not essential in this particular example.

20.3.4. Bending of an Optical Fiber

The position of the centroid (neutral axis) of an optical fiber experiencing bending can be determined from (Figure 20.14):

$$\int_{\hat{A}} E z_1 d\hat{A} = 0, \tag{20.66}$$

where $\hat{A}$ is the cross sectional area and the coordinate z_1 is a distance from the neutral axis. An element of the area can be evaluated by the formula

$$d\hat{A} = \sqrt{r_0^2 - (z_1 + z_c)^2}\, dz, \tag{20.67}$$

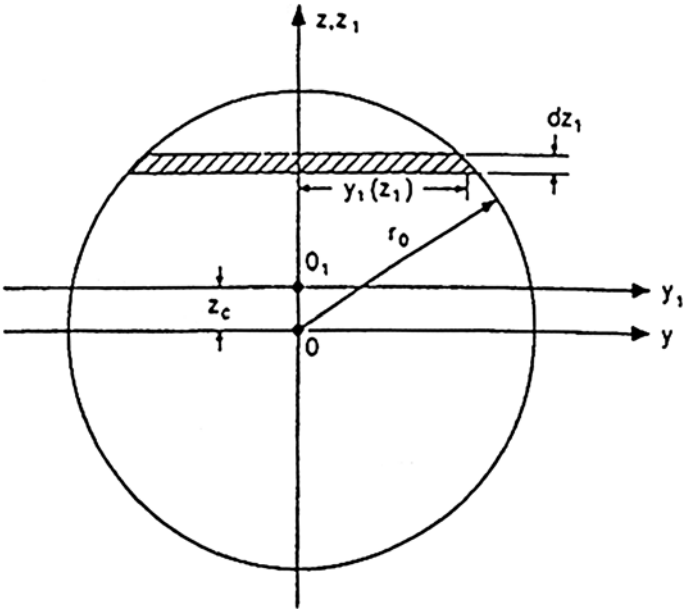


FIGURE 20.14. Silica fiber cross section. Reproduced with permission of Optical Society of America.

where z_c is a deviation of the neutral axis from the geometrical center of the fiber cross section.

The relationship between the strain and the radius of curvature, i.e., $\varepsilon = z_1/R$, can be employed to modify the expression for the tensile modulus obtained from (20.35):

$$E = \frac{d\sigma}{d\varepsilon} = E_0\left(1 + \alpha\frac{z_1}{R}\right). \tag{20.68}$$

The subsequent analysis is conducted by assumption that the effect of the nonlinear stress–strain relationship on the radius of curvature is negligible compared to its effect on the maximum stress. The justification for this assumption is found in [72]. Then, substituting (20.67) and (20.68) into (20.66) one obtains

$$\int_{-(r_0+z_c)}^{r_0-z_c}\left(z_1 + \alpha\frac{z_1^2}{R}\right)\sqrt{r_0^2 - (z_1+z_c)^2}\,dz_1 = 0. \tag{20.69}$$

As is shown in [69], this integral yields the following solution for the position of the neutral axis:

$$z_c = r_0\frac{1-\sqrt{1-(\alpha\varepsilon_0)^2}}{2\alpha\varepsilon_0}, \tag{20.70}$$

where $\varepsilon_0 = r_0/R$ represents a "nominal" linear strain. When this strain is very small, the value of $\alpha\varepsilon_0$ is much smaller than unity and (20.70) can be reduced to

$$z_c = \frac{\alpha\varepsilon_0 r_0}{4}. \tag{20.71}$$

TABLE 20.4.
Stresses, strains and a distance between faceplates in a representative two-point bending test.
Reproduced with permission of ASME and Optical Society of America.

ε_0, %	1	2	3	4	5	6	7	8
ε_t, %	0.985	1.940	2.864	3.7564	4.6162	5.4413	6.2294	6.977
σ_t/E_0	0.01014	0.02053	0.03110	0.04180	0.05255	0.06330	0.07393	0.08437
ε_c, %	1.015	2.060	3.136	4.2436	5.3838	6.5587	7.7707	9.023
σ_c/E_0	0.00984	0.01933	0.02841	0.03703	0.04514	0.05268	0.05959	0.06580
σ_t^0/E_0	0.01030	0.02120	0.03270	0.04480	0.05750	0.07080	0.08470	0.09920
σ_c^0/E_0	0.00970	0.01880	0.02730	0.03520	0.04250	0.04920	0.0553	0.0608
D, mm	14.98	7.49	4.99	3.74	3.00	2.50	2.14	1.87

If $\alpha = 6$, Equation (20.71) yields

$$\frac{z_c}{r_0} = \frac{3}{2}\varepsilon_0. \tag{20.72}$$

Equation (20.72) implies that the relative shift in the position of the centroid due to the nonlinear stress–strain relationship is greater than the nominal bending strain by a factor of 1.5.

The maximum tensile and compressive strains in a glass fiber subjected to bending can be determined accounting for the shift in the centroid due to the nonlinear stress–strain relationship as follows:

$$\varepsilon_t = \frac{r_0 - z_c}{R} = \varepsilon_0\left[1 - \frac{1 - \sqrt{1 - \alpha^2\varepsilon_0^2}}{2\alpha\varepsilon_0}\right],$$
$$\varepsilon_c = \frac{r_0 + z_c}{R} = \varepsilon_0\left[1 + \frac{1 - \sqrt{1 - \alpha^2\varepsilon_0^2}}{2\alpha\varepsilon_0}\right]. \tag{20.73}$$

The corresponding maximum stresses are immediately available using (20.35) and (20.36) as

$$\sigma_t = E_0\varepsilon_t\left(1 + \frac{1}{2}\alpha\varepsilon_t\right),$$
$$\sigma_c = E_0\varepsilon_c\left(1 - \frac{1}{2}\alpha\varepsilon_c\right). \tag{20.74}$$

The strains and stresses calculated using (20.73) and (20.74) are shown in Table 20.4. In this table, ε_0 is the "nominal" linear strain, σ_t/E_0 and σ_c/E_0 are the ratios of the maximum tensile and compressive stresses to the "nominal" (low strain) modulus of elasticity, and σ_t^0/E_0 and σ_c^0/E_0 are similar ratios calculated by (20.74) using $\varepsilon_t = \varepsilon_c = \varepsilon_0$, i.e., neglecting the effect of the nonlinear stress–strain relationship.

The results obtained in this section can be particularly useful for the evaluation of the maximum stresses in optical fibers subjected to large deformations during a two-point bending test (Figure 20.15). This technique [71] involves constraining a bent loop of the fiber between two faceplates which are then brought together until the desirable gap is

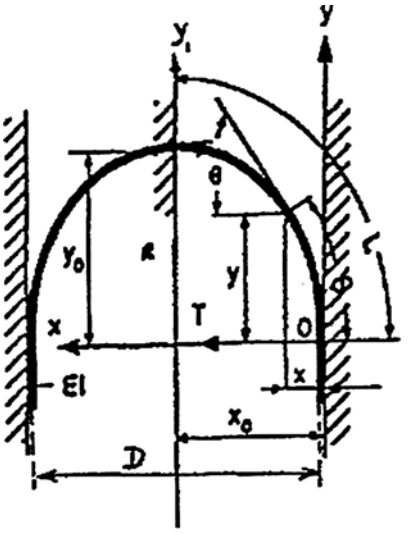


FIGURE 20.15. Silica fiber subjected to two-point bending. Reproduced with permission of Optical Society of America.

achieved, or until the fiber breaks. A modification of this technique involves inserting a U-shaped bend of a fiber into a glass tube of the given inner diameter. The minimum radius of curvature at the midpoint of the fiber bend (R) is related to the distance between the fiber axes in the region beyond the bend (D) by $R = 0.4173D$ [69]. As a result, the "nominal" strain can be found in the form

$$\varepsilon_0 = \frac{r_0}{R} = 2.4\frac{r_0}{D}. \tag{20.75}$$

The calculated values of D are shown in Table 20.4 for a 125 μm glass fiber of the 0.0625 mm radius. The relationship between the ratio of the induced stresses to the "nominal" (low strain) modulus of elasticity are plotted in Figure 20.16 as functions of the distance between faceplates or the tube diameter D. As is evident from these data, the nonlinear stress–strain relationship can have an appreciable effect on the maximum stresses, especially in the region of relatively high bending strains (small tube diameter). An approach which considers the effect of the nonlinear stress–strain relationship on the effective modulus of elasticity, but does not account for this effect on the shift in the position of the centroid of the fiber cross section is inconsistent. Such approach results in an overestimation of the maximum tensile stresses and in an underestimation of the maximum compressive stresses.

20.4. CONCLUSIONS

The chapter illustrates the effect of physical nonlinearity, i.e., nonlinear stress–strain relationships on the response of typical micro- and opto-electronic systems. Two examples are considered, including piezoelectric transducers and optical fibers. In both examples, neglecting physical nonlinearity may result in significant errors in the predicted response.

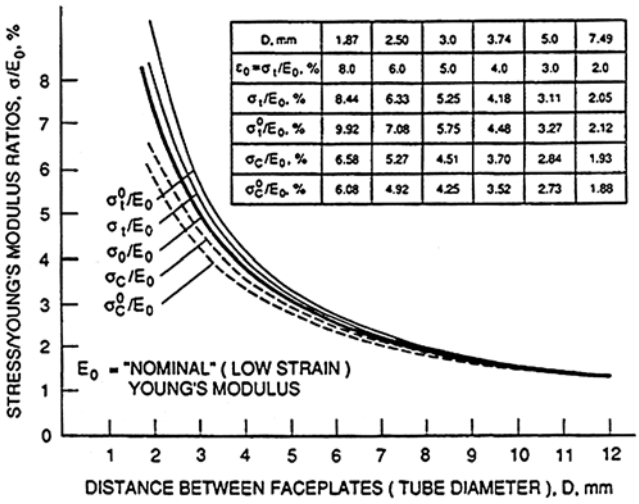


FIGURE 20.16. Stresses and strains in a silica fiber subject to a two-point bending: σ_t, maximum tensile stress; σ_c, maximum compressive stress; σ_t^0, maximum tensile stress without considering the shift in the centroid; σ_c^0, maximum compressive stress without considering the shift in the centroid. Reproduced with permission of ASME and Optical Society of America.

The effect of physical nonlinearity on the dynamic response of piezoelectric rods polarized in the axial direction and subject to harmonic electric field acting in this direction investigated in the chapter is important in a number of applications employing piezoelectric transducers. It is shown that physically nonlinear effects become more pronounced at higher fields. Mathematically, the presence of physical nonlinearity results in additional terms in the system of equations of motion, so that these equations change from the equations for forced vibrations in the physically linear case to Mathieu-Hill equations.

Vibrations of the rods driven by a harmonic electric field were numerically investigated in the vicinity of the fundamental frequency. As follows from the representative examples, the error due to neglecting physical nonlinearity increases in the close vicinity to the resonant frequency. Damping has a noticeable effect on vibrations of piezoelectric rods in the vicinity of the resonant frequency. However, the frequency range where the effect of damping is significant is very narrow. Within this range, damping results in finite amplitude of vibrations, while without damping, the amplitude is infinite, similarly to the situation encountered in the problem of forced vibrations.

The nonlinear stress–strain relationship in silica materials can have a significant effect on the mechanical behavior of optical fibers experiencing static or dynamic loading. A number of problems considered in the chapter include stability, vibration, bending and two-point testing of optical fibers. In most typical situations, a physically nonlinear behavior of silica under both tensile and compressive loads could not be disregarded without sacrificing the accuracy of the solution.

ACKNOWLEDGMENT

This research was partially supported by the Army Research Office Contract W911NF-04-1-0192. The support and encouragement of the Program Director, Dr. Gary

L. Anderson are greatly appreciated. The authors are also grateful to Professors George J. Simitses (Georgia Institute of Technology), Liviu Librescu (Virginia Polytechnic Institute and State University) and Daniel S. Stutts (University of Missouri-Rolla) for valuable discussions of various aspects of the problem relevant to the response of piezoelectric transducers.

REFERENCES

1. G.A. Maugin, J. Pouget, R. Drouot, and B. Collet, Nonlinear Electromechanical Couplings, Wiley, Chichester, 1992.
2. D. Berlington and H.H.A. Krueger, Domain processes in lead titanate zirconate and barium titanate ceramics, Journal of Applied Physics, 30, pp. 1804–1810 (1959).
3. R.S. Woolett and G.L. Leblanc, Ferroelectric nonlinearities in transducer ceramics, IEEE Trans. on Sonics and Ultrasonics, SU-20, pp. 24–31 (1973).
4. H.H.A. Krueger, Mechanical properties of ceramic barium titanate, Physics Reviews, 93, p. 362 (1954).
5. H.H.A. Krueger, Stress sensitivity of piezoelectric ceramics: Part 1. Sensitivity to compressive stress parallel to the polar axis, Journal of the Acoustical Society of America, 42, pp. 636–645 (1967).
6. H.H.A. Krueger, Stress sensitivity of piezoelectric ceramics: Part 2. Heat treatment, Journal of the Acoustical Society of America, 43, pp. 576–582 (1968).
7. H.H.A. Krueger, Stress sensitivity of piezoelectric ceramics: Part 3. Sensitivity to compressive stress perpendicular to the polar axis, Journal of the Acoustical Society of America, 43, pp. 583–591 (1968).
8. R.F. Brown and G.W. McMahon, Material constants of ferroelectric ceramics at high pressure, Canadian Journal of Physics, 40, pp. 672–674 (1962).
9. R.F. Brown and G.W. McMahon, Properties of transducer ceramics under maintained planar stress, Journal of Acoustical Society of America, 38, pp. 570–575 (1965).
10. I.J. Fritz, Uniaxial stress effects in a 95/5 lead zirconate titanate ceramic, Journal of Applied Physics, 49, pp. 4922–4928 (1978).
11. H.F. Tiersten, On the nonlinear equations in thermoelectroelasticity, International Journal of Engineering Science, 9, pp. 587–604 (1971).
12. H.F. Tiersten, Equations for the extension and flexure of relatively thin electroelastic plates undergoing large electric fields, Mechanics of Electromagnetic Materials and Structures, AMD-Vol. 161/MD-Vol. 42, ASME, New York, 1993, pp. 21–34.
13. H.F. Tiersten, Electroelastic equations for electroded thin plates subject to large driving voltages, Journal of Applied Physics, 74, pp. 3389–3393 (1993).
14. Y. Nakagawa, K. Yamanouchi, and K. Shibayama, Third-order elastic constants of lithium niobate, Journal of Applied Physics, 44, pp. 3969–3974 (1973).
15. Y. Cho and R. Yamanouchi, Non-linear elastic, piezoelectric, electrostrictive and dielectric constants of lithium niobate, Journal of Applied Physics, 61, pp. 875–887 (1987).
16. P.J. Chen and S.T. Montgomery, A macroscopic theory for existence of the hysteresis and butterfly loops in ferroelectricty, Ferroelectrics, 23, pp. 199–208 (1980).
17. P.J. Chen and M.M. Madsen, One dimensional polar responses of the electrooptic ceramic PLZT 7/65/35 due to domain switching, Acta Mechanica, 41, pp. 255–264 (1981).
18. E. Bassiouny, A.F. Ghaler, and G.A. Maugin, Thermodynamical formulation for coupled electromechanical hysteresis effects—I. Basic equations, International Journal of Engineering Science, 26, pp. 1279–1295 (1988).
19. T. Leigh and D. Zimmermann, An implicit method for the non-linear modeling and simulation of piezoceramic actuators displaying hysteresis, Smart Materials and Structures, AD-Vol. 24/AMD-Vol. 123, ASME, ASME Press, New York, 1991, pp. 57–63.
20. P. Ge and M. Jovuaneh, Modeling hysteresis in piezoceramic actuators, Precision Engineering, 17, pp. 211–221 (1995).
21. H. Beige and G. Schmidt, Electromechanical resonances for investigating linear and non-linear properties of dielectrics, Ferroelectrics, 41, pp. 39–39 (1982).
22. H. Beige, Elastic and dielectric non-linearities of piezoelectric ceramics, Ferroelectrics, 51, pp. 113–119 (1983).
23. U. von Wagner and P. Hagedorn, Piezo-beam-systems subjected to weak electric field: experiments and modeling of non-linearities, Journal of Sound and Vibration, 256, pp. 861–872 (2002).

24. U. von Wagner, Nonlinear longitudinal vibrations of piezoceramics excited by weak electric fields, International Journal of Non-Linear Mechanics, 38, pp. 565–574 (2003).
25. U. von Wagner, Non-linear longitudinal vibrations of non-slender piezoelectric rods, Journal of Non-Linear Mechanics, 39, pp. 673–688 (2004).
26. U. von Wagner and P. Hagedorn, Nonlinear effects of piezoceramics excited by weak electric fields, Journal of Nonlinear Dynamics, 31, pp. 133–149 (2003).
27. A. Chattopadhyay, H. Gu, and Q. Liu, Modeling of smart composite box beams with nonlinear induced strain, Composites: Part B, 30, pp. 603–612 (1999).
28. R.P. Thornburgh and A. Chattopadhyay, Nonlinear actuation of smart composites using a coupled piezoelectric-mechanical model, Smart Materials and Structures, 10, pp. 743–749 (2001).
29. S. Sherrit, R.B. Stimpson, H.D. Wiederick, and B.K. Mukherjee, Stress and temperature dependence of the direct piezoelectric charge coefficient in lead zirconate titanate ceramics, in V.K. Aarte, V.K. Varadan, and V.V. Varadan, Eds., Smart Materials, Structures and MEMS, Proceedings of SPIE, 3321, pp. 104–113 (1996).
30. H.S. Wiederick, S. Sherrit, R.B. Stimpson, and B.K. Mukherjee, An optical lever measurement of the piezoelectric charge coefficient, Ferroelectrics, 186, pp. 25–31 (1996).
31. S. Sherrit, H.D. Wiedenrick, B.K. Mukherjee, and M. Sayer, Field dependence of the complex piezoelectric, dielectric, and elastic constants of motorola PZT 3203 HD ceramic, in W.C. Simmons, et al., Eds., Smart Materials Technologies, SPIE Proceedings 3040, pp. 99–109 (1997).
32. G. Yang, S.-F. Liu, W. Ren, and B.K. Mukherjee, Uniaxial stress dependence of the piezoelectric properties of lead zirconate titanate ceramics, in C.S. Lynch, Ed., Active Materials: Behavior and Mechanics, Proceedings SPIE, 3992, pp. 103–1013 (2000).
33. B.K. Mukherjee, W. Ren, G. Yang, S.F. Liu, and A.J. Masys, Nonlinear properties of piezoelectric coefficients, in C.S. Lynch, Ed., Active Materials: Behavior and Mechanics, Proceedings SPIE, 4333, pp. 41–54 (2001).
34. W. Ren, A.J. Masys, G. Yang, and B.K. Mukherjee, The field and frequency dependence of the strain and polarization in piezoelectric and electrostrictive ceramics, Presented at the 3rd Asian Meeting on Ferroelectrics (AMF 3), Hong Kong, December 12–15, 2000.
35. D.P. Wang and G.P. Carman, Evaluating the behavior of piezoelectric ceramics subjected to thermal fields, Adaptive Material Systems, Summer Symposium of ASME at Los Angeles, AMD-Vol. 206, MD-Vol. 58, 1995, pp. 33–47.
36. P.M. Chaplya and G.P. Carman, Compression of piezoelectric ceramic at constant electric field: energy absorption through Non-180 domain-wall motion, Journal of Applied Physics, 92, pp. 1504–1510 (2002).
37. R. Barrett, Design and manufacturing of adaptive composites for active flight control surfaces, Presented at the Second International Conference on Composites Engineering, New Orleans, LA, August 21–24, 1995.
38. C.W. Bert and V. Birman, Stress dependency of the thermoelastic and piezoelectric coefficients, AIAA Journal, 37, pp. 135–137 (1999).
39. C.W. Bert and V. Birman, Effects of stress and electric field on the coefficients of piezoelectric materials: one-dimensional formulation, Mechanics Research Communications, 25, pp. 165–169 (1998).
40. S.P. Joshi, Non-linear constitutive relations for piezoceramic materials, Smart Materials and Structures, 1, pp. 80–83 (1992).
41. Q.-M. Wang, Q. Zhang, B. Xu, R. Liu, and E.L. Cross, Nonlinear piezoelectric behavior of ceramic bending mode actuators under strong electric fields, Journal of Applied Physics, 86, pp. 3352–3361 (1999).
42. A. Achuthan, A.K. Keng, and W.C. Ming, Shape control of coupled nonlinear piezoelectric beams, Smart Materials and Structures, 10, pp. 914–924 (2001).
43. S. Sherrit, G. Catoiu, and K.B. Mukherjee, The characterization and modeling of electrostrictive ceramics for transducers, Ferroelectrics, 228, pp. 167–196 (1999).
44. R.B. Williams, Nonlinear mechanical and actuation characterization of piezoceramic fiber composites, Ph.D. thesis, Virginia Polytechnic Institute and State University, Blacksburg, Virginia, March 22, 2004.
45. R. Barrett, Private communication, 2002.
46. S. Priya, D. Viehland, A.V. Carazo, J. Ryu, and K. Uchino, High-power resonant measurements of piezoelectric materials: importance of elastic nonlinearities, Journal of Applied Physics, 90, pp. 1469–1479 (2001).
47. L. Li and N.R. Sottos, Predictions of static displacements in 1–3 piezocomposites, Journal of Intelligent Material Systems and Structures, 6, pp. 169–180 (1995).
48. L. Li and N.R. Sottos, A design for optimizing the hydrostatic performance of 1–3 piezocomposites, Ferroelectric Letters, 21, pp. 41–46 (1996).
49. L. Li and N.R. Sottos, Measurement of surface displacements in 1–3 and 1–1–3 piezocomposites, Journal of Applied Physics, 79, pp. 1707–1712 (1996).

50. O. Sigmund, S. Torquato, and I.A. Aksay, On the design of 1–3 piezocomposites using topology optimization, Journal of Material Research, 13, pp. 1038–1048 (1998).
51. P. Tan and L. Tong, A one-dimensional model for non-linear behaviour of piezoelectric composite materials, Composite Structures, 58, pp. 551–561 (2002).
52. P. Tan and L. Tong, Micromechanics models for non-linear behavior of piezo-electric fiber reinforced composite materials, International Journal of Solids and Structures, 38, pp. 8999–9032 (2001).
53. C.-H. Huang, Y.-C. Lin, and C.-C. Ma, Theoretical analysis and experimental measurement for resonant vibration of piezoceramic circular plates, IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, 51, pp. 12–24 (2004).
54. O. Guillon, F. Thiebaud, P. Delobelle, and D. Perreux, Tensile behavior of PZT in short and open-circuit conditions, Materials Letters, 58, pp. 986–990 (2004).
55. G.J. Simitses, An Introduction to the Elastic Stability of Structures, Robert Krieger Publishing Company, Malabar, Florida, 1986.
56. J.C. Houbolt and G.W. Brooks, Differential equations of motion for combined flapwise bending, chordwise bending, and torsion of twisted nonuniform rotor blades, NACA Report 1346, 1958.
57. V.V. Bolotin, The Dynamic Stability of Elastic Systems, Holden-Day, San Francisco, 1964.
58. C.S. Hsu and W.-H. Cheng, Steady state response of a dynamical system under combined parametric and forcing excitations, Journal of Applied Mechanics, 41, pp. 371–378 (1974).
59. D.V. Nguyen, Interaction between parametric and forced oscillations in multidimensional systems, Journal of Technical Physics, 16, pp. 213–225 (1975).
60. H. Trogerand and C.S. Hsu, Response of a nonlinear system under combined parametric and forcing excitation, ASME Journal of Applied Mechanics, 44, pp. 179–181 (1977).
61. N. HaQuang, D.T. Mook, and R.H. Plaut, Non-linear structural vibrations under combined parametric and external excitations, Journal of Sound and Vibration, 118, pp. 291–306 (1987).
62. N. HaQuang, D.T. Mook, and R.H. Plaut, A non-linear analysis of the interactions between parametric and external excitations, Journal of Sound and Vibration, 118, pp. 425–439 (1987).
63. R.H. Plaut, J.J. Gentry, and D.T. Mook, Non-linear structural vibrations under combined multi-frequency parametric and external excitations, Journal of Sound and Vibration, 140, pp. 381–390 (1990).
64. P.H. Nguyen and J.H. Ginsberg, Vibration control using parametric excitation, ASME Journal of Vibration and Acoustics, 123, pp. 359–364 (2001).
65. E. Mettler, Biegeschwingungen eines Stabes mit kleiner vorkrummung, exzentrisch angreifender publsierender Axiallast und statischer Querbelastung, Forshungshefte aus d. Geb. d. Stahlbaues, 4, pp. 1–23 (1941).
66. F.P. Mallinder and B.A. Proctor, Elastic constants of fused silica as a function of large tensile strain, Physics and Chemistry of Glasses, 5, pp. 91–103 (1964).
67. J.T. Krause, L.R. Testardi, and R.N. Thurston, Deviations from linearity in the dependence of elongation upon force for fibers of simple glass formers and of glass optical lightguides, Physics and Chemistry of Glasses, 20, pp. 135–139 (1979).
68. G.S. Glaesemann, S.T. Gulati, and J.D. Helfinistine, Effect of strain and surface composition on Young's modulus of optical fibers, 11th Optical Fiber Comm. Conference, Technical Digest, TUG5, 26, 1988.
69. E. Suhir, The effect of the nonlinear stress-strain relationship on the mechanical behavior of optical glass fibers, ASME Journal of Electronic Packaging, 114(2) (1992).
70. E. Suhir, Elastic stability, free vibrations, and bending of optical glass fibers: effect of the nonlinear stress-strain relationship, Applied Optics, 31, pp. 5080–5085 (1992).
71. E. Suhir, Effect of the nonlinear stress-strain relationship on the maximum stress in silica fibers subjected to two-point bending, Applied Optics, 32, pp. 1567–1572 (1993).
72. S.K. Sheem and T.G. Giallorenzi, Single-mode fiber-optical power divider: encapsulated etching technique, Optical Letters, 4, p. 29 (1979).
73. R.A. Bergh, G. Kotler, and H.J. Shaw, Single-mode fibre optic directional coupler, Electronics Letters, 16, pp. 260–261 (1980).
74. C.A. Villarruel and R.P. Moeller, Fused single-mode fibre access couplers, Electronics Letters, 17, pp. 243–244 (1981).
75. F. Bilodeau, et al., Compact, low-loss fused biconical taper couplers: overcoupled operation and antisymmetric supermode cutoff, Optical Letters, 12, pp. 634–636 (1987).

PHYSICAL DESIGN

1

Analytical Thermal Stress Modeling in Physical Design for Reliability of Micro- and Opto-Electronic Systems: Role, Attributes, Challenges, Results

E. Suhir

University of California, University of Maryland, and ERS/Siloptix Co., Los Altos, CA 94024, USA

"*Mathematical formulas have their own life, they are smarter than we, even smarter than their authors, and provide more than what has been put into them*"
Heinrich Hertz, German Physicist

"*If my theory is in conflict with the experiment, I pity the experiment*"
Friedrich Hegel, German Philosopher

"*A formula longer than three inches is most likely wrong*"
Unknown Reliability Engineer

1.1. THERMAL LOADING AND THERMAL STRESS FAILURES

Various areas of engineering differ, from the Structural Analysis and Structural Reliability point of view, by the employed materials, typical structures used, and the nature of the applied loads. The most typical microelectronic (ME) and optoelectronic (OE) structures are bodies made of a large variety of dissimilar materials. The most typical loads are thermal loads. These are caused by CTE mismatch and/or by temperature gradients [1–8].

Thermal loading takes place during the normal operation of the system, as well as during its fabrication, testing, or storage. Thermal stresses, strains and displacements are the major contributor to the finite service life and elevated failure rate of ME and OE equipment. Examples are ductile rupture, brittle fracture, thermal fatigue, creep, excessive deformation or displacement, stress relaxation (that might lead to excessive displacements), thermal shock, stress corrosion.

Elevated thermal stresses and strains can lead not only to structural ("physical") failure, but also to functional (electrical or optical) failure. If the heat, produced by the chip, cannot readily escape, then the high thermal stress in the IC can result in failure of the p-n junction [9]. Low temperature microbending (buckling of the glass fiber within the low modulus primary coating) in dual-coated optical fibers, although might be too small

to lead to appreciable bending stresses and delayed fracture ("static fatigue"), can result in appreciable added transmission losses. Loss in optical coupling efficiency can occur, when the displacement in the lateral (often less than 0.2 micrometers) or angular (often less than a split of one percent of a degree) misalignment in the gap between two light-guides or between a light source and a light-guide becomes too large, because of thermally induced deformations or because of thermal stress relaxation in a laser weld. Small lateral or angular displacements in MEMS-based photonic systems (such as, say, some types of tunable lasers) can lead to a complete optical failure of the device. Tiny temperature-change-induced changes in the distance between Bragg gratings "written" on an optical fiber can be detrimental to its functional performance. For this reason thermal control of the ambient temperature is often needed to ensure sufficient protection provided to an optical device, whose performance is sensitive to the change in temperature.

As a matter of fact, the requirements for the mechanical behavior of the materials and structures in OE are often based on the functional (optical) performance of the device/system, rather than on its mechanical (structural) reliability. The requirements for the structural reliability might be much less stringent.

The thermally induced stresses and displacements in ME and OE systems can be linearly or nonlinearly elastic (reversible) or plastic (residual, irreversible), or can be caused by time dependent effects, such as creep, stress relaxation, visco-elastic or visco-plastic phenomena, aging, etc.

The ability to understand the sources of the thermal stresses and strains in ME and OE structures is of significant practical importance, and so is the ability to predict/model/simulate and possibly minimize, if necessary, the induced stresses and displacements.

1.2. THERMAL STRESS MODELING

Thermally induced failures in ME and OE equipment can be prevented only if predictive modeling is consistently used in addition (and, desirably, prior) to experimental investigations and reliability testing [10,11]. Such testing could be carried out on the design (product development) stage, during qualification and manufacturing of the product (qualification testing), or during accelerated or highly accelerated life testing (ALT and HALT).

Accelerated testing, which is the major experimental approach in ME and OE, cannot do without simple and meaningful predictive models. It is on the basis of these models that a reliability engineer decides which parameter should be accelerated, how to process/interpret the experimental data and how to bridge the gap between what one "sees" as a result of accelerated testing and what he/she will supposedly "get" in the actual use condition.

Modeling is the basic approach of any science, whether "pure" or applied [12]. Research and engineering models can be experimental or theoretical. Experimental models are typically of the same physical nature as the actual phenomenon or the object. Theoretical models represent real phenomena and objects by using abstract notions. Such models typically employ more or less sophisticated mathematical methods of analysis, and can be either analytical ("mathematical") or numerical (computational). The today's numerical models are, as a rule, computer-aided, and finite-element analyses (FEA) are widely used in the stress–strain evaluations and physical design of ME and OE structures [13]. Experi-

mental and theoretical models should be viewed, of course, as equally important and indispensable tools for the design of a viable, reliable and cost-effective product [10,11,14,15].

1.3. BI-METAL THERMOSTATS AND OTHER BI-MATERIAL ASSEMBLIES

Pioneering work in modeling of thermal stress in bodies comprised of dissimilar materials was carried out by Timoshenko [16] and Aleck [17]. Timoshenko based his treatment of the problem on a structural analysis (strength-of-materials) approach. Aleck applied theory-of-elasticity method. Both approaches were later extended in application to structures employed in various fields of engineering, including ME and OE [19–45]. Chen and Nelson [18], Chang [19], Suhir [20–22] used structural analysis approach, although the interfacial compliance introduced by Suhir was evaluated based on a theory-of-elasticity method [20,21]. Zeyfang [23], Eischen et al. [24], Kuo [25], Yamada [26] and others used the theory-of-elasticity treatment of the problem.

The application of the structural analysis approach [27] enables one to determine, often with sufficient accuracy and always with extraordinary simplicity, the stresses acting in the constituent materials, as well as the interfacial shearing and through-thickness ("peeling") stresses. This approach results in closed form solutions and in easy-to-use formulas (see, for instance [20]). It can be (and, actually, has been) effectively employed as a part of the physical design process to select the appropriate materials, establish the feasible dimensions of the structural elements, compare different designs from the standpoint of the induced stresses and deformations, etc. On the other hand, the theory-of-elasticity method is based on rather general hypotheses and equations, and provides a rigorous treatment of the problem. Typically, it requires, however, additional use of computers to obtain the final solution to the given problem. The theory-of-elasticity approach is advisable, when there is a need for the most accurate evaluation of the induced stresses. Applied within the framework of linear elasticity, this approach leads, in the majority of cases, to a singularity at the assembly edges or at the corner of a structural element. For this reason its application has been found particularly useful when there is intent to further proceed with fracture analysis of interfacial delaminations, crack initiation and propagation at the corners, etc.

The structural analysis (strength-of-materials) and theory-of-elasticity approaches should not be viewed as "competitors," but rather as different tools, which have their merits and shortcomings, and their areas of application. These two analytical approaches should complement each other in any comprehensive engineering analysis and physical design effort.

1.4. FINITE-ELEMENT ANALYSIS

Finite-element analysis (FEA) has become, since the mid-1950s, the major resource for computational modeling in engineering, including the area of ME and OE (see, for instance, Lau [28], Glaser [29], Akay and Tong [30]). The today's powerful and flexible FEA computer programs enable one to obtain, within a reasonable time, a solution to almost any stress–strain-related problem.

Broad application of computers has, however, by no means, made analytical solutions unnecessary or even less important, whether exact, approximate, or asymptotic. Simple analytical relationships have invaluable advantages, because of the clarity and "compactness"

of the obtained information and clear indication of the role of various factors affecting the given phenomenon or the behavior of the given system. These advantages are especially significant when the parameter under investigation depends on more than one variable. As to the asymptotic techniques and formulas, analytical modeling can be successful in those cases, in which there are difficulties in the application of computational methods, e.g., in problems containing singularities.

But, even when the application of numerical methods encounters no significant difficulties, it is always advisable to investigate the problem analytically before carrying out computer-aided analyses. Such a preliminary investigation helps to reduce computer time and expense, develop the most feasible and effective preprocessing model and, in many cases, avoid fundamental errors. Those that have a hands-on experience in using FEA, know very well that it is easy to obtain *a* solution based on the FEA software, but it might be not that easy to obtain *the* right solution.

Preliminary analytical modeling can be very helpful in creating a meaningful and economic preprocessing simulation model. This is particularly true in OE and photonics, where high accuracy is usually required. Special attention should be paid and special effort should be taken to make the existing FEA programs accurate enough to be suitable for the evaluation of the tiny thermo-mechanical displacements in an OE or a photonic system. Another challenge has to do with the necessity to consider visco-elastic and time-dependent behavior of photonic materials, so that the long-term reliability of the device is not compromised.

It is noteworthy that FEA was originally developed for structures with complicated geometry and/or with complicated boundary conditions, when it might be difficult to apply analytical approaches. Consequently, FEA is especially widely used in those areas of engineering, in which structures of complex configuration are typical: aerospace, marine and offshore structures, some complicated civil engineering structures, etc. In contrast, a relatively simple geometry and simple configurations usually characterize ME and OE assemblies and structures. Owing to that, such structures can be easily idealized as beams, flexible rods, circular or rectangular plates, frames, or composite structures of relatively simple geometry, thereby lending themselves to analytical modeling.

1.5. DIE-SUBSTRATE AND OTHER BI-MATERIAL ASSEMBLIES

The mechanical behavior of bonded bi-material assemblies, and particularly die-substrate assemblies, was addressed in numerous studies [31–40]. Typical failure modes in die-substrate assemblies are [31]:

(1) adherend (die or substrate) failure: a silicon die can fracture in its midportion or at its corner located at the interface;
(2) cohesive failure of the bonding material (i.e., failure in the bulk of the die-attach material); and
(3) adhesive failure of the bonding material (i.e., failure at the adherend/adhesive interface).

An adhesive failure is not expected to occur in a properly fabricated joint. If such a failure takes place, it usually occurs at a very low load level, at the product development stage, and should be regarded as a manufacturing or a quality control problem, rather than a material's or structural one.

A crack on the upper ("free") surface of the die is typically due to the normal stress acting in the die cross-sections. This stress is more or less uniformly distributed throughout the die and drops to zero at its ends. The crack at the die's corner at its interface with the substrate should be attributed to the interfacial (shearing and "peeling") stresses. These stresses concentrate, for sufficiently long assemblies with stiff enough interfaces, at the assembly ends and are next-to-zero in its midportion.

Measures that could be taken to bring the induced stresses down depend on the type/category of the stress responsible for the particular failure mode. In the case of a crack in the midportion of the die, it is the improved thermal match between the die and the substrate materials and/or a lower bow of the assembly that can improve the situation. In the case of a crack at the die's corner, the employment of a thicker and/or lower modulus adhesive can be helpful.

Die-substrate assemblies, as well as many other bi-material assemblies, are characterized by a substantial thermal expansion (contraction) mismatch of the adherend (silicon and the substrate) materials, as well as by thin and low modulus adhesive (die-attach) layers, compared to the thickness and Young's moduli of the adherends. Such a situation results in the fact that the attachment (die-attach) material experiences shear only, and also in the fact that only the interfacial compliance of the adhesive (die-attach) layer, and not its coefficient of expansion, is important [20]. It has been shown also [20,22,31] that thermally induced elastic stresses in sufficiently large bi-material assemblies do not increase with a further increase in the assembly size, and that a substantial relief in the interfacial stresses can be achieved by using thick and low modulus adhesives. This can result also in an improved adhesive and cohesive strength of the adhesive (die-attach) material, and reduce the likelihood of occurrence of the brittle crack at the chip's corner.

In the case of small-size assemblies (e.g., those with chips, not exceeding, say, 5 mm), thick and low modulus adhesives can lead to the decrease in the stress in the chip itself as well. For large chips, however (larger than, say, 10 mm), other measures should be taken, if there is a need to bring down the stresses in the midportion of the die: a substrate material with a better match with silicon should be used; a die-attach material with a lower curing temperature (and/or lower glass transition temperature) could be employed; application of a flexible substrate could be considered, etc.

As to the thermal stress modeling in bi-material assemblies, it has been found [32] that the approach, previously used in application to a thin film structure [33], can be successfully employed to further simplify thermal stress prediction in a bi-material assembly as well. This approach suggests that the interfacial shearing stress can be evaluated using an assumption that this stress is not affected by (not coupled with) the "peeling" stress. Such an assumption is conservative, i.e., results in a reasonable overestimation of the maximum shearing stress compared to the stress level obtained from the coupled equations [21]. After the shearing stress function is determined, the "peeling" stress can be computed from an equation that is similar to the equation of bending of a beam lying on a continuous elastic foundation [27].

The developed models were applied to many problems in ME and OE and beyond. The model suggested in [20] was applied by Hall et al. [34] and other researchers [35] to tri-material assemblies. An assembly should be treated as a tri-material one (as opposite to an adhesively bonded assembly), if the adhesive layer in it is not thin and/or if its Young's modulus is not significantly lower than the Young's modulus of the adherend materials. In such a situation the CTE of the bonding material has to be accounted for. An analytical model for a tri-material assembly, in which all the materials are treated as "equal

constituents/members" of the assembly, i.e., in which the geometries (thicknesses) of all the assembly components and material's properties (elastic constants and CTEs) of all the materials are important, was developed in [36].

Luryi and Suhir [33] applied the model developed in [20] to the case of semiconductor crystal growth. They suggested a new approach to the high quality (dislocation free) epitaxial growth of lattice-mismatched materials. The authors have shown that lattice-mismatch strain can be simply added to the thermal-mismatch strain and, owing to that, can be easily and naturally incorporated into stress analysis models developed earlier for thermally induced strains. This paper has triggered a substantial experimental effort (numerous citations of it could be found in the literature) and is still widely referenced in the physical literature (see, for instance, [38]).

Suhir and Sullivan [39] have developed an axisymmetric version of the model suggested in [20] and applied it for the evaluation of the adhesive strength of epoxy molding compounds used in plastic packaging of IC devices.

Suhir [22] and Cifuentes [40] addressed plastic and elastoplastic deformations in the solder layer and in the beams experiencing thermal loading.

1.6. SOLDER JOINTS

Numerous models have been developed for the evaluation of thermal stresses in, and prediction of the lifetime of, solder joint interconnections (see, for instance, [41–46]). Typically, the stresses in the solder joints are caused by the thermal expansion (contraction) mismatch of the chip and the substrate materials, or, in assemblies of ball-grid-array (BGA) type, by the mismatch of the package structure and the PCB (system's substrate).

The majority of the suggested models are based on the prediction and improving of the solder joint fatigue, which is caused by the accumulated cyclic strain in the solder material. This strain is due to the temperature fluctuations resulting from either the changes in the ambient temperature (temperature cycling) or from heat dissipation in the package (power cycling). Various (viscoplastic) models for the prediction of the fatigue life-time of the solder material were suggested by Akay and Tong [30], Morgan [41], Hwang [42], Ianuzzelli, Pittaresi and Prakash [43] and many others.

The ultimate strength of solder joint interconnections is typically measured by using shear-off tests. A new, "twist-off," technique for testing of solder joint interconnections was suggested in [44]. It enables one to mimic best the actual state of stress of such interconnections. The technique was developed in application to flip-chip (FC) and ball-grid-array (BGA) assemblies.

One effective way to bring down the thermal stresses in solder joints is by employing a flex circuit [45,46]. The developed models can be used to assess the incentive in the application of such circuits, as well as the expected stress relief. Juskey and Carson [47] suggested that flex circuits be used as carriers for the direct chip attachment (DCA) technology. Flex circuitry offers a low cost and reliable system with a low thermal stress level. The flexible material of choice for today's manufacturing environment is polyamide. This material is able to withstand high temperatures during reflow soldering, and possesses good electrical and mechanical characteristics.

Solder materials and solder joints are as important in photonics, as they are in microelectronics. There are, however, a number of specific requirements for the photonics solder materials and joints: ability to achieve high alignment, requirement for a low creep,

etc. [48]. “Hard” (high modulus) solder materials (such as, say, gold-tin eutectics) are thought to have better creep characteristics than “soft” (such as, say, silver-tin) solders. It should be pointed out, however, that “hard” solders can result in significantly higher thermally induced stresses than “soft” solders [49], and therefore their ability to withstand creep might be not as good as expected, not to mention the short-term reliability of the material.

Thermally induced stresses in optical fibers soldered into ferrules were modeled in [49]. Modeling was based on the solution to the axisymmetric theory-of-elasticity problem for an annular composite structure comprised of the metalized silica fiber, the solder ring, and the ferrule. The obtained relationships enable one to design the joint in such a way that the solder ring is subjected to relatively low compressive stresses. It has been shown that neither low expansion ferrules, nor high expansion ones, might be suitable for a particular solder material and particular thickness of the solder ring.

Solders are often used as continuous attachment layers in ME and OE assemblies. Stress concentration at the ends of such attachments can lead to plastic deformations of the solder material. These deformations were addressed by Suhir [22] and Cifuentes [40]. This problem has become recently of significant importance in connection with using Indium or Indium-based alloys as suitable attachments of the quantum wells of GaAs lasers to metal substrates. The developed models enable one to assess the size of the zone, in which the plastic deformations are possible, and to assess the effect on the state of the inelastic strain in the device. The plastic stresses will not propagate inwards the assembly, if its length exceeds appreciably the total length of the areas occupied by the elevated stresses. This consideration provides a practically useful criterion for the selection, if possible, of the length of the continuous solder layer in the application in question.

1.7. DESIGN RECOMMENDATIONS

Based on the modeling of thermal stresses in typical adhesively bonded or soldered assemblies, i.e., in assemblies with appreciable CTE mismatch of the adherends and a homogeneous adhesive or solder layer, the following general recommendations, aimed at the improvement of the ultimate and fatigue strength, of the assemblies, have been developed:

- equalize the in-plane and bending stiffness of the adherends, and use identical adherends, if possible;
- use low modulus adhesives; as an alternative to using a low modulus adhesive throughout the joint, use such an adhesive only at the ends of the joint, i.e., in the region of high interfacial stresses, while a higher modulus adhesive could be used in the midportion of the assembly;
- vary, if possible, the adherend thickness along the assembly in a proper way and/or slant the adherends edges for a thicker adhesive layer at the assembly ends;
- keep the stresses within the elastic range, if possible;
- minimize “peeling” (in the case of multimaterial and thin film structures) and axial (in the case of solder joints) stresses.

1.8. "GLOBAL" AND "LOCAL" MISMATCH AND ASSEMBLIES BONDED AT THE ENDS

In those cases when the adhesive (solder) layer is not homogeneous, or when the components are just partially bonded or soldered to each other, both "global" and "local" mismatch loading takes place [50–53]. The "local" mismatch loading is due to the mismatch of the dissimilar materials within the bonded or soldered region, while the "global" mismatch loading is caused by the mismatch in the unbonded region. Examples are: solder joint interconnections, optical glass fiber interconnects adhesively bonded or soldered at their ends into a ferrule or a capillary; optical glass fibers in micromachined (MEMS) optical switches packaged into a dual-in-line package, etc.

The interaction of the interfacial shearing stresses caused by the "global" and "local" mismatches in a typical bi-material assembly adhesively bonded or soldered at the ends [51–53] can be qualitatively summarized as follows:

- The interfacial shearing stresses caused by the "local" mismatch are antisymmetric with respect to the mid-cross-section of the bonded area: these stresses are equal in magnitude and opposite in directions (signs).
- The "local" shearing stresses concentrate at the ends of the bonded area and, for sufficiently long bonded joints and stiff interfaces (thin and high-modulus adhesive layer), are next-to-zero in the midportion of the bonded area.
- For short-and-compliant bonded areas, the "local" shearing stresses are more-or-less linearly distributed over the length of the bonded area, and their maxima at the assembly ends can be significantly lower than the maximum stresses in long-and-stiff bonded joints.
- The shearing stresses caused by the "global" mismatch act in the same direction over the entire length of the bonded joint. This direction is such, that in the inner portions of the joints (i.e., in the portions located closer to the mid-cross-section of the unbonded region, which is also the mid-cross-section of the assembly as a whole), the total interfacial stress should be computed as the difference between the "local" and the "global" stress. In the outer portions of the bonded joints, the total stress should be computed as the sum of the "local" and the "global" stress.
- In the case of short-and-compliant joints, when both the "local" and the "global" stresses are more or less uniformly distributed over the joint's length, the total stress is indeed larger than each of these stress categories. Since, however, both the "local" and the "global" stresses in short-and-compliant joints can be very low compared to the stresses in long-and-stiff assemblies, the total stress can be very low as well, despite the fact that, for the outer (peripheral) portions of the bonded joints, this stress is obtained as a sum of the "local" and the "global" stresses.
- In the case of long-and-stiff joints, the "global" stresses concentrate at the inner edges of the bonded joints and rapidly decrease with an increase in the distance of the given cross-section from these edges. In such a situation, the interaction of the "local" and "global" stresses is always favorable, i.e., at the inner edge, results in the total stress, obtained as a difference between the "local" and the "global" stress, and, at the outer edge, is due to the "local" stress only.
- For sufficiently long-and-stiff bonded joints, the magnitude of the "global" stress at the inner boundary is equal to the magnitude of the maximum "local" stress, so that the total shearing stress is zero.

Although substantial relief in the total stress can be achieved by employing bonded joints with short-and-compliant attachments, this approach usually cannot be recommended, because insufficient "real estate" of the bonded areas does not allow one to produce reliable enough joints. However, in the case of solder joint interconnections, both the favorable effect of the short-and-compliant joint and the unfavorable effect of the summation of the "local" and the "global" thermal stresses/strains at the assembly ends should be considered.

The interaction of the "local" and the "global" stresses, with consideration of the effect of the coefficient of thermal expansion (contraction) of the epoxy material itself, was studied in [51] for a glass fiber interconnect whose ends are epoxy bonded into capillaries. The necessity of taking into account the CTE of the adhesive material was due to the fact that the cross-sectional area of the adhesive ring was considerably larger than the cross-sectional area of the glass fiber. Therefore the longitudinal compliance of the adhesive ring was comparable with the compliance of the fiber and could not be neglected.

Understanding of the interaction of the "global" and "local" stresses is particularly important in connection with the ME and OE assemblies bonded at the ends. In some ME assemblies of the flip-chip type the solder joint stand-off is so low that it is practically impossible to bring in the underfill material underneath the chip, especially if the chip is large. On the other hand, there might be no need for that, since the underfill material works effectively only at its peripheral portions [51,52]. Modeling of the mechanical behavior of such an assembly is crucial in order to establish the adequate width of the adhesive layer: this width should be large enough to provide sufficient bonding strength of the assembly. The stresses in such an assembly will not be higher than in an assembly with a continuous underfill.

1.9. ASSEMBLIES WITH LOW MODULUS ADHESIVE LAYER AT THE ENDS

Interfacial shearing and peeling stresses in adhesively bonded or soldered assemblies concentrate at the assembly ends. These stresses can be reduced by employing a low modulus material at the assembly ends [54] and/or by slanting the edges of the assembly components [55], thereby increasing the thickness of the adhesive layer at the assembly ends. The stresses at the ends of polymer-coated optical fibers can be reduced by using a low modulus coating at the fiber ends [56,57]. The mechanical behavior of such ME and OE structures is, in a sense, opposite to the situation that takes place in an assembly adhesively bonded at the ends. Indeed, in an assembly with a low modulus adhesive/coating at the end, it is the midportion of the assembly that is characterized by an elevated Young's modulus of the adhesive (coating), while in the case of an assembly bonded at the ends, it is its midportion that is characterized by a "low" (actually, zero) Young's modulus of the "attachment."

1.10. THERMALLY MATCHED ASSEMBLIES

There is an obvious incentive to employ thermally matched materials in ME and OE assemblies. It is this assembly, which is used in a Si-on-Si flip-chip (FC) design [58,59], in a ceramic Cerdip/Cerquad ME package [60]. On the other hand, some photonics structures (say, holographic memory assemblies) are made of identical adherends and a compliant (thick and low modulus) adhesive [61–63].

There is substantial difference in the mechanical behavior of the assemblies with an appreciable mismatch in the CTE of the adherends and the thermally matched assemblies,

particularly those with identical adherends. While in assemblies with an appreciable thermal mismatch of the adherends, the CTE of the adhesive material (as long as this material is thin and/or has a low Young modulus) does not affect the mechanical behavior of the assembly, in assemblies with identical adherends the mismatch between the adhesive and the adherends' materials the CTE of the adhesive is definitely important. In addition, the mechanical behavior and reliability of the adhesive material is quite different. In assemblies with mismatched adherends, and thin and low modulus adhesives, the adhesive layer is primarily subjected to the interfacial shear, while in the case of matched assemblies (identical adherends) the adhesive layer elevated experiences both shearing and performs similarly a thin film fabricated on a thick substrate and tensile (compressive) stresses.

Thermal stresses in solder joints in thermally matched silicon-on-silicon flip-chip assemblies achieve their maximum values at the interfaces and concentrate at the joints' corners [58,59]. The stresses, acting in the axial direction (these stresses are analogous to the "peeling" stresses in thin film structures), are the highest, and it is these stresses and strains that are primarily responsible for the joint's reliability.

The case of identical ceramic adherends was considered in connection with choosing an adequate coefficient of thermal expansion for a solder (seal) glass in a ceramic package design [60]. It has been found that the best result can be achieved by using a probabilistic approach, in which the coefficient of thermal expansion of the solder glass is treated as a random variable. The package manufactured in accordance with the developed recommendations exhibited no failures. Based on the performed analysis, it has been concluded that in order to successfully apply a probabilistic approach (see, for instance, [64]) customers should require that vendors provide information concerning both the mean value and the standard deviation of the parameter of interest.

Several thermoelastic models [61–63] were developed for the prediction of the mechanical behavior of the adhesive material in adhesively bonded assemblies with identical nondeformable adherends of different configurations. The analyses were carried out in application to assemblies used in advanced holographic memory devices. It has been shown, particularly, that the interfacial compliance of the adhesive layer, in the case of sufficiently large-and-thin assemblies with thermally matched adherends, is half the magnitude of the interfacial compliance in the case of assemblies with mismatched adherends. It has been shown also that the elevated interfacial shearing stresses are somewhat higher for a circular assembly than for a rectangular one. These stresses also occupy a narrower zone around the assembly edge. An inhomogeneous adhesive layer, which is important for the considered application, was examined. The developed models enable one to establish the conditions, at which the requirement for the undistorted boundaries of the inner "pieces" of the adhesive is fulfilled. This requirement is important from the stand point of the satisfactory optical performance of the assembly.

1.11. THIN FILMS

Typical thermal stress failures in thin films fabricated on thick substrates are interfacial delaminations (including delamination buckling), and film cracking and blistering. Numerous investigators [65–71] analyzed thermal stresses in thin films. Based on the obtained results, practical recommendations for a physical design of a reliable thin film structure have been formulated. Particularly, it has been found that

- the thermal stress in the given film layer of a multilayer film structure is due to the thermal expansion mismatch of this layer with the substrate, and not with the adjacent film layers [68];

- the edge stresses in the film are affected by the edge configuration [70]: circular assemblies are somewhat "stiffer" than the rectangular ones, i.e., result in higher stresses that concentrate at a narrower peripheral ring;
- stress in a thin film, which does not experience bending stresses, is not affected by the assembly bow, while the assembly bow and the stresses in the substrate are strongly affected by the stresses in the film [71].

The effect of lattice mismatch of semiconductor materials during crystal growth of thin Germanium films on a thick Silicon substrate was addressed, along with the effect of thermal mismatch, by Luryi and Suhir [37]. It has been shown that by using a "tower-like" surface of the substrate (such a surface can be achieved by high-resolution lithography, by employment of porous silicon, etc.) one can grow dislocation free semiconductor films.

1.12. POLYMERIC MATERIALS AND PLASTIC PACKAGES

Polymeric materials are widely used in ME and OE engineering (see, for instance, [72–78]). Examples are: plastic packages of integrated circuit (IC) devices, adhesives, various enclosures and plastic parts, polymeric coatings of optical silica fibers, and even polymeric lightguides. There are numerous and rapidly growing opportunities for the application of polymers for diverse functions in the "high-technology" field. Polymeric materials are inexpensive and lend themselves easily to processing and mass production techniques. The reliability of these materials, however, is usually not as high as the reliability of inorganic materials and is often insufficient for particular applications, thereby limiting the area of the technical use of polymers. There exists a crucial necessity for the advancement of the experimental and theoretical methods, techniques and approaches, aimed at the prediction and improvement of the short/long-term performance of polymeric materials for various ME and OE applications.

Recent improvements in the mechanical properties of molding compounds, plastic package designs, and manufacturing technologies have resulted in substantial increase in the reliability of plastic packages. There is, however, one major industry-wide concern associated with these packages—their moisture-induced failures ("popcorn" cracking). Such failures typically occur during surface mounting the packages onto printed circuit boards by means of high temperature reflow soldering. "Popcorn" cracking is usually attributed to the elevated pressure of the water vapor, generated due to a sudden evaporation of the absorbed moisture [73]. It is believed that thermal stresses also play an important role, both directly, due to their interaction with mechanical vapor-pressure-induced stresses in the underchip portion of the molding compound, and indirectly, by triggering the initiation and facilitating the propagation of the interfacial delaminations.

It has been suggested [74] that constitutive equations, obtained as a generalization of von-Karman's equations for large deflections of plates, be used as a suitable analytical stress model for the prediction and prevention of structural failures in moisture-sensitive plastic packages. Such a generalization accounts for the combined action of the lateral pressure, caused by the generated water vapor, and the thermally induced loading. The developed model can be used for the selection of the low stress molding compounds, for comparing different package design from the standpoint of their propensity to "popcorn"-cracking, in the development of "figures-of-merit" [75], which would enable one to separate packages that need to be "baked" and "bagged" from those that do not, etc. This loading is due to both the temperature gradients and the thermal expansion (contraction)

mismatch of the dissimilar materials in the package. Since the coefficient of thermal expansion and Young's modulus of the molding compound are temperature dependent, the constitutive equations account for this dependence. The developed equations were applied to the delaminated underchip layer of the molding compound. This layer is treated as a thin rectangular plate clamped at the support contour. It has been shown, in particular, that, from the standpoint of structural analysis, the distinction between "thick" and "thin" packages should be attributed primarily to the level of the in-plane ("membrane"), normal stresses in the underchip portion of the compound: in "thick" packages this portion exhibits bending only, while in "thin" packages it is subjected to both bending and in-plane loading. The obtained data, which are in good agreement with experimental observations, have indicated that the geometric characteristics of the package (the underchip layer thickness, chip and paddle size, etc.) have a strong effect on the package propensity to failure.

The obtained results have been used to develop guidelines ("figures-of-merit"), which enable one to separate packages that need to be "baked" and "bagged" from those that do not, as well as for guidelines aimed at the preliminary selection of the feasible molding compound for the given package design.

1.13. THERMAL STRESS INDUCED BOWING AND BOW-FREE ASSEMBLIES

Thermal stress induced bowing can prevent further processing of BGA packages or of thin (TSOP) plastic packages [79], can lead to cracking of ceramic substrates in thin overmolded packages [80,81], or can have another adverse effect on the design or processing of plastic packages of IC devices. It has been shown [79–81] that employment of additional (surrogate) layers can dramatically improve the situation.

There is an obvious incentive for the use of bow-free (temperature change insensitive) assemblies in ME and OE packaging. It has been shown [83,84] that this can be achieved if a thick enough bonding layer is introduced to produce an appreciable axial (in-plane) force. This force is necessary to create a bending moment that would be able to equilibrate the thermally induced moment produced by the dissimilar adherends. A statically determinate bi-material assembly (i.e., an assembly with a very thin and/or a very low-modulus bonding layer) cannot be made bow free. The thermally induced forces acting in the components of a bi-material assembly are equal in magnitude and opposite in sign, and create a bending moment that can be equilibrated by the elastic moment only. This inevitably leads to non-zero deflections, whether large or small. To be bow-free, a multi-material assembly should be made statically indeterminate. It should contain, therefore, at least three dissimilar materials, so that the resulting bending moment, caused by the induced forces in all the three materials, is zero.

A sufficiently large axial force in the bonding material can be created by one or a combination of two or more of the following measures:

- by using a bonding material with a high elastic modulus;
- by using a bonding materials with a significant thermal mismatch with the adherends;
- by using a bonding material with a high curing temperature, and/or
- by making the bonding material thick.

It is only the last measure, however, that, while resulting in a desirable elevated thermally induced force in the bonding material, does not necessarily lead to an elevated axial stress in it. Computations based on the developed analytical models [83,84] have indicated that the "thick" bonding layer in a bow free assembly can still be made thin enough (about 4 mils or so) to be effective, provided that the material and/or the thickness of at least one of the adherends is adequately chosen.

1.14. PROBABILISTIC APPROACH

Probabilistic models might be very useful in situations, in which the "fluctuations" from the mean values are significant and in which the variability, change and uncertainty play a vital role (see, for instance, [60,64,85,86]). In the majority of such situations the product will most likely fail, if these uncertainties are ignored. So far, probabilistic (statistical) models are used in "high-tech" engineering primarily for the design and analyses of experiments. They are very seldom used yet as a physical design tool. In this connection we would like to emphasize that wide and consistent use of probabilistic models would not only enable one to establish the scope and the limits of the application of deterministic solutions, but can provide a solid basis for a well-substantiated and goal-oriented accumulation, and effective utilization of empirical data. Probabilistic models enable one to quantitatively assess the degree of uncertainty in various factors, which determine the performance of a product. Then a reliability engineer can design a product with a predictable and low probability of failure. A good illustration to these statements is the success of the design described in [60]).

1.15. OPTICAL FIBERS AND OTHER PHOTONIC STRUCTURES

Various problems of the thermal stress modeling in bare and coated optical silica fibers were addressed in [3,7,87–98].

Low temperature microbending can result in substantial added transmission losses in dual-coated optical fibers. Based on the developed analytical stress models [90], it has been shown that the initial curvatures can play an important role in the low temperature behavior of a dual-coated silica fiber and that certain curvature lengths are less favorable than others from the standpoint of the possible fiber buckling. It has been shown also [91] that the magnitude of the spring constant of the elastic foundation provided by the primary coating layer could have a significant effect on the buckling conditions, and that, in the case of thick and relatively low modulus secondary coatings, both coating layers should be considered when evaluating the spring constant. For thin and high modulus secondary coatings, however, only the primary coating material could be considered when evaluating the spring constant of the elastic foundation for the silica glass fiber.

Application of a mechanical approach to the evaluation of low temperature added transmission losses [92] enables one, based on the developed analytical stress model, to evaluate the threshold of the low temperature added transmission losses from purely mechanical calculations, without resorting to optical calculations or measurements. The model (confirmed by actual optical measurements) presumes that the threshold of the elevated added transmission losses coincides with the threshold of the elevated thermally induced stresses applied by the coating (jacket) to the silica fiber.

Various aspects related to thermal stresses and thermal stress related behavior of solder materials for photonics applications were addressed in [53]. Thermal stresses in, and optimal physical design of, solder joints for metalized optical fibers soldered into ferules were analyzed, based on the developed analytical stress model, in [54]. It has been demonstrated that an adequate modeling of what could be expected in an actual joint is a must: the selection of the right enclosing material and the right thickness of the solder preform (for the given solder material) should be conducted and decided upon prior to manufacturing of the joint.

Mechanical behavior and elastic stability of bare, polymer coated and metalized optical fiber interconnects was modeled in [93–95]. It has been shown that low modulus polymer coatings have significant advantages over high modulus metalizations, as far as the stresses in the coating (metalization) are concerned, and typically should be preferred despite of their sensitivity to moisture penetration. An analytical stress model developed in [94] enables pone to select the appropriate enclosure material for minimizing the thermally induced bending stresses in an optical fiber interconnect experiencing ends-offset and thermally induced compressive loading because of its mismatch with the enclosure material. In those cases when low buckling stresses are a problem, thicker polymer coatings can be used to improve the elastic stability of a polymer-coated fiber [95].

Suhir and Vuillamin [96] have demonstrated, based on the developed analytical and FEA models, that the gradient in the distribution of the CTE along one of the diameters of a glass fiber cross-section can be responsible for the undesirable "curling" phenomenon that often occurs during drawing of optical silica fibers.

Optical silica fiber materials exhibit highly nonlinear (but still elastic) behavior when subjected to tension or compression. This effect was considered, along with the effect of the nonprismaticity, in the model [97] developed for a fused biconical taper (FBT) coupler experiencing thermally induced tension. An effective method for thermostatic compensation of temperature sensitive devices was suggested in [98] in application to Bragg gratings. It has been shown that there is no need to use, for particular applications, mechanically vulnerable ceramic materials with a negative CTE: regular and more mechanically reliable materials can be successfully used for the objective in question.

It has been recently demonstrated [99,100] that a newly developed nano-particle material (NPM) can make a substantial difference in the state-of-the-art of coated optical fibers: this material has all the merits of the polymer coated and metalized optical fibers without having their drawbacks.

1.16. CONCLUSION

- Predictive modeling is an effective tool for the prediction and prevention of mechanical and functional failures in microelectronics and photonics materials, structures, packages and systems, subjected to thermal loading.
- Experimental and theoretical models should be viewed as equally important and equally indispensable to the design of a viable, reliable and cost-effective product. The same is true for analytical and numerical (FEA) models.
- Special effort should be taken to make the existing FEA program accurate enough to be suitable for the evaluation of the thermal stresses and displacements in photonics structures. In such a situation, analytical modeling of a simplified structure of interest can be very useful for the selection and mastering of the preprocessing FEA model.

- Application of the probabilistic approach enables one to quantitatively assess the role of various uncertainties in the materials properties, geometrical characteristics and loading conditions, and, owing to that, to design and manufacture a viable and reliable product.
- A newly developed nanomaterial can make a substantial difference in the state-of-the-art of coated optical fibers: this material has all the merits of the polymer coated and metalized optical fibers without having their drawbacks.

REFERENCES

Thermal loading and thermal stress failures

1. J.H. Lau, Ed., Thermal Stress and Strain in Microelectronics Packaging, Van-Nostrand Reinhold, New York, 1993.
2. E. Suhir, R.C. Cammarata, D.D.L. Chung, and M. Jono, Mechanical behavior of materials and structures in microelectronics, Materials Research Society Symposia Proceedings, Vol. 226, 1991.
3. E. Suhir, M. Fukuda, C.R. Kurkjian, Eds., Reliability of photonic materials and structures, Materials Research Society Symposia Proceedings, Vol. 531, 1998.
4. E. Suhir, M. Shiratori, Y.C. Lee, and G. Subbarayan, Eds., Advances in Electronic Packaging—1997, Vols. 1 and 2, ASME Press, 1997.
5. E. Suhir, Thermal stress failures in microelectronic components—review and extension, in A. Bar-Cohen and A.D. Kraus, Eds., Advances in Thermal Modeling of Electronic Components and Systems, Hemisphere, New York, 1988.
6. E. Suhir, B. Michel, K. Kishimoto, and J. Lu, Eds., Mechanical Reliability of Polymeric Materials and Plastic Packages of IC Devices, ASME Press, 1998.
7. E. Suhir, Thermal stress failures in microelectronics and photonics: prediction and prevention, Future Circuits International, issue 5, 1999.
8. E. Suhir, Microelectronics and photonics—the future, Microelectronics Journal, 31(11–12) (2000).
9. G.A. Lang, et al., Thermal fatigue in silicon power devices, IEEE Transactions on Electron Devices, 17 (1970).

Thermal stress modeling

10. E. Suhir, Accelerated Life Testing (ALT) in microelectronics and photonics: its role, attributes, challenges, pitfalls, and interaction with qualification tests, Keynote address at the SPIE's 7-th Annual International Symposium on Nondestructive Evaluations for Health Monitoring and Diagnostics, 17–21 March, San Diego, CA, 2002.
11. E. Suhir, Reliability and accelerated life testing, Semiconductor International, February 1, 2005.
12. E. Suhir, Modeling of the mechanical behavior of microelectronic and photonic systems: attributes, merits, shortcomings, and interaction with experiment, Proceedings of the 9-th Int. Congress on Experimental Mechanics, Orlando, FL, June 5–8, 2000.
13. E. Suhir, Analytical stress–strain modeling in photonics engineering: its role, attributes and interaction with the finite-element method, Laser Focus World, May 2002.
14. E. Suhir, Thermomechanical stress modeling in microelectronics and photonics, Electronic Cooling, 7(4) (2001).
15. M. Schen, H. Abe, and E. Suhir, Eds., Thermal and Mechanical Behavior and Modeling, ASME, AMD-Vol, 1994.

Bi-metal thermostats and other bi-material assemblies

16. S. Timoshenko, Analysis of bi-metal thermostats, Journal of the Optical Society of America, 11 (1925).
17. B.J. Aleck, Thermal stresses in a rectangular plate clamped along an edge, ASME Journal of Applied Mechanics, 16 (1949).
18. W.T. Chen and C.W. Nelson, Thermal stresses in bonded joints, IBM Journal, Research and Development, 23(2) (1979).

19. F.-V. Chang, Thermal contact stresses of bi-metal strip thermostat, Applied Mathematics and Mechanics, 4(3) Tsing-hua Univ., Beijing, China (1983).
20. E. Suhir, Stresses in bi-metal thermostats, ASME Journal of Applied Mechanics, 53(3) (1986).
21. E. Suhir, Interfacial stresses in bi-metal thermostats, ASME Journal of Applied Mechanics, 56(3) (1989).
22. E. Suhir, Calculated thermally induced stresses in adhesively bonded and soldered assemblies, Proc. of the Int. Symp. on Microelectronics, ISHM, 1986, Atlanta, Georgia, Oct. 1986.
23. R. Zeyfang, Stresses and strains in a plate bonded to a substrate: Semiconductor devices, Solid State Electronics, 14 (1971).
24. J.W. Eischen, C. Chung, and J.H. Kim, Realistic modeling of the edge effect stresses in bimaterial elements, ASME Journal of Electronic Packaging, 112(1) (1990).
25. A.Y. Kuo, Thermal stress at the edge of a bi-metallic thermostat, ASME Journal of Applied Mechanics, 57 (1990).
26. S.E. Yamada, A bonded joint analysis for surface mount components, ASME Journal of Electronic Packaging, 114(1) (1992).
27. E. Suhir, Structural analysis in microelectronic and fiber optic systems, Vol. 1, Basic Principles of Engineering Elasticity and Fundamentals of Structural Analysis, Van Nostrand Reinhold, New York, 1991.
28. J.H. Lau, A note on the calculation of thermal stresses in electronic packaging by finite-element method, ASME Journal of Electronic Packaging, 111(12) (1989).
29. J.C. Glaser, Thermal stresses in compliantly joined materials, ASME Journal of Electronic Packaging, 112(1) (1990).
30. J.U. Akay and Y. Tong, Thermal fatigue analysis of an smt solder joint using FEM approach, Journal of Microcircuits and Electronic Packaging, 116 (1993).
31. E. Suhir, Die attachment design and its influence on the thermally induced stresses in the die and the attachment, Proc. of the 37th Elect. Comp. Conf., IEEE, Boston, MA, May 1987.
32. V. Mishkevich and E. Suhir, Simplified approach to the evaluation of thermally induced stresses in bi-material structures, in E. Suhir, Ed., Structural Analysis in Microelectronics and Fiber Optics, ASME Press, 1993.
33. E. Suhir, Approximate evaluation of the elastic interfacial stresses in thin films with application to high-Tc superconducting ceramics, Int. Journal of Solids and Structures, 27(8) (1991).
34. P.M. Hall, et al., Strains in aluminum-adhesive-ceramic trilayers, ASME Journal of Electronic Packaging, 112(4) (1990).
35. E.K. Buratynski, Analysis of bending and shearing of tri-layer laminations for solder joint reliability, in E. Suhir, et al., Advances in Electronic Packaging 1997, ASME Press, 1997.
36. E. Suhir, Analysis of interfacial thermal stresses in a tri-material assembly, Journal of Applied Physics, 89(7) (2001).
37. S. Luryi and E. Suhir, A new approach to the high-quality epitaxial growth of lattice—mismatched materials, Applied Physics Letters, 49(3) (1986).
38. S.C. Lee et al., Strain-reliweved, dislocation-free $In_xGa_{1-x}As/GaAs(001)$ heterostructure by nanoscale-patterned growth, Applied Physics Letters, 85(18) (2004).
39. E. Suhir and T.M. Sullivan, Analysis of interfacial thermal stresses and adhesive strength of bi-annular cylinders, Int. Journal of Solids and Structures, 26(6) (1990).
40. A.O. Cifuentes, Elastoplastic analysis of bimaterial beams subjected to thermal loads, ASME Journal of Electronic Packaging, 113(4) (1991).

Solder Joints

41. H.S. Morgan, Thermal stresses in layered electrical assemblies bonded with solder, ASME Journal of Electronic Packaging, 113(4) (1991).
42. J.S. Hwang, Modern Solder Technology for Competitive Electronics Manufacturing, McGraw-Hill, New York, 1996.
43. R.J. Iannuzzelli, J.M. Pitarresi, and V. Prakash, Solder joint reliability prediction by the integrated matrix creep method, ASME Journal of Electronic Packaging, 118 (1996).
44. E. Suhir, Twist-off testing of solder joint interconnections, ASME Journal of Electronic Packaging, 111(3) (1989).
45. E. Suhir, Stress relief in solder joints due to the application of a flex circuit, ASME Journal of Electronic Packaging, 113(3) (1991).
46. E. Suhir, Flex circuit vs regular substrate: predicted reduction in the shearing stress in solder joints, Proc. of the 3-rd Int. Conf. on Flexible Circuits FLEXCON 96, San-Jose, CA, Oct. 1996.

47. F. Juskey and R. Carson, DCA on flex: A low cost/stress approach, in E. Suhir, et al., Eds., Advances in Electronic Packaging, ASME Press, 1997.
48. E. Suhir, Solder materials and joints in fiber optics: reliability requirements and predicted stresses, Proc. of the Int. Symp. on Design and Reliability of Solders and Solder Interconnections, Orlando, FL, Febr. 1997.
49. E. Suhir, Thermally induced stresses in an optical glass fiber soldered into a ferrule, IEEE/OSA Journal of Lightwave Technology, 12(10) (1994).

"Global" and "Local" Mismatch and Assemblies Bonded at the Ends

50. E. Suhir, "Global" and "local" thermal mismatch stresses in an elongated bi-material assembly bonded at the ends, in E. Suhir, Ed., Structural Analysis in Microelectronic and Fiber-Optic Systems, Symposium Proceedings, ASME Press, 1995.
51. E. Suhir, Thermal stress in a bi-material assembly adhesively bonded at the ends, Journal of Applied Physics, 89(1) (2001).
52. E. Suhir, Bi-material assembly adhesively bonded at the ends and fabrication method, U.S. Patent #6,460,753, 2002.
53. E. Suhir, Predicted thermal mismatch stresses in a cylindrical bi-material assembly adhesively bonded at the ends, ASME Journal of Applied Mechanics, 64(1) (1997).

Assembly with Low Modulus Adhesive Layer at the Ends

54. E. Suhir, Thermal stress in an adhesively bonded joint with a low modulus adhesive layer at the ends, Applied Physics Journal, (April) (2003).
55. E. Suhir, Electronic assembly having improved resistance to delamination, U.S. Patent #6,028,772, 2000.
56. E. Suhir, Thermal stress in a polymer coated optical glass fiber with a low modulus coating at the ends, Journal of Materials Research, 16(10) (2001).
57. E. Suhir, Coated optical fiber, U.S. Patent #6,647,195, 2003.

Thermally Matched Assembliess

58. E. Suhir, Axisymmetric elastic deformations of a finite circular cylinder with application to low temperature strains and stresses in solder joints, ASME Journal of Applied Mechanics, 56(2) (1989).
59. E. Suhir, Mechanical reliability of flip-chip interconnections in silicon-on-silicon multichip modules, IEEE Conference on Multichip Modules, IEEE, Santa Cruz, Calif., March 1993.
60. E. Suhir and B. Poborets, Solder glass attachment in cerdip/cerquad packages: thermally induced stresses and mechanical reliability, Proc. of the 40th Elect. Comp. and Techn. Conf., Las Vegas, Nevada, May 1990, see also: ASME Journal of Electronic Packaging, 112(2) (1990).
61. E. Suhir, Adhesively bonded assemblies with identical nondeformable adherends: predicted thermal stresses in the adhesive layer, Composite Interfaces, 6(2) (1999).
62. E. Suhir, Adhesively bonded assemblies with identical nondeformable adherends and inhomogeneous adhesive layer: predicted thermal stresses in the adhesive, Journal of Reinforced Plastics and Composites, 17(14) (1998).
63. E. Suhir, Adhesively bonded assemblies with identical nondeformable adherends and "piecewise continuous" adhesive layer: predicted thermal stresses and displacements in the adhesive, Int. Journal of Solids and Structures, 37 (2000).
64. E. Suhir, Applied Probability for Engineers and Scientists, McGraw Hill, New York, 1997.

Thin Films

65. K. Roll, Analysis of stress and strain distribution in thin films and substrates, Journal of Applied Physics, 47(7) (1976).
66. G.H. Olsen and M. Ettenberg, Calculated stresses in multilayered heteroepitaxial structures, Journal of Applied Physics, 48(6) (1977).
67. J. Vilms and D. Kerps, Simple stress formula for multilayered thin films on a thick substrate, Journal of Applied Physics, 53(3) (1982).
68. E. Suhir, An approximate analysis of stresses in multilayer elastic thin films, ASME Journal of Applied Mechanics, 55(3) (1988).

69. T.-Y. Pan and Y.-H. Pao, Deformation of multilayer stacked assemblies, ASME Journal of Electronic Packaging, 112(1) (1990).
70. E. Suhir, Approximate evaluation of the elastic thermal stresses in a thin film fabricated on a very thick circular substrate, ASME Journal of Electronic Packaging, 116(3) (1994).
71. E. Suhir, Predicted thermally induced stresses in, and the bow of, a circular substrate/thin-film structure, Journal of Applied Physics, 88(5) (2000).

Polymeric Materials and Plastic IC Packages

72. E. Suhir, Applications of an epoxy cap in a flip-chip package design, ASME Journal of Electronic Packaging, 111(1) (1989).
73. G.S. Ganssan and H. Berg, Model and analysis for reflow cracking phenomenon in SMT plastic packages, 43-rd IEEE ECTC., 1993.
74. E. Suhir, Failure criterion for moisture-sensitive plastic packages of integrated circuit (IC) devices: application of von-Karman equations with consideration of thermoelastic strains, Int. Journal of Solids and Structures, 34(12) (1997).
75. E. Suhir and Q.S.M. Ilyas, "Thick" plastic packages with "small" chips vs "thin" packages with "large" chips: how different is their propensity to moisture induced failures?, in E. Suhir, Ed., Structural Analysis in Micro-electronics and Fiber Optics, Symposium Proceedings, ASME Press, 1996.
76. M. Uschitsky and E. Suhir, Predicted thermally induced stresses in an epoxy molding compound at the chip corner, in E. Suhir, Ed., Structural Analysis in Microelectronics and Fiber Optics, Symposium Proceedings, ASME Press, 1996.
77. M. Ushitsky, E. Suhir, and G.W. Kammlott, Thermoelastic behavior of filled molding compounds: composite mechanics approach, ASME Journal of Electronic Packaging, 123(4) (2001).
78. D.K. Shin and J.J. Lee, A study on the mechanical behavior of epoxy molding compound and thermal stress analysis in plastic packaging, in E. Suhir, et al., Advances in Electronic Packaging 1997, Vol. 1, ASME Press, 1997.

Thermal Stress Induced Bowing and Bow-Free Assemblies

79. E. Suhir, Predicted bow of plastic packages of integrated circuit devices, in J.H. Lau, Ed., Thermal Stress and Strain in Microelectronic Packaging, Van Nostrand Reinhold, New York, 1993.
80. E. Suhir and J. Weld, Electronic package with reduced bending stress, U.S. Patent #5,627,407, 1997.
81. E. Suhir, Arrangement for reducing bending stress in an electronics package, U.S. Patent #6,180,241, 2001.
82. E. Suhir, Device and method of controlling the bowing of a soldered or adhesively bonded assembly, U.S. Patent #6,239,382, 2001.
83. E. Suhir, Bow free adhesively bonded assemblies: predicted stresses, Electrotechnik & Informationtechnik, 120(6) (2003).
84. E. Suhir, Bow-free assemblies: predicted stresses, Therminic'2004, Niece, France, Sept. 29–Oct. 1, 2004.

Probabilistic Approach

85. E. Suhir, Probabilistic approach to evaluate improvements in the reliability of chip-substrate (chip-card) assembly, IEEE CPMT Transactions, Part A, 20(1) (1997).
86. E. Suhir, Thermal stress modeling in microelectronics and photonics packaging, and the application of the probabilistic approach: review and extension, IMAPS International Journal of Microcircuits and Electronic Packaging, 23(2) (2000).

Optical Fibers and Other Photonic Structures

87. E. Suhir, Stresses in dual-coated optical fibers, ASME Journal of Applied Mechanics, 55(10) (1988).
88. E. Suhir, Fiber optic structural mechanics—brief review, editor's note, ASME Journal of Electronic Packaging, September 1998.
89. E. Suhir, Polymer coated optical glass fibers: review and extension, Proceedings of the POLYTRONIK'2003, Montreaux, October 21–24, 2003.
90. E. Suhir, Effect of initial curvature on low temperature microbending in optical fibers, IEEE/OSA Journal of Lightwave Technology, 6(8) (1988).

91. E. Suhir, Spring constant in the buckling of dual-coated optical fibers, IEEE/OSA Journal of Lightwave Technology, 6(7) (1988).
92. E. Suhir, Mechanical approach to the evaluation of the low temperature threshold of added transmission losses in single-coated optical fibers, IEEE/OSA Journal of Lightwave Technology, 8(6) (1990).
93. E. Suhir, Coated optical fiber interconnect subjected to the ends off-set and axial loading, International Workshop on Reliability of Polymeric Materials and Plastic Packages of IC Devices, Paris, Nov. 29–Dec. 2, 1998, ASME Press, 1998.
94. E. Suhir, Optical fiber interconnect with the ends offset and axial loading: what could be done to reduce the tensile stress in the fiber? Journal of Applied Physics, 88(7) (2000).
95. E. Suhir, Critical strain and postbuckling stress in polymer coated optical fiber interconnect: what could be gained by using thicker coating? International Workshop on Reliability of Polymeric Materials and Plastic Packages of IC Devices, Paris, Nov. 29–Dec. 2, 1998, ASME Press, 1998.
96. E. Suhir and J.J. Vuillamin, Jr., Effects of the CTE and Young's modulus lateral gradients on the bowing of an optical fiber: analytical and finite element modeling, Optical Engineering, 39(12) (2000).
97. E. Suhir, Predicted stresses and strains in fused biconical taper couplers subjected to tension, Applied Optics, 32(18) (1993).
98. E. Suhir, Apparatus and method for thermostatic compensation of temperature sensitive devices, U.S. Patent #6,337,932, 2002.
99. E. Suhir, Polymer coated optical glass fiber reliability: could nano-technology make a difference? Polytronic'04, Portland, OR, September 13–15, 2004.
100. D. Ingman and E. Suhir, Nanoparticle material for photonics applications, Patent pending, 2001.

2

Probabilistic Physical Design of Fiber-Optic Structures

Satish Radhakrishnan[a], Ganesh Subbarayan[a], and Luu Nguyen[b]

[a]*Purdue University, West Lafayette, IN 47907, USA*
[b]*National Semiconductor Corporation, Santa Clara, CA 95052, USA*

2.1. INTRODUCTION

Performance of a fiber-optic system depends on the coupling efficiency and the alignment retention capability. Fiber-optic systems experience performance degradation due to uncertainties in the alignment of the optical fibers with the laser beam. The laser devices are temperature sensitive, generate large heat fluxes, are prone to mechanical stresses induced and require stringent alignment tolerance due to their spot sizes. The performance of a photonic system is also affected by many other factors such as geometric tolerances, uncertainties in the properties of the materials, optical parameters such as Numerical Aperture etc. In this chapter, we apply systematic, formal procedures for designing the system in the presence of the above mentioned uncertainties. A low-cost generic fiber-optic package is used as a demonstration vehicle for the design procedures that are described in this chapter.

We begin the chapter by describing a representative fiber-optic system design (a low-cost VCSEL transreceiver), which is used as a vehicle in the development of the design and analysis techniques. An optical model describing the coupling efficiency between the laser and the fiber is next developed. We then develop a general mathematical formulation for maximizing the coupling efficiency and robustness of the system. The optimal geometry of the elements of the system that yield the maximum coupling efficiency and robustness is determined using the developed procedures.

Since the number of uncertain parameters that influence a photonic system can be large, an assessment of inter-relationship between the parameters with a view to identifying the critical parameters is essential. Towards this end we develop system representation and formal graph partitioning strategies to decompose the system into different subsystems. Through this process, we identify critical variables that have a larger, system-level influence. We also develop a simple to implement simulated annealing algorithm for carrying out the system partitioning. The results of system decomposition using graph partitioning and simulated annealing techniques are compared.

In general, the uncertainty in the performance of a photonic system arises mainly due to idealizations in geometry, material behavior, and loading history. Uncertainties in geometry can be predicted and controlled using tighter tolerances. However, the models currently used to describe material behavior are mostly deterministic. To predict the coupling efficiency of a photonic system to greater degree of confidence stochastic analysis procedures are necessary. As part of this analysis, the behavior of materials must be stochastically characterized. We present extensive experimental data on thermally and UV-cured type of epoxies typically used in photonic packages to enable stochastic analysis. The test data includes the viscoelastic behavior. We present an analytical model to obtain the stochastic variation in the displacement of the bonded VCSEL devices resulting from the stochastic viscoelastic behavior of the bond epoxies. We utilize the analytical model to predict the uncertainty in the optical coupling efficiency.

We finally describe efficient stochastic modeling techniques to determine the uncertainties in the alignment of the laser beams with the optical fibers. These include First Order Second Moment and Second Order Second Moment method. We then describe a stochastic design procedure where the uncertainty in behavior and geometry are considered during the design stage.

2.1.1. Demonstration Vehicle

The methodologies described in this chapter for the design of fiber-optic systems in the presence of uncertainty are demonstrated through a proposed photonic system (Figure 2.1). The system described in the example consists of a base, to which a 12×1 VCSEL array is bonded using epoxy. The curved reflector is positioned over the VCSEL leaving a small projection on the rear side to connect the anodes. The positioning of the reflector is such that the optical beam emerging out of the VCSEL is reflected into the optical fiber positioned horizontally. As the optical beam emerges out of the VCSEL, it diverges. The shape and size of the curved reflector should be such that the diverging optical beam is reflected into the optical fiber with the least coupling loss. The divergence angle of the

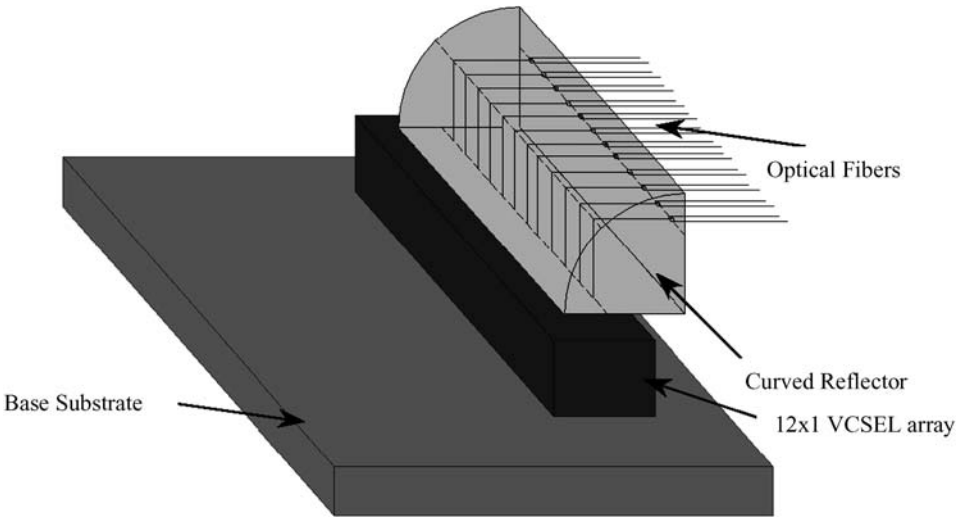


FIGURE 2.1. The fiber-optic system used as a demonstration vehicle in the study.

optical beam after reflection off the curved reflector should be such as to be less than the acceptance angle of the optical fiber to obtain the maximum possible coupling efficiency.

2.2. OPTICAL MODEL

There are many methods such as the Finite Element method, Beam Propagation method, Ray Tracing, etc. that have been applied thus far in the literature for optical modeling. Bierhoff et al. [1,2] describes that methods such as Finite Element Method (FEM), Beam Propagation Method (BPM), etc. can be used very efficiently in modeling single-mode interconnects, but these methods are not applicable for optical multi-mode waveguides guiding more than 1000 propagating modes. Due to the resulting numerical complexity in using FE-BPM, methods based on geometrical optics, called ray tracing, are more effective for modeling multi-mode waveguides. The ray tracing method approximates the output beam of the laser diode by a finite number (N) of rays, each determined by a starting point, a direction, and the time dependent optical power it carries. The propagation of the rays from the laser diode through various optical components as it enters the optical fiber is traced geometrically. This modeling approach leads to an optical model of a laser-diode providing N outputs and an input model of an optical fiber providing M inputs (Figure 2.2).

Kurzweg et al. [3] explains that the ray or geometric optics are the simplest of the optical modeling methods and have the smallest computation time when it comes to modeling and simulation. In the present chapter, the optical beam emerging from the VCSEL is approximated as being Gaussian in character. The laser-fiber coupling analysis of these Gaussian beams is performed using ray tracing analysis. Saleh and Teich, [4] explain that the Gaussian beams are quick to solve, since no explicit integration is necessary to calculate the resulting Gaussian beam at the interface of adjacent components. Ray tracing method has been widely applied in the literature [5–11] to model and to calculate the coupling efficiency in coupling the Gaussian beams with optical fibers and waveguides. The ray tracing method is used for optical modeling in this chapter.

The optical model is described below for the example problem. The equations for the coupling efficiency are symbolically solved in Mathematica [12] using an analytical de-

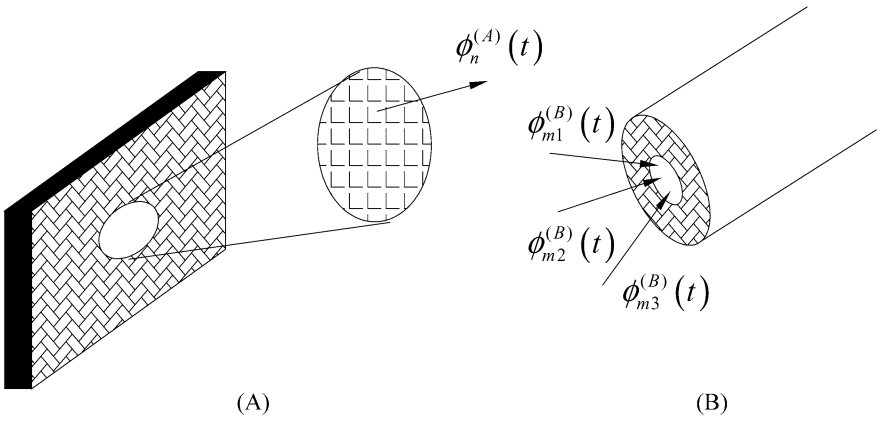


FIGURE 2.2. Equivalent modeling by ray tracing of (A) the output behavior of laser diodes and (B) the input behavior of the optical fibers.

scription of the rays emanating from the VCSEL. The assumptions made in developing the optical model are described below. The optical beam emerging from the VCSEL is assumed to be Gaussian. The curved reflector is assumed to be coated with a highly reflective material with negligible reflective losses. The numerical aperture of the optical fiber is taken as 0.2 for multi-mode operation and 0.14 for single-mode operation. The initial shape of the curved reflector is taken as cylindrical (only a quarter of the cylinder is taken as the curved reflector) with initial radius chosen as 300 microns. The center of the curved reflector is positioned 200 microns away from the center of the opening of the VCSEL. The optical fiber end is placed at the vertical axial plane of the curved reflector. The candidate material for the curved reflector is a cyclic olefin copolymer called Topas [13]. The refractive index of Topas used for the optical model is 1.53 [13].

A 12×1 array multi-mode VCSEL operating at 850 nm [14] and single-mode VCSEL operating at 1550 nm [15] was used. Calculations carried out in this study suggest that for single-mode operation, the wavelength of the laser beam has to be greater than 1462 nm, and hence the decision to use a VCSEL operating at 1550 nm. The opening diameter of the optical beams as they emerge out of the VCSEL was 16 microns for multi-mode operation and 8 microns for single-mode operations.

2.2.1. Mode Field Diameter

For a Gaussian beam, mode field diameter is defined as the diameter within which 85% of the power of an optical beam is contained when the light travels in free space. It is also defined as the diameter at which the electric and magnetic fields are reduced to $1/e$ of their maximum values, i.e., the diameter at which the power reduces to $1/e^2$ of the maximum power, because the power is proportional to square of the field strength. Since the optical beam has a Gaussian profile, 100% power would be contained in a beam only when the diameter is infinity. Thus the diameter within which 85% of the power is contained is taken for power calculations.

2.2.1.1. Divergence Angle The ray tracing method is based on geometric analysis and hence the rays propagating from the laser are taken as a set of n rays bounded by two rays tracing the mode-field diameter in two dimensions. The divergence angle corresponding to the mode-field diameter ($1/e^2$ diameter) specified by the VCSEL manufacturers [14,15] are used for ray tracing analysis.

2.2.1.2. Acceptance Angle of the Optical Fiber The optical beam entering the optical fiber propagates along the optical fiber due to total internal reflection. However, only portions of the optical beam directed at the optical fiber will propagate along its length. This is because, for a beam to undergo total internal reflection inside an optical fiber, the incident angle should be less than a critical value known as the acceptance angle. Thus, only the beams incident on the core of the optical fiber at an angle less than the acceptance angle would propagate.

For a fiber to act in single-mode, the numerical aperture (NA), which is a characteristic property of the optical fiber, should be as small as possible. The NA values for commercially available single-mode fibers are currently no lower than 0.14 [16]. Hence a NA value of 0.14 is used for calculating the acceptance angle for single-mode fiber. For multi-mode operation, NA value of 0.2 (a common choice in industry) is used.

The numerical aperture of the optical fiber is directly proportional to the sine of the acceptance angle. Thus, corresponding to the above NA values, the acceptance angle for multi-mode operation is 23° and for single-mode operation it is 16°.

2.2.1.3. V-Number The V-number or the V-parameter is the most important parameter which determines whether an optical fiber would act as a single-mode fiber or a multi-mode fiber. It also indicates the number of modes that propagate along a fiber when operated in multi-mode. For an optical fiber to transmit light in a single-mode, the V-number should be less than 2.405. The optical fiber would transmit multiple modes when the V-number exceeds 2.405. The number of modes N that propagate along the optical fiber is given by $N \approx V^2/2$.

V-number is defined as $\nu = \dfrac{2\pi a}{\lambda}(\text{NA})$.

Thus, for the V-number to be as low as possible the NA value and the core radius (a) of the optical fiber should be as low as possible and the wavelength of the optical beam should be large. The core diameter of the optical fiber is fixed at 8 μm for single-mode fibers and, as stated earlier, 0.14 is the lowest value of NA currently available for single-mode fibers. Thus, the value of the critical wavelength that enables single-mode operation is 1462 nm. Hence, a VCSEL operating at 1550 nm was chosen for the system and used in the calculations of optical behavior.

2.2.2. Refraction and Reflection Losses

The optical beam emerging from the VCSEL passes through the base of the curved reflector and, after reflection, emerges out through the side of the curved reflector. The refractive index of the curved reflector is 1.53 and thus the optical beam entering the curved reflector would undergo a change in the divergence angle and more importantly, it would cause refractive losses. The optical beam would then reflect off the curved section of the reflector. The curved reflector is assumed to have negligible losses on reflection. The reflected beams further undergo refractive losses as they exit out of the reflector. The angle with which the optical beam would exit the reflector would further change due to refraction before entering the optical fiber. A small air gap is assumed between the interfaces of the VCSEL and the curved reflector and between curved reflector and the optical fiber. The magnitude of the refractive loss was estimated at 8.77%.

2.2.2.1. Angle at Entry to the Optical Fiber As explained in the section above, the angle of the optical beam changes as it enters and exits the curved reflector. The divergence angle at the entry to the optical fiber for the initial design was determined using ray tracing to be 30° for both multi-mode and single-mode operations.

2.2.3. Calculations for Coupling Losses

The coupling power of a fiber-optic system can be defined as the ratio of power of the optical beam that enters the optical fiber (when coupled) to the output power of the VCSEL.

$$\text{Coupling Power} = \frac{\text{Power into Fiber}}{\text{Laser Power}}. \tag{2.1}$$

For a Gaussian beam, the power into the fiber is given by $P_0(1 - e^{-2(r/w)^2})$, where, r is the radius of the diverging optical beam corresponding to the acceptance angle of the optical

TABLE 2.1.
Estimated divergence angle and the resulting coupling efficiency at the initial design.

Wavelength	V-number	Divergence angle at VCSEL end	Divergence angle at the optic fiber end	Optical fiber diameter (μm)	Coupling loss (%)	Coupling efficiency (%)
For Multi mode operation (NA = 0.2), Acceptance angle = 23°						
$\lambda = 850$ nm	11.8	27° ($1/e^2$)	30°	50	10	75
For Single mode operation (NA = 0.14), Acceptance angle = 16°						
$\lambda = 1550$ nm	2.27	15° (FWHM)	30°	8	12	73

fiber and w is the mode field radius of the diverging optical beam when it enters the optical fiber. The laser power is taken as the power contained within the mode field diameter of the optical beam [$(P_0(1 - e^{-2})$] as it exits the VCSEL. The summary of the results for the optical model is given in Table 2.1.

2.2.4. *Coupling Efficiency*

For the initial design, the coupling efficiency values for both multi-mode and single-mode operation are low (Table 2.1). This is mainly due to the angle of the optical beam at the entry to the optical fiber being much larger than the acceptance angle. One can obtain the maximum possible coupling efficiency when this divergence angle is less than the acceptance angle of the optical fiber. The maximum possible coupling efficiency for both multi-mode and single-mode operations is 83%. The 17% loss is due to refraction as the optical beam traverses through the curved reflector.

2.2.4.1. Minimum Spot Size and Beam Shift The rays tracing the mode filed diameter of the Gaussian beams converge and diverge linearly. However the beams do not converge to a single point. The Gaussian ray profile changes from linear to parabolic shape as the beam converges to result in a minimum spot size (Figure 2.3). The minimum spot size is dependent on the wavelength of the optical beam and is given by $\lambda/\pi\theta$. For the single mode operation, this minimum spot size is 7 microns. The fiber-optic system experiences beam shift due to geometric and assembly tolerances as well due to the material behavior at high temperatures (Figure 2.4). The effect of the beam shift or the misalignment of the optical beam on the coupling efficiency can be seen in Figure 2.5. Here the coupling efficiency is taken as 100% when the beam is aligned without shift (i.e., the divergence angle of the optical beam is equal to the acceptance angle at the fiber). The coupling efficiency value does not change initially for small beam shift since the optical beam usually has a smaller spot size (W_0) as compared to the optical fiber (W) and slowly decreases with increasing beam shift. It can reduce to as low as 30% when the normalized beam shift with respect to the optical fiber is 0.5.

The maximum possible coupling efficiency can be obtained by the optimal design of the shape and size of the curved reflector. The optimized shape and size of the curved reflector should be such that the angle of the diverging beam at the entry to the optical fiber should be within the acceptance angle of the optical fiber. The procedures for achieving such designs are described later in this chapter.

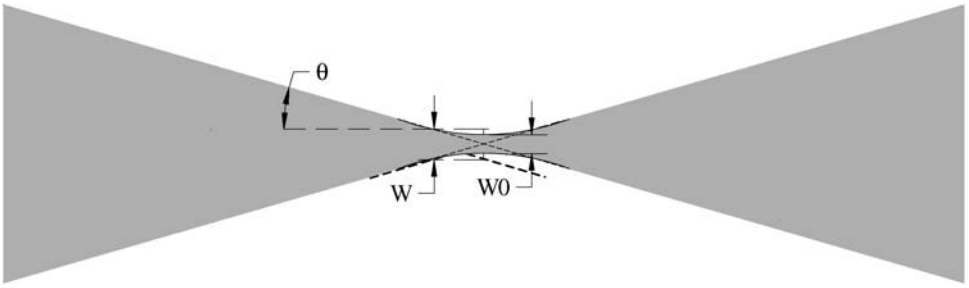


FIGURE 2.3. Gaussian Beams possess a minimum spot size, which is illustrated pictorially here.

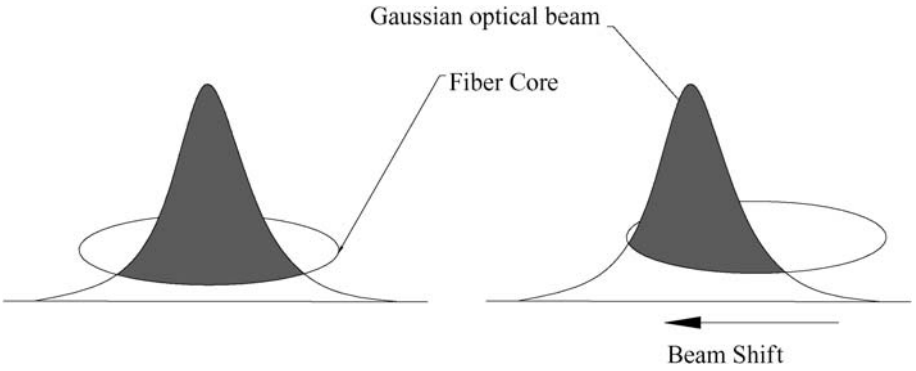


FIGURE 2.4. Shift of Gaussian Beams and the resulting coupling loss.

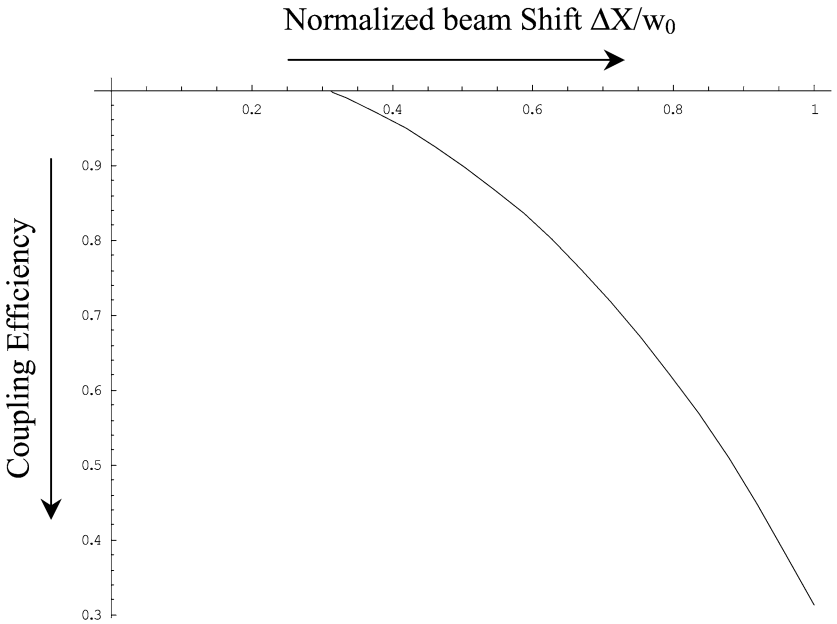


FIGURE 2.5. Coupling efficiency change due to normalized beam shift.

2.3. INTERACTIONS IN SYSTEM AND IDENTIFICATION OF CRITICAL VARIABLES

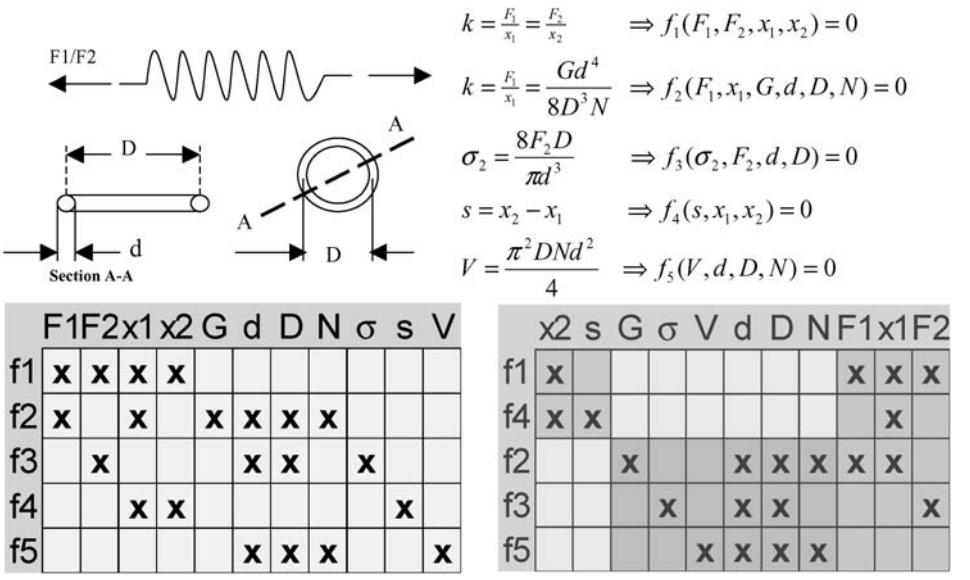
In general, stochastic analysis of any complex system requires an understanding of the inter-relationships between various elements that together make the complex system. The identification of critical variables to the system would reduce the complexity and the computational expense of the problem. Here, we use the coupling efficiency as the performance measure to evaluate the degradation caused by the misalignment of the laser beam with the optical fiber. To determine the uncertainty in the coupling efficiency of the system, the uncertainties in the variables that influence the coupling efficiency need to be determined first. However, uncertainties of all the variables that affect the coupling efficiency would not contribute equally to the overall uncertainty of the system. A systems analysis procedure based on the function–variable matrix is described here to identify the critical variables, and through their uncertainty determine the uncertainty in the coupling efficiency.

2.3.1. Function Variable Incidence Matrix

System decomposition procedures to identify critical system level parameters have been widely used in the literature. The first step of such a procedure is the formal representation of the functions and the variables that affect them. One such commonly used representation is a table referred to in the literature as the function dependence table or the function–variable incidence matrix [17,18]. The function–variable incidence matrix (referred in the rest of the chapter as the function–variable matrix) is a mechanism for the formal representation of the inter-relationships in the system. The function–variable matrix is partitioned into different subsystems using formal system decomposition techniques to identify the variables that are critical to the system. In other words, variables that strongly tie the functions are the system-level "linking" variables that belong to the whole system as opposed to any one sub-system.

Most systems in general can be characterized as being either hierarchical or non-hierarchical [17]. Starting from a function–variable matrix of the form shown in Figure 2.4, an ideal system would partition into completely independent subsystems with no interaction, but most commonly, engineering systems can only be partitioned into the block angular form shown in Figure 2.6. A system that can be decomposed into independent sub-systems is hierarchical in nature, while a system that partitions into sub-systems that are inter-dependent through a set of system-level "linking" variables is termed a non-hierarchical one. The key feature of non-hierarchical systems is the emergence of two kinds of variables: local variables, which come into play only when describing the functions of specific subsystems, and linking variables, which are shared among subsystems. The linking variables serve to describe the interactions among the function groups.

The systems-level analysis begins with the partitioning of the function–variable matrix into subsystems with local and linking variables. Optimal partitioning into two subsystems will yield the least number of linking variables. However, determining the optimal partition is an NP (Nondeterministic Polynomial time) complete problem [19]. Therefore, commonly, heuristic algorithms are used to carryout the partitioning. These partitioning algorithms and codes based on these algorithms often rely on a graph representation [20] of the system, and therefore there is a need to convert the function–variable matrix into an undirected graph.



	F1	F2	x1	x2	G	d	D	N	σ	s	V
f1	x	x	x	x							
f2	x		x		x	x	x	x			
f3		x				x	x		x		
f4			x	x						x	
f5						x	x	x			x

	x2	s	G	σ	V	d	D	N	F1	x1	F2
f1	x								x	x	x
f4	x	x								x	
f2			x			x	x	x	x	x	
f3				x		x	x				x
f5					x	x	x	x			

FIGURE 2.6. Function–variable matrix of a simple spring system and its partitioning into two subsystems is illustrated in the figure. Optimal partitioning leads to subsystems that are least interactive. The linking variables F_1, x_1 and F_2 are system-level variables since they determine the nature of the interaction between the subsystems.

System decomposition using graph partitioning techniques is a two step process. The function–variable matrix is first converted to a graph with vertexes and weighted edges. Michelena and Papalamabaros [19] suggested a procedure to convert the function–variable matrix to a hypergraph, which in turn is converted to an undirected graph. In Table 2.2, the function–variable matrix developed for the system described as test vehicle in this chapter is shown. The variables that lead to misalignment of the optical fiber with the laser beam can be characterized as being geometrical, optical, thermo-mechanical, and environmental in nature. The geometric and optical variables affecting the system were obtained from [21]. Not all the variables mentioned above affect the coupling efficiency directly. There are also intermediate functions that in turn affect the final coupling efficiency. These intermediate functions are acceptance angle, divergence angle at laser end, divergence angle at the optical fiber end, fiber offset and ray offset. The intermediate functions become variables for the description of higher level functions. There are thus a total of 7 functions and 23 variables.

The sensitivity values of the geometric variables were calculated based on the optical model described earlier (presented in greater detail in [21]). Some of the variables such as the environmental variables and thermo-mechanical variables do not have any analytical descriptions that enable the easy calculation of the sensitivity values. For these variables, (what was believed to be) a reasonable estimate of the sensitivity value was used.

2.3.2. Function Variable Incidence Matrix to Graph Conversion

Michelena and Papalambros [19] developed techniques to convert the function–variable matrix to graph assuming equal effect of each variable on respective functions and calculating the edge weights of the graph depending on the number of variables on which each function depends. Such an assumption of equal effect of variables on their respective

TABLE 2.2.
Function–variable matrix with sensitivity values for the photonic system used in the present study.

	Load on the epoxy between VCSEL and the base	Load on the epoxy between molded block and the base	Thickness of epoxy between VCSEL and base	Thickness of epoxy between molded block and base	Position of optic fiber in X, Y and Z direction	Position of curved reflector (X_0, Y_0)	Dimensions of curved reflector (a, b)	Position of VCSEL with respect to the base	Numerical aperture of fiber	Wavelength of laser beam	Initial diameter of the laser beam	Core diameter of optical fiber	Young's modulus of molded block	Viscoelasticity property of the epoxy between molded block and the base	Viscoelasticity property of the epoxy between the VCSEL and the base	Thermal conductivity of molded block	CTE value of the molded block	Acceptance angle	Divergence angle at laser end	Divergence angle at optical fiber end	Fiber offset	Ray offset	Viscoelasticity behavior
Acceptance angle									0.507									1					
Divergence angle at laser end										1	1								1	1			
Divergence angle at optical fiber end	0.5	0.5	0.5	0.5		0.5	0.5	0.5		1	1									1			
Fiber offset		1		1	1							0.5	0.2	1		0.5	0.5				1		
Ray offset	1		1			0.5	0.5	1							1							1	
Viscoelasticity behavior	0.4	0.4	0.2	0.2		1							0.5	1	1	0.5	0.5				1	1	1
Coupling efficiency					0.0155	0.015	0.02	0.015	0.02	0.02	0.02	0.02						0.04	0.0204	0.04	0.02	0.0204	
	Environmental effects		Geometrical tolerance						Optical parameters									Thermo-mechanical effects					

functions is not always valid, especially when the degree of dependence of the function on different variables varies substantially. The method to calculate the edge weights of the graph is modified in this chapter to include the sensitivity values of the function with respect to each of the variables.

The conversion of the function–variable matrix to graph first consists of representing the function–variable matrix in terms of a hypergraph as shown in Figure 2.7. A hypergraph may be visualized as consisting of solid linkages each representing one variable. The vertexes of the linkage represent the functions that the variables affect and the edges are the hyperedges [19]. The number of edges in a linkage is denoted as p. Assuming equal effect of all variables on a function, one simple estimate of the weight of a hyperedge is $w = 1/(p-1)$. Since the minimum number of edges needed to be cut to partition the vertex set of a p-hyperedge is $(p-1)$, the total weight of the cut edges will be one [19] (Figure 2.7).

The sensitivity values of the variables with respect to the functions can be used to accurately calculate the edge weight. The modified procedure for calculating the edge weight of the graph for the example problem is shown in Figure 2.8. The edges of the hypergraph and graph are exactly as shown in Figure 2.7. However, the weights of hyperedges are calculated differently. The weight of each hyperedge of a variable is different and depends only on the sensitivity value of the functions (vertexes) that the hyperedge connects. The hyperedge weight is calculated as the product of the sensitivity values of the functions connected by the hyperedge with respect to the variable. This can be written mathematically as

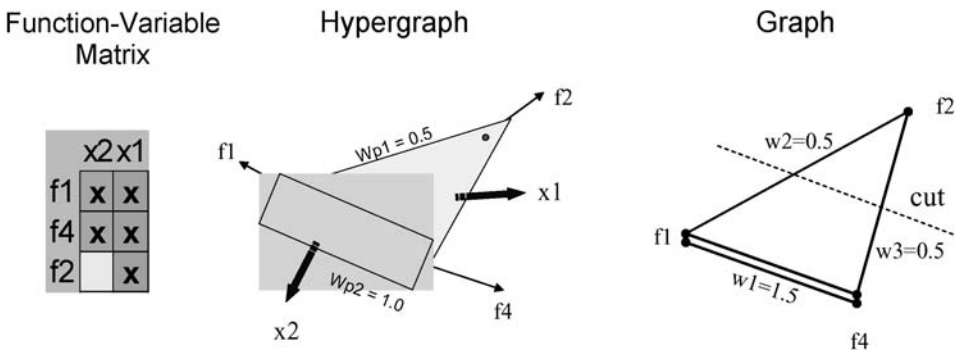


FIGURE 2.7. A schematic illustration of the conversion from function–variable matrix to hypergraph to graph for the application of graph partitioning algorithms [19].

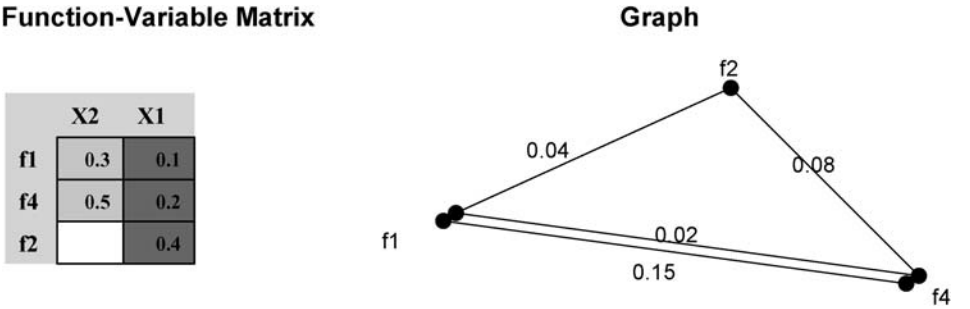


FIGURE 2.8. A schematic illustration of the calculation of edge weights based on sensitivity values.

$$w(i,j)_{i\neq j} = \sum_{k=1}^{n} f(i,k)\cdot f(j,k), \quad i,j = 1,2,\ldots m, \tag{2.2}$$

where, m = number of functions, n = number of variables. $w(i,j)$ represents the weight of graph edge connecting functions i and j while $f(i,j)$ represents the element of the function–variable matrix. The value of $f(i,j)$ is zero when a variable j does not affect the function i thus resulting in zero edge weight.

The algorithm to convert the function–variable matrix to graph was implemented in Java based on the work in reference [19] using the modified procedure to include sensitivities in calculating the edge weight. The output of the Java code is inline with the input file required by the Chaco program [10] to partition the graph. The graph partitioning algorithms are discussed next.

2.3.3. Graph Partitioning Techniques

The graph partitioning as a means for system decomposition is widely applied in a variety of fields such as power networks [22,23], finite element analysis [24], VLSI design [25], and materials development [26] among others. The algorithms for graph partitioning are well established and largely heuristic in nature owing to the NP completeness of the partitioning problem. The popular Kernighan-Lin algorithm [20] as well as its variant by Fiduccia and Mattheyses [27] are implemented in codes such as Chaco developed by Hendrickson and Leland [28].

The Kernighan-Lin algorithm though popular and widely applied has two major drawbacks. Firstly it needs the system to be represented in the form of a graph. This graph is then partitioned to obtain the optimal system decomposition. Secondly Kernighan-Lin algorithm does not result in optimal partition when the ratio of number of edges to the number of vertexes is low. The algorithm performs poorly when this ratio is less than three and produces nearly optimal solutions when the ratio is higher than five [29].

The graph partitioning algorithms (for example those described in reference [20] and the ones used in this study) enable one to partition (into sub-graphs) any combinatorial problem expressed as a graph. Their most common use has thus far been to divide a finite element domain into sub-domains for efficient parallel computation. Here, such an approach is used for assigning system-level design and material parameters into appropriate sub-systems for the test vehicle described earlier in an automated manner. The common heuristic algorithms such as the Kernighan-Lin algorithm are available in the Chaco program [28] used in the present study. In addition to partitioning a system into two subsystems, simultaneous k-partitioning is possible in Chaco when the number of partitions, k, is specified by the user.

2.3.4. System Decomposition using Simulated Annealing

Another method to partition a graph is to use a globally optimal search strategy such as Simulated Annealing. Graph partitioning using simulated annealing is demonstrated in the literature in references [30–32], but such a partitioning procedure is not common. System decomposition using simulated annealing is a two step process where in the system must be represented in the form of a network or graph and the graph is then partitioned optimally.

Simulated annealing is a generalization procedure similar to Monte Carlo simulation for examining the equation of states and frozen states of n-body systems [33]. The algorithm is inspired by the manner in which liquid freezes or metals recrystallize during the process of annealing. In an annealing process a melt, initially at high temperature and in disordered state, is slowly cooled so that the system at any time is approximately in thermal equilibrium. As cooling proceeds, the system becomes more ordered and approaches a stable ground state at $T = 0$. Hence, the process can be thought as an adiabatic path to the lowest energy state. If the initial temperature of the system is too low or cooling is done fast, the system may become quenched forming defects or freezing out in a metastable state (i.e., trapped in local minimum energy state).

In the original Metropolis scheme [33], an initial state of a thermodynamic system was chosen at energy E and temperature T. Holding T constant, the initial configuration is perturbed and the change in energy ΔE is computed. If the change in energy is negative the new configuration is accepted. If the change in energy is positive it is accepted with a probability given by the Boltzmann factor $\exp(E/kT)$. This process is then repeated sufficient number of times to give good sampling statistics for the current temperature, and then the temperature is decremented and the entire process repeated until a frozen state is achieved at $T = 0$. The flow of control during solution using simulated annealing is described in the flowchart of Figure 2.9.

By analogy, the generalization of this Monte Carlo approach to combinatorial optimization problem is straightforward [32–35]. The current state of the thermodynamic system is analogous to the current solution to the optimization problem, the energy equation for the thermodynamic system is analogous to the objective function, and the ground state is analogous to the global minimum. The major difficulty in implementing the algorithm is the lack of analogy for the temperature T. Whether the final minima obtained is global or local depends on the "annealing schedule," the choice of initial temperature, how many iterations are performed at each temperature, and the magnitude of the temperature decrement as the cooling proceeds.

Many different schemes have been proposed for the length of the Markov chain and for updating the temperature. Aarts [36] suggests that for discrete valued design variables, every possible combination of design variables in the neighborhood of a steady state design variable should be visited at least once with a probability of P. Thus, if there are S neighboring designs then the length of Markov chain is given by

$$M = S \ln\left(\frac{1}{1-P}\right),$$

where $P = 0.99$ for $S > 100$, and $P = 0.995$ for $S < 100$. For discrete valued variables, there are many options for defining the neighborhood of the design. One possibility is to define it as all the designs that can be obtained by changing one design variable to its next higher or lower value [37]. For an n variable design problem, the immediate neighborhood has $S = 3n - 1$ points.

Many different schemes have been proposed for updating the temperature. A frequently used rule is a constant cooling update [36,37].

$$T_{k+1} = \alpha T_k, \quad k = 0, 1, 2, \ldots . \tag{2.3}$$

where, $0.5 \leq \alpha \leq 0.95$.

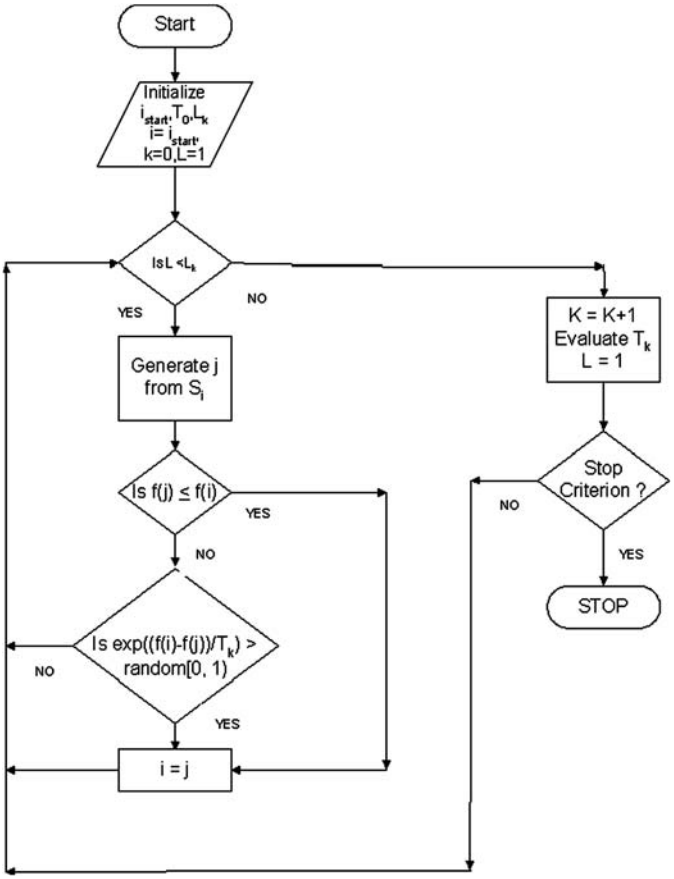


FIGURE 2.9. A flow chart describing the simulated annealing procedure.

In addition to using heuristic graph-partitioning strategies, in the present study, system decomposition is demonstrated by implementing the simulated annealing algorithm. The main reason for using simulated annealing is to reduce the complexity of the partitioning process by converting the two-part operation (conversion of function–variable matrix to graph followed by partitioning of the graph) into one operation. Though the computational time for simulated annealing is large due to the number of iterations required to achieve the convergence criteria, the time required for partitioning does not increase significantly as the number of subsystems increases. Also, as the system size increases, the time taken for partitioning does not increase significantly. Thus for a large system, the time taken for partitioning a system with simulated annealing will compare well with that obtained using the Chaco program. In the present work, the simulated annealing algorithm was implemented using Mathematica [12].

The objective function for this problem can be stated as

$$\min \sum_{j=1}^{n} \sum_{i=1}^{m} L(i, j), \tag{2.4}$$

where m = number of functions, n = number of linking variables. $L(i, j) = n \times m$ is the matrix containing the sensitivity values of all the functions with respect to the linking variables. The goal of the partitioning problem is to rearrange the rows and columns such that the number of linking variables is a minimum. The double summation in the objective function above is defined as the cut-size of the partition. Thus, the cutsize resulting from the partition must be minimized to obtain the optimal partition of the system.

For the photonic system considered in this chapter, the function–variable matrix was first converted to a graph and the edge weights were calculated based on the methodology described earlier. The Chaco program [20] was then used to partition the graph and thereby partition the system into subsystems. A similar partitioning was carried out using the simulated annealing algorithm. The system decomposition results obtained using the simulated annealing method matched exactly with that obtained using the Chaco program.

The result of partitioning the system into two subsystems is shown in Table 2.3. In the table, the seven functions were split into two groups of three and four functions each. We further partitioned the second subsystem into two lower-level sub-systems. The two lower-level sub systems consisted of two functions each. For the first lower-level sub system, there were no local variables. The second partitioning resulted in five linking variables for two lower-level sub systems. The final output of the graph partitioning analysis is shown in Table 2.4. Thus, after partitioning the system into three sub systems, we find that the system is naturally decomposed into largely domain-specific subsystems consisting of optical parameters, material parameters and loading and a few geometric parameters. The partitioning resulted in 15 linking variables for the whole system.

The linking variables were mainly geometric variables: position of optical fiber in the x, y and z direction; the position of the curved reflector in the x and y direction; the dimensions of the curved reflector; the position of the VCSEL with respect to the base; the initial diameter of the laser beam; the core diameter of the optical fiber; the divergence angle at the optical fiber end; the fiber offset and the ray offset; the thickness of the epoxy between base VCSEL and the base; the thickness of the epoxy between the molded block and the base. The only optical parameter among the linking variables was the wavelength of the laser beam, while the load/environmental parameters were the load on the epoxy between VCSEL and the base and the load on epoxy between the molded block and the base. Finally, the parameter defining the viscoelastic behavior of the epoxy between the VCSEL and the base was also a linking variable. These are the most critical variables that describe the interaction between the subsystems identified by the partitioning process. In the following sections, we evaluate the effect of uncertainty in the above identified linking variables on the coupling efficiency of the system.

2.4. DETERMINISTIC DESIGN PROCEDURES

The coupling efficiency for the initial design was determined to be very low as compared to its maximum achievable value (Table 2.1). As mentioned earlier, the coupling efficiency can be maximized through the optimal design of the shape and size of the curved reflector. A system fabricated per the specified design, however, is not guaranteed to achieve the optimal coupling efficiency due to the uncertainties caused by the fabrication/assembly processes as well as the material behavior under the use environment. Ideally, the fiber-optic system should be robust against the uncertainties in the variables. The following sections describe the design procedures for the optimal and robust design of the fiber-optic system.

TABLE 2.3.
The function–variable matrix for the photonic system obtained after the first partition.

	Numerical aperture of fiber	Acceptance angle	Divergence angle at laser end	Load on the epoxy between VCSEL and the base	Load on the epoxy between molded block and the base	Thickness of epoxy between VCSEL and base	Thickness of epoxy between molded block and base	Young's modulus of molded block	Viscoelasticity property between molded block and base	Viscoelasticity property of the epoxy between VCSEL and the base	Thermal conductivity of molded block	CTE value of the molded block	Viscoelasticity behavior	Position of optic fiber in X, Y and Z direction	Position of curved reflector (X_0, Y_0)	Dimensions of curved reflector (a, b)	Position of VCSEL with respect to the base	Initial diameter of the laser beam	Core diameter of optical fiber	Divergence angle at optic fiber end	Ray offset	Fiber offset	Wavelength of laser beam
Acceptance angle	✓	✓																					✓
Divergence angle at laser end			✓																				✓
Coupling efficiency	✓	✓	✓			✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Divergence angle at optical fiber end				✓	✓	✓	✓	✓	✓		✓	✓		✓	✓	✓	✓	✓		✓			
Fiber offset					✓	✓	✓	✓	✓		✓	✓		✓	✓	✓			✓		✓		
Ray offset				✓	✓	✓		✓	✓	✓	✓	✓		✓	✓	✓	✓				✓		
Viscoelasticity behavior				✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

TABLE 2.4.
The function–variable matrix for the photonic system obtained after the second partition.

	Numerical aperture of fiber	Acceptance angle	Divergence angle at laser end	Young's modulus of molded block	Viscoelasticity property of the epoxy between molded block and the base	Thermal conductivity of molded block	CTE value of the molded block	Viscoelasticity behavior	Load on the epoxy between VCSEL and the base	Load on the epoxy between molded block and the base	Thickness of epoxy between VCSEL and base	Thickness of epoxy between molded block and base	Viscoelasticity property of the epoxy between the VCSEL and the base	Position of optic fiber in X, Y and Z direction	Position of curved reflector (X_0, Y_0)	Dimensions of curved reflector (a, b)	Position of VCSEL with respect to the base	Initial diameter of the laser beam	Core diameter of optical fiber	Divergence angle at optic fiber end	Ray offset	Fiber offset	Wavelength of laser beam
Acceptance angle	✓																						
Divergence angle at laser end		✓	✓															✓	✓	✓			✓
Coupling efficiency	✓	✓	✓																✓	✓		✓	✓
Divergence angle at optical fiber end									✓	✓	✓							✓		✓			✓
Ray offset									✓	✓			✓	✓	✓	✓	✓				✓		
Fiber offset				✓	✓	✓	✓			✓	✓	✓		✓				✓				✓	
Viscoelasticity behavior				✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓						✓	✓	
	Optical parameters			Material parameters					Geometry and loading parameters					Geometry tolerances									

The procedures are deterministic since the design problem is formulated through deterministic treatment of functions and variables. In other words, while the design procedure is deterministic, the expectation is that the resulting designs are robust to uncertainties in the variables.

2.4.1. Optimal and Robust Design

In general, to maximize the coupling efficiency of the fiber-optic system, the laser beam shape and size entering the optical fiber needs to be optimized. The objective for the optimization procedure is to ensure that the i) diameter of the beam at the entry to the optical fiber is less than the diameter of the core of the optical fiber, and, ii) divergence angle at the entry to the optical fiber is less than the acceptance angle of the optical fiber.

The specific form of objective function for both single and multi-mode operation chosen in the present study is:

$$\max f \equiv \left(1 - e^{-2(r/w)^2}\right), \tag{2.5}$$

where r is the radius of the diverging optical beam desired to be less than the acceptance angle of the optical fiber and w is the mode field radius of the optical beam when it enters the fiber. The radius and the divergence angle of the optical beam are calculated using the parameters y_1, y_2, θ_1 and θ_2 as shown in Figure 2.10.

Due to the inherent uncertainties in the design variables, in practice, the function value will vary in the neighborhood of the optimum design. Thus, the optimal design may not necessarily be the most robust one or, the optimal design need not necessarily be least sensitive to perturbations in the design. Thus, we also develop in the present study a mathematical formulation to identify the robust design that will minimize the performance variability while achieving (to the extent possible) the maximum coupling efficiency.

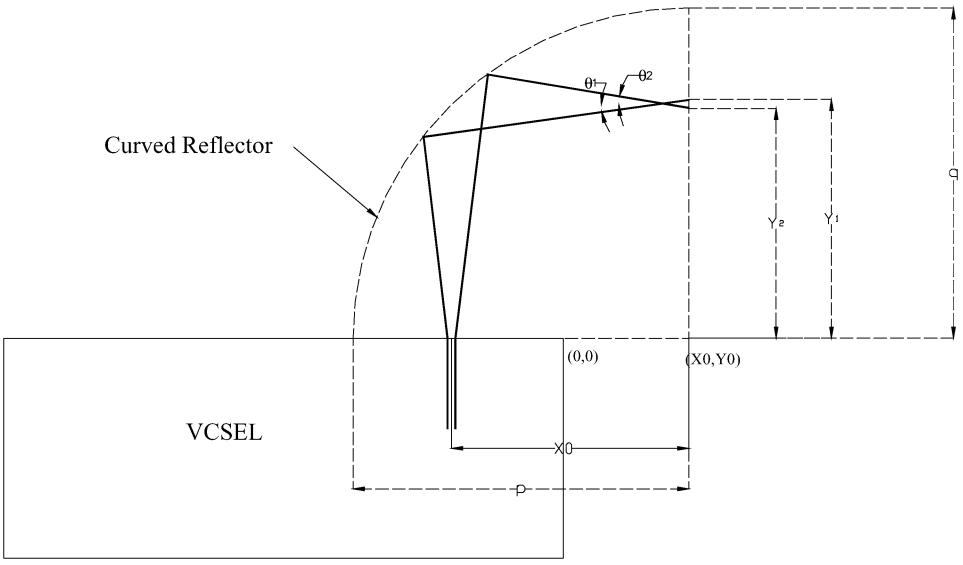


FIGURE 2.10. Parameters used in the optimal design of the reflector.

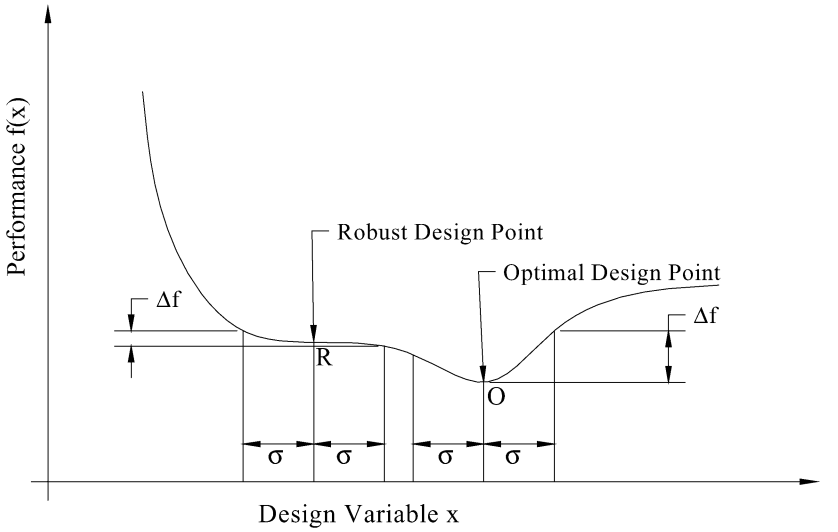


FIGURE 2.11. An illustration of optimal and robust design points.

The difference between the optimal design point and the robust design point can be seen in Figure 2.11. Points R and O are the robust and optimal design points, respectively. For a perturbation of the order of the standard deviation σ in the design variable, it can be seen that the change in the function value Δf for an optimal design point could be high, while for a robust design point, the change in function value Δf is minimal. The extension of these ideas to multiple dimensions is straightforward.

The methodologies developed thus far in the literature [38–43] for robust design are suitable only for smooth functions since these are based on determining sensitivities. Therefore, the methodologies cannot handle steep changes in function values. In this study, we formulate a more general robust optimal design problem using a min-max optimization formulation. The objective function for robust design is as follows:

$$\min_{x} \max_{\Delta x} \left[f - \frac{1}{\Delta x} \int_{x}^{x+\Delta x} f\,dx \right]^2. \tag{2.6}$$

In the above expression, the integral represents the average value of the function over the interval Δx at x. The squared difference between the function value and the average value is first maximized for a particular value of Δx to identify the variation of the function at any x. This function is then minimized with respect to the variable x.

The integral in Equation (2.6) was evaluated using a numerical quadrature due to the complexity of the function. The built-in numerical integration function in Mathematica, while accurate, required a great deal of computational time. Therefore, a simpler trapezoidal rule was used for integration. Since the trapezoidal rule involves the evaluation of the function values at specific intervals, the computation time was reduced by an order of magnitude. The final form of the objective function for solving Equation (2.6) was as follows:

$$\min_{x} \max_{\Delta x} \left[f - \frac{h}{2\Delta x} \left(f_0 + f_n + 2 \sum_{i=1}^{n-1} f_i \right) \right]^2, \quad n = \left(\frac{\Delta x}{h} \right). \tag{2.7}$$

In general, the robust design determined by solving Equation (2.7) could result in a coupling efficiency that is lower than the optimal value. The robust design objective function can be modified to include the optimal design function to yield a robust design with an acceptable coupling efficiency value.

$$\max_x \left\{ c_1 f - c_2 \max_{\Delta x} \left[f - \frac{h}{2\Delta x} \left(f_0 + f_n + 2 \sum_{i=1}^{n-1} f_i \right) \right]^2 \right\}, \quad c_1 + c_2 = 1. \tag{2.8}$$

Through the parameters c_1 and c_2, one can effect a trade-off between the optimal and robust designs. Clearly, when c_2 is unity, the intent is to obtain a robust design and when the parameter is set to zero, the intent is to obtain an optimal design. For other combinations of the parameters c_1 and c_2, one obtains designs with characteristics in between those of optimal and robust designs.

2.4.2. A Brief Review of Multi-Objective Optimization

The design formulation in Equation (2.8) is a constrained multi-objective optimization problem. The general form of bound constrained multi-objective function is:

$$\text{Min: } c_1 f_1(x) + c_2 f_2(x) + \cdots \tag{2.9}$$

$$\text{Subject to: } \underline{x_l} \le \underline{x} \le \underline{x_u} \quad 0 \le c_1, c_2, c_3, \ldots \le 1.$$

The concept of constrained multi-objective optimization can be explained using Figure 2.12. Let f_1 and f_2 be two objective functions dependent on variables x_1 and x_2. Let x_{m_1} and x_{m_2} be the points that minimize f_1 and f_2, respectively. The constant function value point sets around x_{m_1} and x_{m_2} are also illustrated in the figure. The two curves intersecting the constant value function curves represent the constraints. The curve connecting x_{m_1} and x_{m_2} represents the set of points that minimize $f = c_1 f_1(x) + c_2 f_2(x)$ for

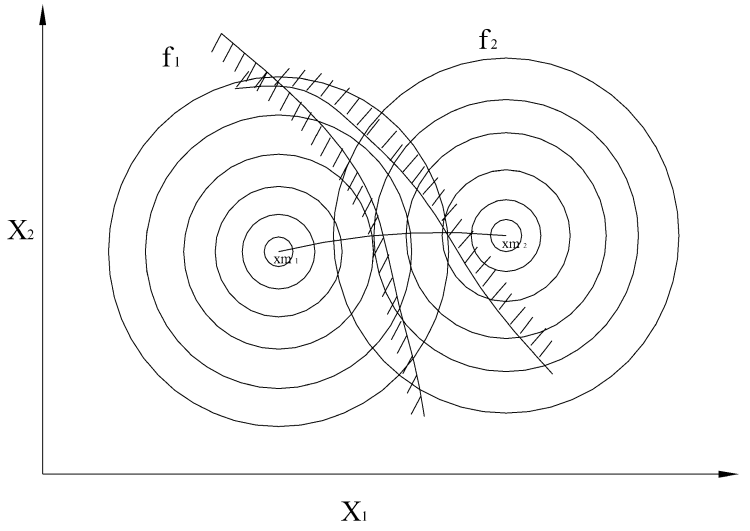


FIGURE 2.12. A schematic illustration of multi-objective optimization.

various combinations of the parameters c_1 and c_2 in the range of [0, 1]. In general, the optimum in a multi-objective problem is an $(n-1)$ dimensional hypersurface of "non-dominated" points from which it is not possible to simultaneously decrease all the function values. For an unconstrained problem, any point in the curve connecting x_{m_1} and x_{m_2} would contribute to a solution set. However, the constraints reduce the solution set size limiting the possible acceptable combinations of c_1 and c_2. Thus for a multi-objective optimization problem, the constraints could result in a design with one or more objective function given greater importance as compared to the other objective functions.

2.4.3. Implementation

Traditional gradient-based search methods such as the quasi-Newton method implemented in the "*find minimum*" function of Mathematica [12] were first used to determine the optimum solution. The "*find minimum*" function in Mathematica, since it uses a quasi-Newton algorithm, is only assured of converging to a local minimum. Even this convergence depends on the satisfaction of "guaranteed descent" conditions such as the Goldstein-Armijo condition [44]. Different starting values for "*find minimum*" yield different solutions indicating multiple local minima for the multi-objective function. For problems with a large number of variables and with potentially large number of local minima, there is no assurance that the global minimum will be identified even if many runs, each with a different starting point, are carried out. Common algorithms that are capable of identifying the global minima use search techniques that randomly sample the design space. Simulated annealing and Genetic algorithms belong to this class of random search algorithms [45]. This procedure of random search, though very effective in identifying the global minimum, becomes computationally expensive for optimization problems with a large number of variables, especially when the cost of analysis is high. In the present study, owing to the relatively low cost of analysis, we chose simulated annealing [46–49] as the optimization algorithm. We used the "*Nminimize*" function implemented within Mathematica to use the simulated annealing optimization algorithm to solve Equation (2.8).

The inner maximization function in Equation (2.8) was solved using the "*Nminimize*" function of Mathematica. However, the output of the inner maximization function could not be used as an input for the outer maximization function within Mathematica. Thus, a simulated annealing code was written to solve the outer maximization function. The computational time to solve Equation (2.8) was about 40 hours.

2.4.4. Results

The procedures described in the previous subsections were applied to the test vehicle. The cross-section of the curved reflector was taken as an ellipse for the purposes of the design. The objective functions in Equation (2.8) was solved using the simulated annealing algorithm to determine the global optimum values of major radius (p), minor radius (q) and the position of the VCSEL with respect to the axis of the curved reflector (x_0) for the respective design objectives. The lower and upper limits on both the major and minor radii of the curved reflector for both single-mode and multi-mode operation were 250 microns and 600 microns respectively. The lower limit of 250 microns was chosen to ensure that it was larger than the outer diameter of the cladding of the optical fibers which is 125 microns. The upper limit of 600 microns was chosen to limit the overall size of the package.

The results for optimal and robust designs for both multi-mode and single-mode operation are tabulated in Tables 2.5 and 2.6 respectively. The values in the last columns in

TABLE 2.5.
Deterministic optimal and robust designs for multi-mode operation.

Method	p (μm)	q (μm)	x_0 (μm)	Coupling efficiency %
Initial design	300	300	200	90
Deterministic optimal design	600	308	530	100
Deterministic robust design	598	331	520	100

TABLE 2.6.
Deterministic optimal and robust designs for single-mode operation.

Method	p (μm)	q (μm)	x_0 (μm)	Coupling efficiency %
Initial design	300	300	200	88
Deterministic optimal design	600	270	545	100
Deterministic robust design	600	275	543	100

Tables 2.5 and 2.6 are the coupling efficiency values for the respective methods. It can be seen that both the methods result in an improved coupling efficiency value with respect to the initial design. Both the designs yielded 100% coupling efficiency for both multi- and single-mode operation.

2.5. STOCHASTIC ANALYSIS

In this section, we develop techniques for the stochastic analysis of the photonic systems. Traditionally, uncertainty analysis has been based on the sampling methods. These methods involve evaluating functional relationships at a set of sample points, and thereby establishing the output uncertainty through exhaustive evaluation of input uncertainties. One such method is the Monte Carlo Simulation. Monte Carlo method involves random sampling from the distribution of inputs and successive model runs until a statistically significant distribution of outputs is obtained. However, its use is limited when complex engineering analyses are required to predict the output function. Added to that, the probability distribution functions may not be well defined for the input variables. The numerous simulations required by Monte Carlo techniques can lead to very high computational time. This high computational time may not be acceptable due to cost and time constraints. Thus the Monte Carlo method, though very accurate, is not suitable for complex physical systems. Thus, the approximation techniques based on the first and second order analysis of numerical models are more suitable due to lower computational cost. These approximate, but computationally efficient methods for stochastic analysis are described in this section.

2.5.1. The First and Second Order Second Moment Methods

The assumption underlying the First and Second Order Second Moment methods is that the important information about the random variables (or functions) of interest can be summarized with the mean representing the expected value of the variable (or function), and the variance representing the second moment about the mean. The First and Second Order Second Moment methods are based on Taylor Series expansion around either mean

or critical values of one or more variables. The second moment methods can be used for systems with second moment inputs and parameters with relatively small variance. The third and higher moments are usually ignored since they are relatively small compared to the second moment.

The First Order analysis is the analysis of the mean and the variance of a random function based on its first order Taylor Series expansion. Second order analysis is the analysis of the mean and the variance based on the functions second order Taylor Series expansion.

Let $\mathbf{f}(\mathbf{x})$, be a vector function of random variable $\mathbf{x}$. The Taylor Series expansion of the function about say the mean $\hat{\mathbf{x}}$ is

$$f(\mathbf{x}) = f(\hat{\mathbf{x}}) + \nabla f^T(\hat{\mathbf{x}})(\mathbf{x} - \hat{\mathbf{x}}) + \frac{1}{2}(\mathbf{x} - \hat{\mathbf{x}})^T \mathbf{H}(\mathbf{x} - \hat{\mathbf{x}}) + \cdots, \tag{2.10}$$

where, $\mathbf{H}$ is the Hessian matrix containing the second partial derivatives of the function. The first order Taylor Series is then,

$$f(\mathbf{x}) = f(\hat{\mathbf{x}}) + \nabla f^T(\hat{\mathbf{x}})(\mathbf{x} - \hat{\mathbf{x}}). \tag{2.11}$$

Taking the expected value of the above function, the mean of f is estimated to first order by:

$$\begin{aligned} \hat{f} &= E[f(\mathbf{x})] \\ &= E[f(\bar{\mathbf{x}}) + \nabla f^T(\bar{\mathbf{x}})(\mathbf{x} - \bar{\mathbf{x}})] \quad \mathbf{x} \in R^n \\ &= f(\bar{\mathbf{x}}). \end{aligned} \tag{2.12}$$

In the above expression, both $f(\bar{\mathbf{x}})$ and $\nabla f(\bar{\mathbf{x}})$ are evaluated at $\bar{\mathbf{x}}$ and therefore are known, deterministic quantities. The first order estimate of the mean is exactly the value obtained through the application of traditional deterministic approach.

Thus, using the expected value, the variance of f may be approximated to first order by:

$$\begin{aligned} \boldsymbol{var}[f(\mathbf{x})] &= E[(f - \hat{f})^2] \\ &= \nabla f^T(\bar{\mathbf{x}}) E[(\mathbf{x} - \bar{\mathbf{x}})(\mathbf{x} - \bar{\mathbf{x}})^T] \nabla f(\bar{\mathbf{x}}) \\ &= \nabla f^T(\bar{\mathbf{x}}) \boldsymbol{cov}(\mathbf{x}) \nabla f(\bar{\mathbf{x}}), \end{aligned} \tag{2.13}$$

where, $\boldsymbol{cov}(\mathbf{x})$ is the covariance matrix of $\mathbf{x}$. The variance of f is a function of the uncertainty or variability of $\mathbf{x}$, and the sensitivity of f to $\mathbf{x}$ in the neighborhood of $\bar{\mathbf{x}}$.

In a similar manner, using the second order Taylor series expansion of the function, the estimate of the mean can be calculated as:

$$\begin{aligned} \hat{f} &= E\left[f(\bar{\mathbf{x}}) + \nabla f^T(\bar{\mathbf{x}})(\mathbf{x} - \bar{\mathbf{x}}) + \frac{1}{2}(\mathbf{x} - \bar{\mathbf{x}})^T \mathbf{H}(\mathbf{x} - \bar{\mathbf{x}}) \right] \\ &= f(\bar{x}) + \frac{1}{2} \sum_i \sum_j \mathbf{H}_{ij} \boldsymbol{cov}(\mathbf{x})_{ij}, \end{aligned} \tag{2.14}$$

where, $\boldsymbol{cov}(\mathbf{x})_{ij}$ is the ijth element of the covariant matrix. Thus, the second order mean includes an additional correction for the difference between the function evaluated at $\widehat{\mathbf{x}}$ and its true value. It is possible that a first order or deterministic model may yield incorrect results even when the estimate of $\mathbf{x}$ is subject to only small errors. However, if the difference between the first order estimate and the second order estimate is large, then the second order estimate may also be inadequate. In such cases, sampling techniques such as the Monte Carlo methods may be required.

The second order estimate of the variance of f may be derived following the procedure used for first order estimate:

$$\begin{aligned}\boldsymbol{var}[f(\mathbf{x})] &= E\Bigg[\Bigg(\nabla f^T(\widehat{\mathbf{x}})(\mathbf{x}-\widehat{\mathbf{x}}) + \frac{1}{2}(\mathbf{x}-\widehat{\mathbf{x}})^T\mathbf{H}(\mathbf{x}-\widehat{\mathbf{x}}) \\ &\qquad - \frac{1}{2}\sum_i\sum_j \mathbf{H}_{ij}\boldsymbol{cov}(\mathbf{x})_{ij}\Bigg)^2\Bigg] \\ &= \nabla f^T(\widehat{\mathbf{x}})\boldsymbol{cov}(\mathbf{x})\nabla f(\widehat{\mathbf{x}}) - \left[\frac{1}{2}\sum_i\sum_j \mathbf{H}_{ij}\boldsymbol{cov}(\mathbf{x})_{ij}\right]^2 \\ &\quad + \sum_i\sum_j\sum_k \frac{\partial f}{\partial x_i}H_{jk}E[(x_i-\widehat{x}_i)(x_j-\widehat{x}_j)(x_k-\widehat{x}_k)] \\ &\quad + \frac{1}{4}\sum_i\sum_j\sum_k\sum_l H_{ij}H_{kl}E[(x_i-\widehat{x}_i)(x_j-\widehat{x}_j)(x_k-\widehat{x}_k)(x_l-\widehat{x}_l)].\end{aligned} \tag{2.15}$$

The third and fourth moments of $\mathbf{x}$ in the above expression are in general difficult to compute. However, if the variables x_i are independent, then further simplification of the above expressions is possible since for independent variables $x_i, x_j, x_k \ldots$

$$E[(x_i-\widehat{x}_i)(x_j-\widehat{x}_j)(x_k-\widehat{x}_k)\ldots] = 0 \quad \text{if } m \neq n \text{ where } m,n \in \{i,j,k\ldots\}. \tag{2.16}$$

Thus, for independent variables, the above expressions for first order and second order methods reduce to those listed in Table 2.7.

2.6. PROBABILISTIC DESIGN FOR MAXIMUM RELIABILITY

Probabilistic design unlike the deterministic design is aimed at the design of the system in the presence of uncertainty [45,46]. The uncertainties in the design variables are assumed to be known apriori and are included during the design process to account for their effects on the system performance. The quantification of the performance of the system is very critical for fiber-optic systems where the system performance is greatly affected by the uncertainty in the variables. The probabilistic design techniques have found widespread use in the design of structures and mechanical systems [47–52]. Their use for design of electronic packages [53–55] and fiber-optic systems to maximize the system reliability is largely missing.

Deterministic designs may be thought of as specifying a safety factor defined as the ratio of mean strength or allowable limit (μ_S) against the mean system response or

TABLE 2.7.
Expressions for first- and second-order estimates of mean and variance of a general function.

	First order estimate	Second order estimate
Mean (μ_f)	$f(\bar{\mathbf{x}})$	$f(\bar{\mathbf{x}}) + \frac{1}{2}\sum_i \frac{\partial^2 f}{\partial x_i^2}\sigma_i^2$
Variance (σ_f^2)	$\sum_i \left(\frac{\partial f}{\partial x_i}\right)^2 \sigma_i^2$	$\sum_i \left(\frac{\partial f}{\partial x_i}\right)^2 \sigma_i^2 - \left(\frac{1}{2}\sum_i \frac{\partial^2 f}{\partial x_i^2}\sigma_i^2\right)^2 + \sum_i \frac{\partial f}{\partial x_i}\frac{\partial^2 f}{\partial x_i^2} E\left[(x_i - \bar{\mathbf{x}}_i)^3\right] + \frac{1}{4}\sum_i \left(\frac{\partial^2 f}{\partial x_i^2}\right)^2 E\left[(x_i - \bar{\mathbf{x}}_i)^4\right]$

where, all the derivatives are evaluated at $\bar{\mathbf{x}}_i$, the summations are over the number of variables, and σ_i is the standard deviation of x_i. The third and fourth moment terms in the second order estimate are very difficult to obtain for real systems and therefore are usually ignored. Clearly, the stochastic analysis of any system imposes the need to characterize the mean and variance of all input parameters x_i. In the case of a photonic system, of all the variables, the material behavior expected to play a dominant role in determining the uncertainty in performance. The statistical description of the viscoelastic behavior of common epoxies used to bond the VCSEL to the substrate is the focus of the following section.

applied load (μ_L). By considering only the mean values of allowable limit and system response, it is not possible to estimate the reliability of the system, since it is statistically possible for the design to fail even if the mean strength is larger than the applied load. In the Probabilistic design procedure, on the other hand, the design variable values are determined such that they satisfy a probabilistically stated reliability criterion. The values of design variables for maximum reliability or minimum probability of failure are determined using the design procedure. In Figure 2.13, μ_L and μ_S indicate the mean values of the applied load and the strength for a system design respectively. The two curves $f_L(l)$ and $f_S(s)$ are the probability density functions corresponding to the applied load and the strength. The overlapping region indicates the region of failure. For the reliability of the system to be maximum, this overlap needs to be minimized as discussed below.

Considering normal distributions for both S (μ_S, σ_S) and L (μ_L, σ_L), another random variable $Z = S - L$ is first introduced. Assuming that S and L are independent, Z has a normal distribution with a mean value of $\mu_S - \mu_L$ and a standard deviation of $\sqrt{\sigma_S^2 + \sigma_L^2}$. The system failure can now be defined as when Z is less than zero. Thus, the probability of failure can be written as

$$p_f = 1 - \Phi\left[\frac{\mu_S - \mu_L}{\sqrt{\sigma_S^2 + \sigma_L^2}}\right],$$

where Φ is the cumulative distribution function (CDF) of the standard normal variable. Here,

$$\frac{\mu_S - \mu_L}{\sqrt{\sigma_S^2 + \sigma_L^2}}$$

is termed as the safety margin or reliability index and is denoted by β.

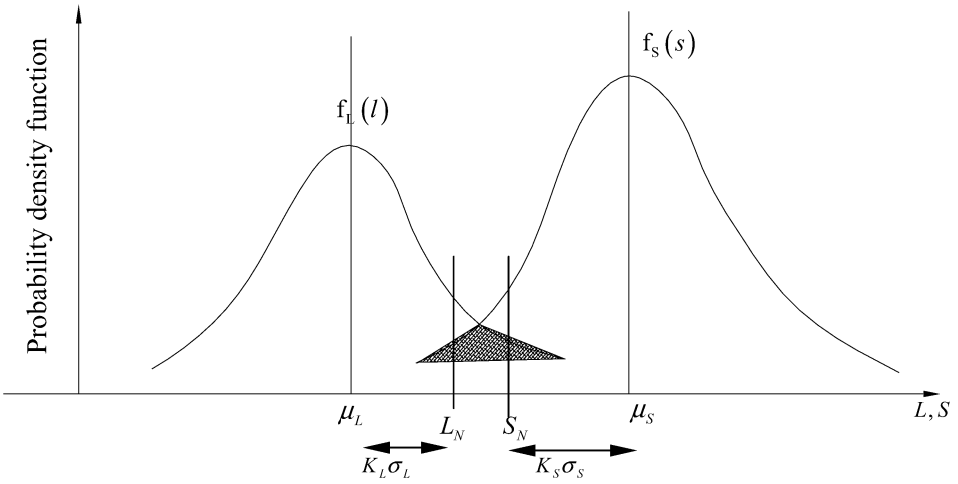


FIGURE 2.13. An illustration of load-strength interference on which the probabilistic design is based.

Thus, the objective function for probabilistic design is formulated as:

$$\max : \beta, \tag{2.17}$$

where, $\beta = g(f > C)$, where f is the coupling efficiency and C is the specified acceptable value of f. Here, we define the failure of the system to occur when the coupling efficiency drops below 100%; the goal of the optimization is thus to minimize the probability that coupling efficiency drops below 100%. The specific form for β can be written as

$$\beta = \frac{\mu_{CE} - 1}{\sqrt{\sigma_{CE}^2 + 0}}.$$

In maximizing β, μ_{CE} is maximized while σ_{CE} is minimized. We state that Equation (2.17) can be reformulated alternatively as the multi-objective optimization problem shown below:

$$\max : c_1\mu_{CE} - c_2\sigma_{CE}. \tag{2.18}$$

The multi-objective formulation possesses the advantage of allowing the tradeoff between the optimal and probabilistic designs. The mean value of coupling efficiency (μ_{CE}) is given by the function f in Equation (2.5). The standard deviation of coupling efficiency (σ_{CE}) is determined by using the first order approximation given by

$$\sigma_{CE} = \sqrt{\sum_i \left(\frac{\partial f}{\partial x_i}\right)\sigma_{x_i}^2}. \tag{2.19}$$

For purposes of illustration, the standard deviation of the design variables (major and minor diameter of the curved reflector and position of curved reflector with respect to the VCSEL) was chosen here to be 10% of the value of the core diameter of the optical fiber

for multi-mode operation and 20% for the single-mode operation respectively. The optimization of the above function was also performed using the simulated annealing algorithm implemented within Mathematica.

2.6.1. Results

The results of the probabilistic design are listed in Tables 2.8 and 2.9. For the sake of easy comparison, the results of deterministic design listed in Tables 2.5 and 2.6 are repeated here. It is clear from the tables that probabilistic optimal design yields a solution that is identical to the deterministic optimal design. Monte Carlo simulation was carried out at the final design to determine the coupling efficiency uncertainty given the uncertainty in the design variables (Figures 2.14 and 2.15). Figures 2.14 and 2.15 consider only the effect of geometric uncertainty on the coupling efficiency. The optimization of the curved reflector is performed in 2-dimensions and accounts for the effect of coupling loss due to misalignment only in the lateral direction. The material uncertainty, however, results in misalignment of the optical fibers along the longitudinal direction of the curved reflector and is thus not included in the optimization of the shape and size of the curved reflector.

The deterministic and probabilistic optimal and robust design all produce a much improved coupling efficiency performance as compared to the initial design. The coupling efficiency varies in the range of 40–100% (considering three-sigma rule) for the initial design. This variation is reduced to 75–100% and 80–100% for the optimal and robust designs respectively. The robust design procedure performs better than optimal design during the early design stage when the uncertainty in the design variables is usually unknown. The probabilistic design can be used to modify the robust design when experimental data are available to quantify the uncertainty in the design variables. The probabilistic design is however a natural choice when the uncertainties in the design variables are known apriori.

Similar comparison plots are shown for c_1 and c_2 values of 0.5 for robust and probabilistic design (Figures 2.16 and 2.17). The increase in standard deviation values for robust

TABLE 2.8.
Results of deterministic/probabilistic optimal, deterministic robust and probabilistic robust designs for multi-mode operation.

Method	p (μm)	q (μm)	x_0 (μm)	Coupling efficiency %
Initial design	300	300	200	90
Deterministic/probabilistic optimal design	600	308	530	100
Deterministic robust design	598	331	520	100
Probabilistic robust design	600	388	501	100

TABLE 2.9.
Results of deterministic/probabilistic optimal, deterministic robust and probabilistic robust designs for single-mode operation.

Method	p (μm)	q (μm)	x_0 (μm)	Coupling efficiency %
Initial design	300	300	200	88
Deterministic/probabilistic optimal design	600	270	545	100
Deterministic robust design	600	275	543	100
Probabilistic robust design	600	300	534	100

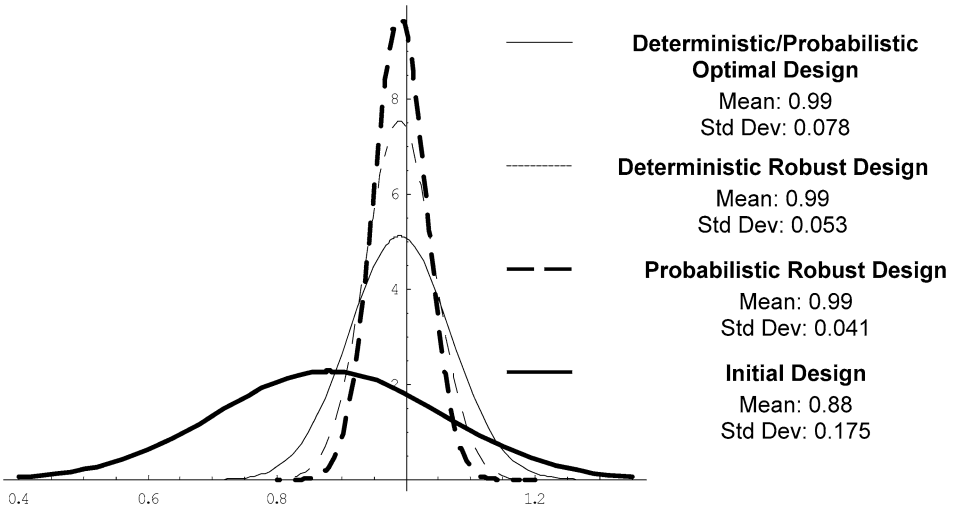


FIGURE 2.14. A comparison of coupling efficiency variation between deterministic/probabilistic optimal, deterministic robust, probabilistic robust and initial design for multi-mode operation.

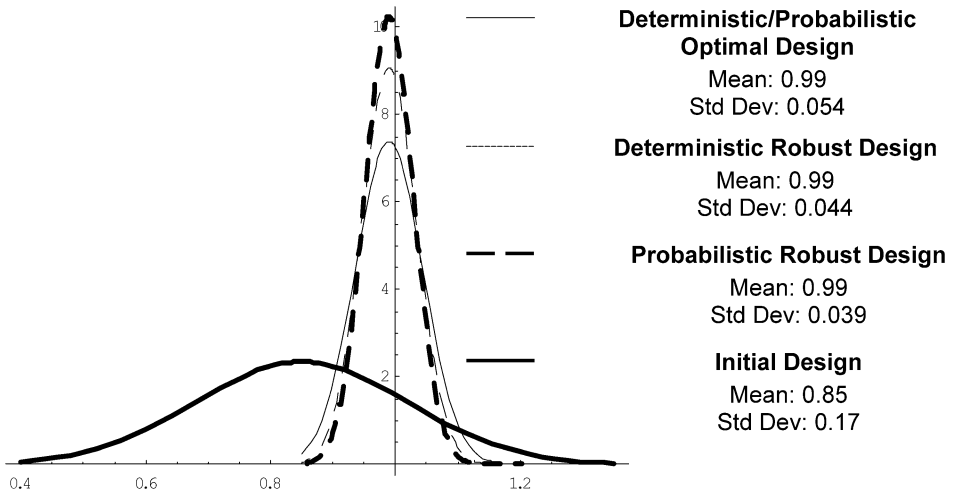


FIGURE 2.15. A comparison of coupling efficiency variation between deterministic/probabilistic optimal, deterministic robust, probabilistic robust and initial design for single-mode operation.

and probabilistic values is a result of trade off with respect to the optimal design. The results indicate that the standard deviation values for all the methods are within a small range. It can be seen that the robust design performs marginally better than the probabilistic design for both the multi-mode and single mode operation. Thus for equal weightage of optimal and robust design, the deterministic design techniques are more desirable as they result in comparable results with respect to probabilistic design and saves time and cost in computing the uncertainty in design variables.

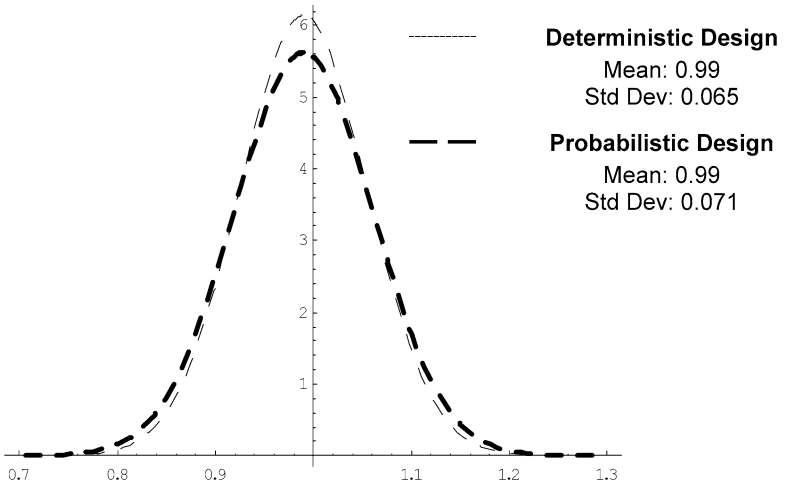


FIGURE 2.16. A comparison of coupling efficiency variation between deterministic and probabilistic designs for multi-mode operation with trade off.

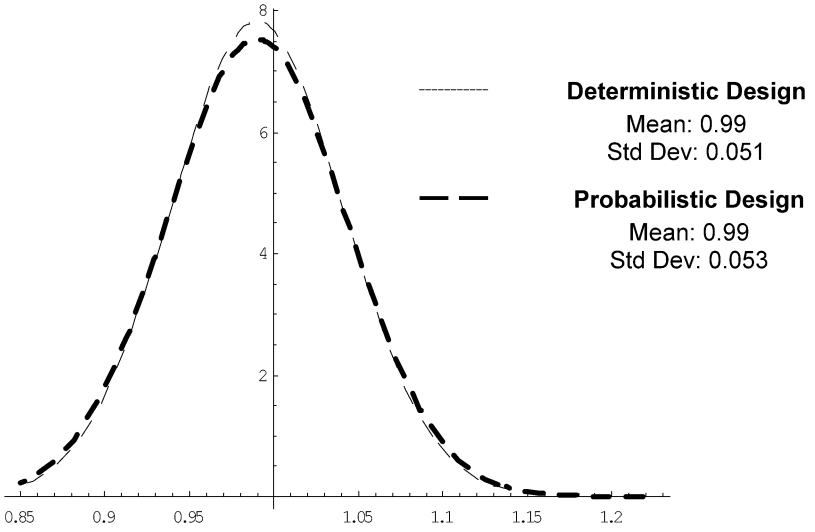


FIGURE 2.17. A comparison of coupling efficiency variation between deterministic and probabilistic designs for single-mode operation with trade off.

2.7. STOCHASTIC CHARACTERIZATION OF EPOXY BEHAVIOR

A typical fiber-optic package experiences high operating temperatures in the range of 70–100°C. The epoxies used to bond the laser to the substrate undergo temperature induced creep behavior due to the operating temperature. The displacement of the laser with respect to the optical fiber induced due to the temperature induced creep behavior of the epoxy could result in significant coupling efficiency loss. A displacement of one radius value of the optical fiber could result in coupling efficiency drop of up to 70% (Figure 2.5). The uncertainty in the material behavior combined with the uncertainty in the geometric

parameters would further affect the coupling loss. A stochastic characterization of material uncertainty is thus required to determine its effect on system behavior and thus to enable design decisions. In the following, a representative epoxy material EMI Emcast 501 is stochastically characterized.

2.7.1. Viscoelastic Models

Material models are necessary to parametrically describe material behavior. Towards this end, we begin by describing popular one-dimensional viscoelastic models. The time dependent (viscoelastic) behavior of the epoxies can be simulated through elastic springs and viscous dashpots. The most common models used to describe the viscoelastic behavior are the Maxwell, Kelvin and the Standard solid model. The spring and the dashpot are set in series configuration for Maxwell model and in parallel configuration for Kelvin model. The standard model consists of either the Kelvin model in series with a spring or a Maxwell model in parallel with a spring. The standard solid model is used in this study since the behavior of the epoxy is more similar to solid than liquid.

The subscripts 2 and 1 used in Figure 2.18 represent the parallel Kelvin arrangement and the single spring element respectively.

For the arrangement in series as shown in Figure 2.18,

$$\begin{aligned} &\sigma = \sigma_1 = \sigma_2, \\ &\varepsilon = \varepsilon_1 + \varepsilon_2. \end{aligned} \tag{2.20}$$

The strains in the spring 1 and in the Kelvin arrangement are given by

$$\varepsilon_1 = \frac{\sigma_1}{E_1}, \quad \varepsilon_2 = \frac{\sigma_2}{E_2 + \eta_2 D}. \tag{2.21}$$

Substituting the above equations into the equation for total strain we get the equation of strain and strain rate related to the stress and stress rate for the standard model shown in Figure 2.18 as

$$\dot{\varepsilon}\eta_2 + \varepsilon E_2 = \left(\frac{E_2 + E_1}{E_1}\right)\sigma + \frac{\eta_2}{E_1}\dot{\sigma}. \tag{2.22}$$

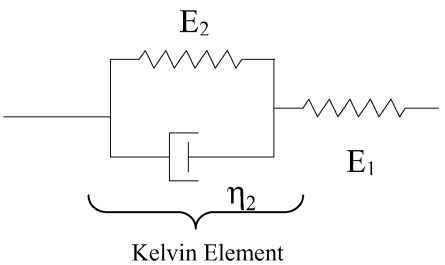


FIGURE 2.18. Standard solid model used to describe EMI Emcast 501 epoxy.

2.7.2. Modeling the Creep Test

The epoxy between the VCSEL and the substrate is subjected to creep induced by high temperatures. Thus the standard solid model described above and governed by Equation (2.22) will have to be used to determine its creep behavior. Creep test is a standard procedure wherein a stress τ_0 is applied at time $t = 0$ on the sample and then maintained constant thereafter. This test can be modeled as a function of time with the aid of the unit step function $[u(t)]$. Thus, during creep test

$$\tau(t) = \tau_0[u(t)]. \tag{2.23}$$

Substituting for $\tau(t)$ in Equation (2.22) and applying Laplace transform, we get

$$\eta_2 s\bar{\varepsilon}(s) + \bar{\varepsilon}(s)E_2 = \left(\frac{E_2 + E_1}{E_1}\right)\bar{\tau}(s) + \frac{\eta_2}{E_1}s\bar{\tau}(s). \tag{2.24}$$

In the case of creep test, $\tau(t) = \tau_0[u(t)]$,

thus $\bar{\tau}(s) = \dfrac{\tau_0}{s}$.

Thus, Equation (2.24) simplifies to

$$\eta_2 s\bar{\varepsilon}(s) + \bar{\varepsilon}(s)E_2 = \left(\frac{E_2 + E_1}{E_1}\right)\frac{\tau_0}{s} + \frac{\eta_2}{E_1}\tau_0,$$

$$\bar{\varepsilon}(s)(\eta_2 s + E_2) = \frac{\tau_0}{E_1}\left[(E_2 + E_1)\frac{1}{s} + \eta_2\right], \tag{2.25}$$

$$\bar{\varepsilon}(s) = \frac{\tau_0}{E_1}\left\{\frac{E_2 + E_1}{\eta_2}\left[\frac{1}{s\left(s + \dfrac{E_2}{\eta_2}\right)}\right] + \frac{1}{\left(s + \dfrac{E_2}{\eta_2}\right)}\right\}, \tag{2.26}$$

$$\bar{\varepsilon}(s) = \frac{\tau_0}{E_1}\left[\frac{E_2 + E_1}{E_2}\left(\frac{1}{s} - \frac{1}{s + \dfrac{E_2}{\eta_2}}\right) + \frac{1}{s + \dfrac{E_2}{\eta_2}}\right]. \tag{2.27}$$

Now, applying inverse Laplace transform, we get

$$\varepsilon(t) = \frac{1}{E_1}\left[\frac{E_2 + E_1}{E_2}\left(1 - e^{-(\frac{E_2}{\eta_2})t}\right) + e^{-(\frac{E_2}{\eta_2})t}\right]\tau_0[u(t)], \tag{2.28}$$

$$\varepsilon(t) = \frac{1}{E_1}\left[\frac{E_2 + E_1}{E_2} - \frac{E_1}{E_2}e^{-(\frac{E_2}{\eta_2})t}\right]\tau_0[u(t)]. \tag{2.29}$$

Equation (2.29) above gives the strain in the epoxy when a constant shear stress τ_0 is applied. Thus, given the applied stress on the epoxy we can calculate the strain and

hence the misalignment of the laser beams with the optical fiber as a function of time. The stress experienced by the epoxy is caused by the temperature induced differential expansion between the VCSEL and the substrate. This will be determined later for a bonded three layer system.

2.7.3. Dynamic Mechanical Analysis

The epoxy was characterized using a Dynamic Mechanical Analyzer (DMA). The stress–strain behavior is characterized as a function of temperature as well as frequency. A sinusoidal stress or strain is applied to the material and the output response is measured. Viscoelastic materials exhibit a lag in the output response which is characterized by the phase shift δ as shown in Figure 2.19. The DMA provides two outputs namely the storage modulus which indicates the elasticity of the material and the amount of energy it can store when a stress is applied and a loss modulus which indicates the viscous property of the material and the amount of energy lost to friction and internal motions. The phase lag terms relate the storage and the loss moduli. The tangent of the phase angle δ gives the ratio of loss modulus to storage modulus

$$\tan\delta = \frac{E''}{E'}. \tag{2.30}$$

Equation (2.22) was used to fit the two outputs obtained from the DMA tests to obtain the parameters of the model shown in Figure 2.18. The model parameters E_1, E_2 and η_2 are temperature dependent. The material parameters E' and E'' obtained from the DMA tests however depend on both temperature as well as the frequency at which the stress or strain is applied.

The epoxy samples were tested at four different temperatures of 50°C, 80°C, 110°C, and 150°C. A sinusoidal displacement with maximum amplitude of 5 μm was applied over a frequency range of 0.01 Hz to 10 Hz (0.01 Hz is the minimum value of frequency that could be applied with the equipment that was used in this study).

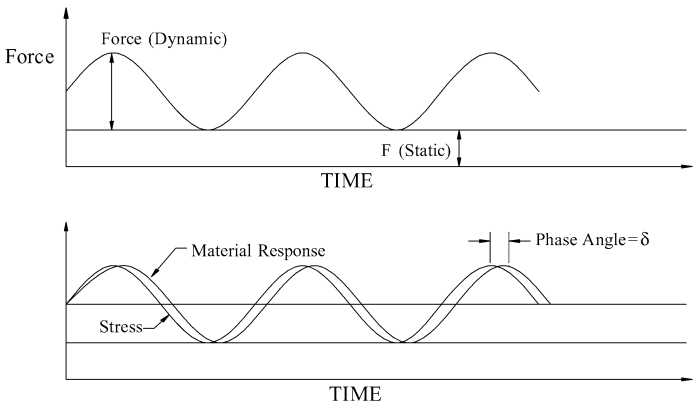


FIGURE 2.19. Principle of DMA.

The equation of modulus for the model in Figure 2.18 can now be obtained by applying a sinusoidal strain to Equation (2.31).

$$\dot{\varepsilon}\eta_2 + \varepsilon E_2 = \left(\frac{E_2 + E_1}{E_1}\right)\sigma + \frac{\eta_2}{E_1}\dot{\sigma}, \tag{2.31}$$

$$\varepsilon = \varepsilon_0 e^{i\omega t}.$$

The resultant stress is then

$$\sigma = \sigma_0 e^{i(\omega t+\delta)}. \tag{2.32}$$

Substituting Equation (2.32) in to Equation (2.31), we get

$$\varepsilon_0 E_2 e^{i\omega t} + i\omega\varepsilon_0\eta_2 e^{i\omega t} = \frac{1}{E_1}\left[\sigma_0(E_1 + E_2)e^{i(\omega t+\delta)} + i\omega\sigma_0\eta_2 e^{i(\omega t+\delta)}\right], \tag{2.33}$$

$$E_c = \frac{\sigma(t)}{\varepsilon(t)} = \frac{E_1(E_2 + i\omega\eta_2)}{(E_1 + E_2 + i\omega\eta_2)},$$

$$E_c = \frac{E_1(E_1E_2 + E_2^2 + w^2\eta_2^2)}{(E_1 + E_2)^2 + w^2\eta_2^2} + i\frac{E_1^2 w\eta_2^2}{(E_1 + E_2)^2 + w^2\eta_2^2}, \tag{2.34}$$

where, E_c is the complex modulus of the viscoelastic material dependent on the frequency of the applied stress or strain. The real part of Equation (2.34) is the storage modulus and the imaginary part is the loss modulus. The constants in Equation (2.34) can be found by fitting the equation to the experimental data with respect to frequency.

2.7.4. Experimental Results

Figures 2.20–2.21 show the data fit and the parameters E_1, E_2 and η_2 obtained for one of the 40 samples tested at 50°C, 80°C, 110°C and 150°C respectively. The curve fit was done with respect to the magnitude of the complex modulus of the experimental data.

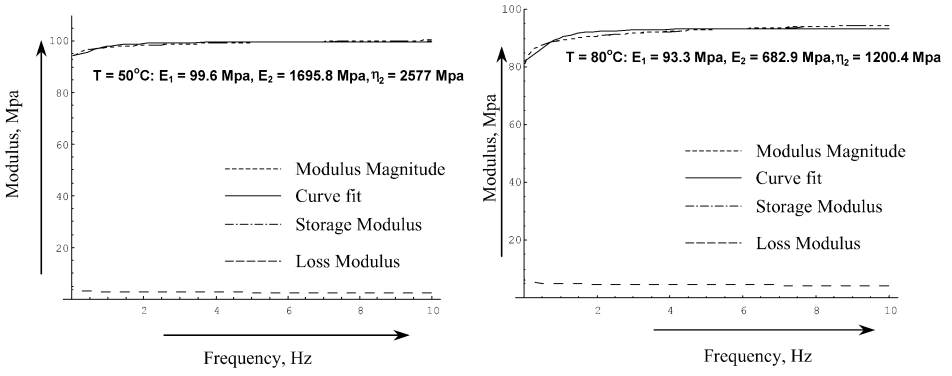


FIGURE 2.20. A fit of the measured modulus value of EMI Emcast epoxy as a function of both frequency and temperature (left plot at 50°C and right plot at 80°C).

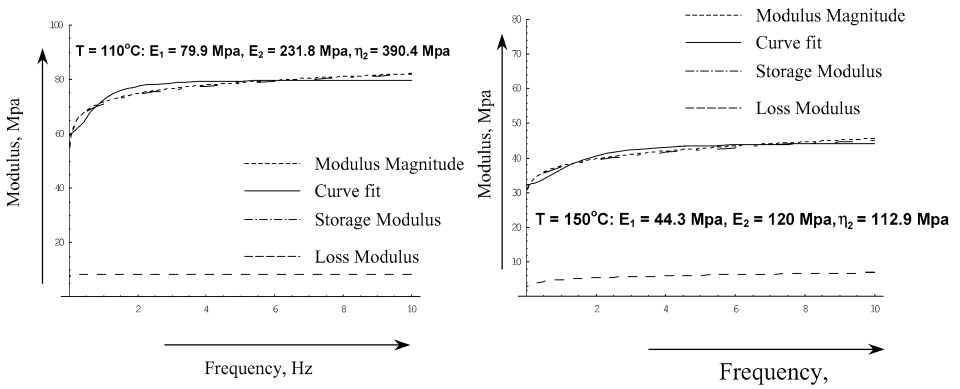


FIGURE 2.21. A fit of the measured modulus value of EMI Emcast epoxy as a function of both frequency and temperature (left plot at 110°C and right plot at 150°C).

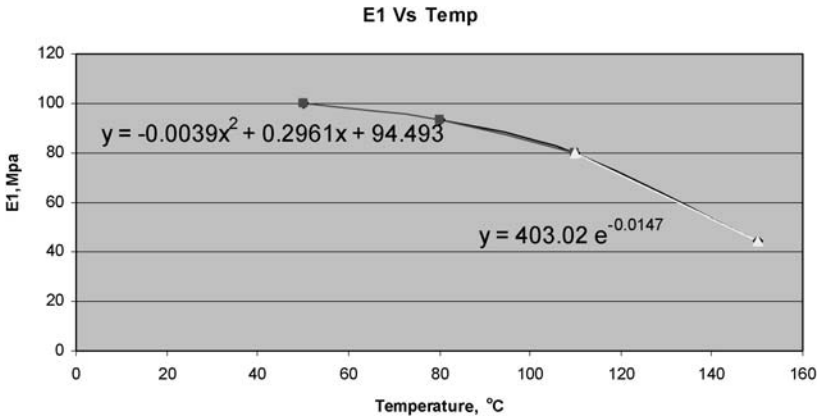


FIGURE 2.22. A fit of the measured value of parameter E_1 as a function of temperature.

The storage modulus was very close to the magnitude of complex modulus as can be seen in Figures 2.20–2.21. The loss modulus values are also plotted in Figures 2.20–2.21. The loss modulus value was in the range of 2.5%–10% of the storage modulus for the temperature range from 50°C to 150°C. The standard solid parameters E_1, E_2 and η_2 were next fitted with respect to temperature to describe the model parameters over the entire temperature range of 50°C to 150°C. Quadratic and an exponential description provided fits with least error for E_1 between 50°C to 110°C and between 80°C and 150°C respectively. A cubic description was the best fit for parameters E_2 and η_2 over the entire temperature range. This data fit with respect to temperature for one of the samples is shown in Figures 2.22–2.24. The data fits shown in Figures 2.20–2.21 are however for only one of the 40 tests we conducted. The stochastic variations of the parameters over the 40 tests are shown in Figures 2.25–2.27.

Similar plots of variation of the material parameters were determined at temperatures 50°C, 110°C and 150°C. The variation in material parameters at intermediate temperatures can be determined using the data fit curves similar to those in Figures 2.22–2.24.

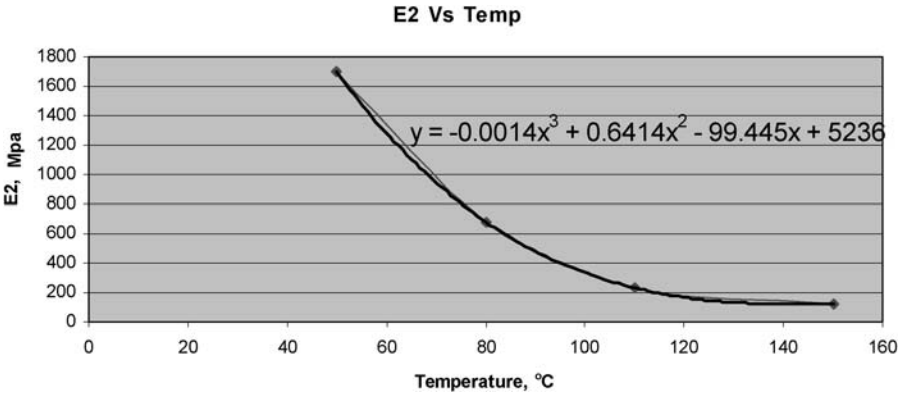


FIGURE 2.23. A fit of the measured value of parameter E_2 as a function of temperature.

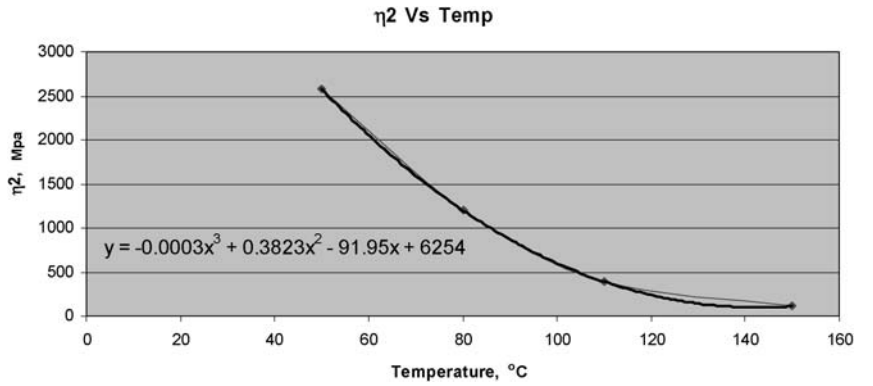


FIGURE 2.24. A fit of the measured value of parameter η_2 as a function of temperature.

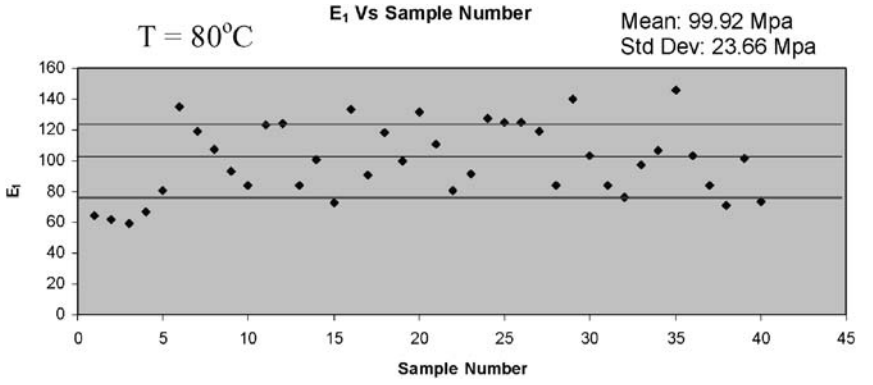


FIGURE 2.25. Variation of E_1 at 80°C.

2.8. ANALYTICAL MODEL TO DETERMINE VCSEL DISPLACEMENT

The creep model developed in the previous section can be used to determine the magnitude of the shear displacement in the epoxy. The calculation of displacement of the epoxy

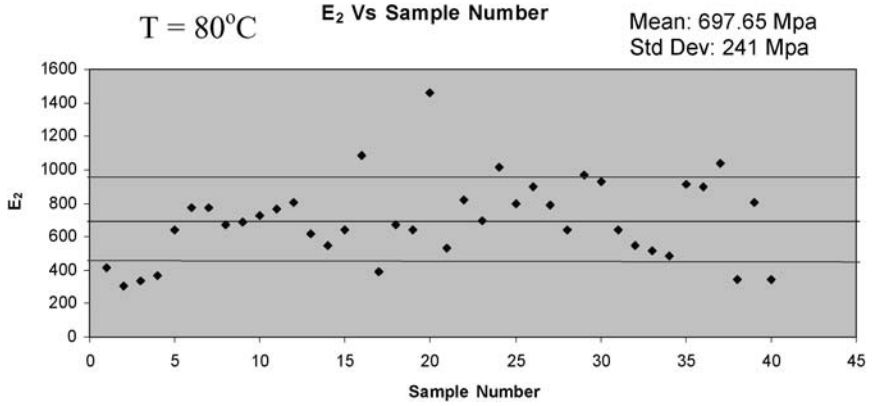


FIGURE 2.26. Variation of E_2 at 80°C.

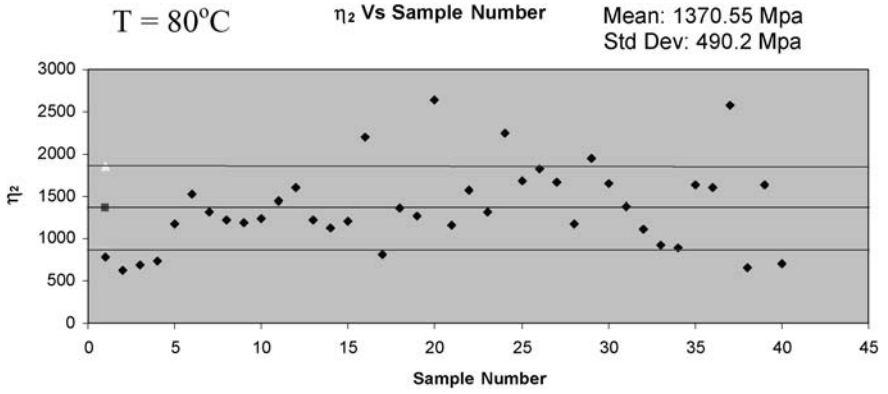


FIGURE 2.27. Variation of η_2 at 80°C.

requires the determination of its stress state. The stresses that are acceptable must satisfy static equilibrium. The simple one-dimensional derivation below determines the displacement by balancing the forces on each layer of the assembly. The derivation is an extension of the classical Timoshenko theory of bi-metal thermostats. The stress distribution analysis of bonded assemblies for electronic packages was first considered by Chen and Nelson [56]. The following derivation was obtained by Suhir [57] for bi-metal thermostats and later extended for tri-material assemblies [58,59].

In Figure 2.28, the 2D analytical model of the VCSEL, epoxy and the substrate are shown. Assuming that the bond epoxy is "soft" relative to the component and the substrate, the forces acting on the component, epoxy and on the substrate can be modeled as shown in Figure 2.29.

The displacement compatibility for the epoxy is given by

$$u_c(x) = u_s(x) - 2\kappa_e \tau(x). \tag{2.35}$$

The last term in Equation (2.35) accounts for the non-uniform distribution of forces and are calculated under the assumption that the corresponding corrections are directly

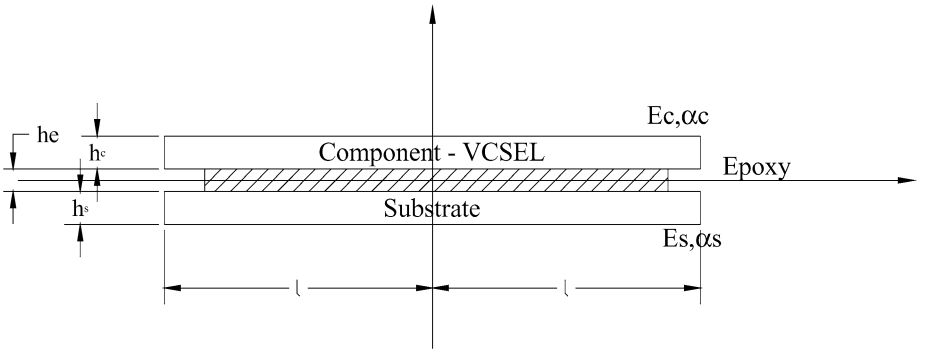


FIGURE 2.28. Two-dimensional model of (the cross-section of) VCSEL, epoxy and substrate.

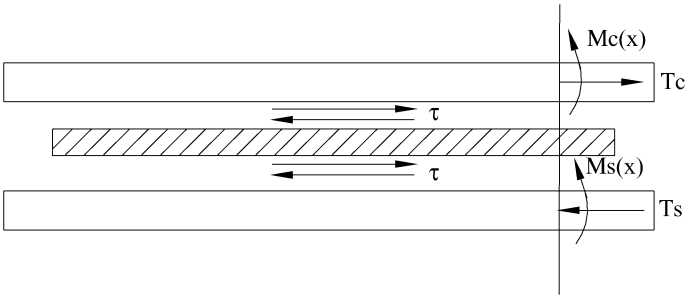


FIGURE 2.29. A free-body diagram of the VCSEL-epoxy-substrate showing the stresses resulting from differential thermal expansion.

proportional to the shearing force in the given cross-section; a further assumption is that the corrections are not affected by the shearing forces in the other cross-sections. Constant κ_e is the interfacial compliance of the epoxy layer.

The displacements $u_c(x)$ and $u_s(x)$ of the component and the substrate are given by

$$u_c(x) = \alpha_c \Delta T x - \lambda_c \int_0^x T_c(\xi) d\xi + \kappa_c \tau(x) + \frac{h_c}{2} \int_0^x \frac{d\xi}{\rho(\xi)}, \tag{2.36}$$

$$u_s(x) = \alpha_s \Delta T x - \lambda_s \int_0^x T_s(\xi) d\xi + \kappa_s \tau(x) + \frac{h_s}{2} \int_0^x \frac{d\xi}{\rho(\xi)}. \tag{2.37}$$

The first terms in Equations (2.36) and (2.37) are the unrestricted thermal expansions of the component and the substrate. The second terms are due to the normal forces and are calculated under the assumption that these forces are uniformly distributed over the thickness. The third terms were motivated earlier in Equation (2.35) and the last terms arise out of the curvature (ρ s the local radius) caused by the bending. Also, due to force balance,

$$T_c(x) = T_s(x) = \int_{-l}^{x} \tau(\xi) d\xi, \tag{2.38}$$

where, $T_c(x)$ and $T_s(x)$ are the normal force (per unit assembly width) along the length direction and α_s and α_c are the thermal expansion coefficients. Further,

$$\lambda_c = \frac{1-\nu_c}{E_c h_c}, \quad \lambda_s = \frac{1-\nu_s}{E_s h_s} \tag{2.39}$$

are the axial compliances of the component and the substrate and

$$\kappa_c = \frac{h_c}{3G_c}, \quad \kappa_s = \frac{h_s}{3G_s} \tag{2.40}$$

are their interfacial compliances,

$$\kappa_e = \frac{h_e}{3G_e} \tag{2.41}$$

is the interfacial compliance of the bond epoxy, h_i are the thicknesses of the three elements of the assembly.

For the earlier described standard viscoelastic solid, the shear modulus G_e of the epoxy is given by Equation (2.29), which is repeated here:

$$\frac{1}{G_e} = \frac{1}{E_2}\left[\frac{E_2+E_1}{E_1} - e^{-(\frac{E_2}{\eta_2})t}\right]. \tag{2.42}$$

The moment balance in the mid-layer of the assembly, at any location x along the length of the assembly yields the following equation:

$$\frac{h_c+h_e}{2}T_c(x) + \frac{h_s+h_e}{2}T_s(x) = M_c(x) + M_s(x), \tag{2.43}$$

where,

$$M_i(x) = \frac{D_i}{\rho(x)}, \quad D_i = \frac{E_i h_i^3}{12(1-\nu_i^2)},$$

where, D_i are the flexural rigidity.

If $h = h_c + h_s$, if $h_e \ll h_c, h_s$, $D = D_c + D_s$:

$$\therefore \ \frac{1}{\rho(x)} = -\frac{h}{2D}T(x). \tag{2.44}$$

Substituting Equations (2.36, 2.37 and 2.44) in Equation (2.35), we get

$$\tau(x) - \frac{1}{L^2}\int_0^x T(\xi)d\xi = \frac{\Delta\alpha\Delta T x}{\kappa}, \tag{2.45}$$

where,

$$\frac{1}{L^2} = \frac{\lambda}{\kappa}, \quad \lambda = \lambda_c + \lambda_s + \frac{h^2}{4D}, \quad \kappa = \kappa_c + \kappa_s + 2\kappa_e.$$

Solving Equation (2.45) and applying the boundary condition, $\tau(0) = 0$, $T(l) = 0$, we get

$$\tau(x) = \frac{L}{\kappa}(\alpha_c - \alpha_s)\Delta T \frac{\sinh\left(\frac{x}{L}\right)}{\cosh\left(\frac{l}{L}\right)}. \tag{2.46}$$

Substituting Equation (2.46) into Equation (2.29), we get

$$\frac{u_c - u_s}{h_e} = \frac{1}{E_2}\left[\frac{E_2 + E_1}{E_1} - e^{-(\frac{E_2}{\eta_2})t}\right]\frac{(\alpha_c - \alpha_s)\Delta T L}{\kappa}\frac{\sinh\left(\frac{x}{L}\right)}{\cosh\left(\frac{l}{L}\right)}. \tag{2.47}$$

Substituting Equations (2.41) and (2.42) into Equation (2.47) and rearranging, we get:

$$\frac{u_c}{L} = \frac{u_s}{L} + \frac{3\kappa_e}{\kappa}(\alpha_c - \alpha_s)\Delta T \frac{\sinh\left(\frac{x}{L}\right)}{\cosh\left(\frac{l}{L}\right)}. \tag{2.48}$$

Equation (2.48) gives the normalized displacement of the VCSEL due to the viscoelastic deformation of the epoxy. This normalized displacement is used to determine the misalignment of the optical beams with respect to the optical fibers. This enables the calculation of the coupling loss along the length of the VCSEL array.

It is of interest to compare the above result against the pioneering derivation for stresses in bonded layers by Chen and Nelson [56]. Chen and Nelson's model considered only the normal stresses on the adherents, only the shear stresses on the adhesives and ignored the peeling stresses and the bending moments on the adherents. The equilibrium equations were solved along with the stress–strain constitutive equations to determine the displacements and stresses on the bonded layers. The shear stress and displacement of VCSEL derived using Chen and Nelson model in a manner analogous to the above derivation are:

$$\tau = \frac{\tilde{L}}{3\kappa_e}(\alpha_c - \alpha_s)\Delta T \frac{\sinh\left(\frac{x}{\tilde{L}}\right)}{\cosh\left(\frac{l}{\tilde{L}}\right)}, \tag{2.49}$$

and

$$\frac{u_c}{\tilde{L}} = \frac{u_s}{\tilde{L}} + (\alpha_c - \alpha_s)\Delta T \frac{\sinh\left(\frac{x}{\tilde{L}}\right)}{\cosh\left(\frac{l}{\tilde{L}}\right)}, \tag{2.50}$$

where,

$$\frac{1}{\tilde{L}^2} = \frac{G_e}{h_e}\left(\frac{1}{E_s h_s} + \frac{1}{E_c h_c}\right). \tag{2.51}$$

A comparison of the results for the maximum shear stress value for an elastic system using Chen and Nelson [56] model, Suhir's [57–59] model and Finite Element Analysis (FEA) is given below. The comparison was made for varying thickness and modulus values of the epoxy layer.

It can be seen from Figures 2.30 and 2.31 that the maximum shear stress value predicted by the Suhir model agrees more closely with the FEA results as compared to Chen and Nelson's model for a wide range thickness and modulus values of the epoxy. Chen

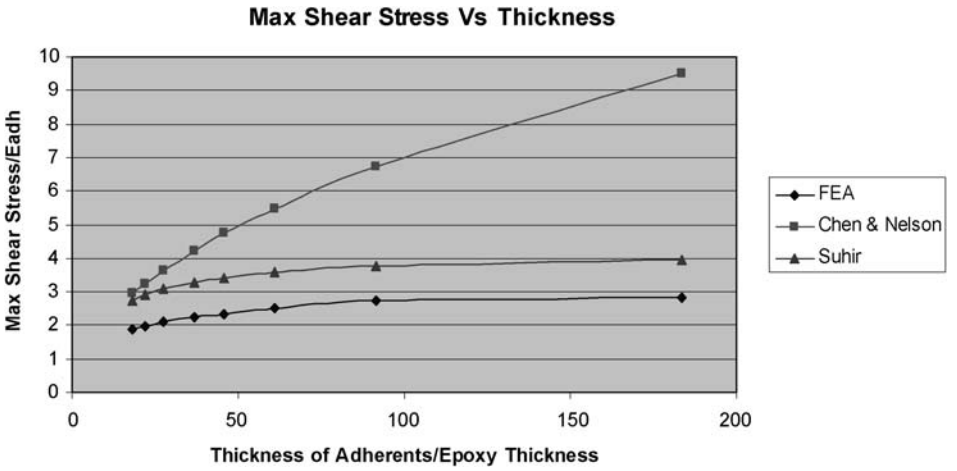


FIGURE 2.30. A plot of maximum shear stress value with respect to the thickness of the epoxy layer determined using Chen and Nelson's model [56], Suhir's model [57–59] and FEA.

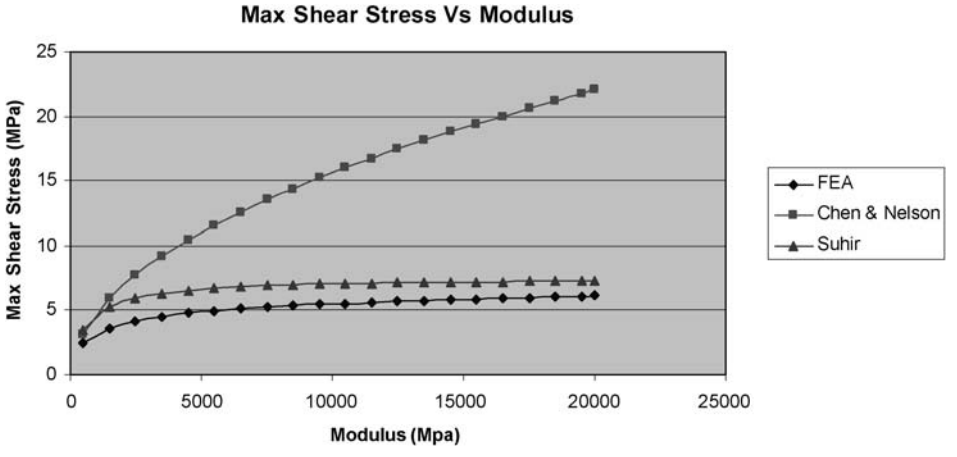


FIGURE 2.31. A plot of maximum shear stress value with respect to the modulus of the epoxy determined using Chen and Nelson's model [56], Suhir's model [57–59] and FEA.

and Nelson model, however, does yield results that are close to those predicted through FEA when the modulus of the epoxy is low and when the thickness of the epoxy layer is relatively large (but smaller than the thickness of the adherents). This is because Chen and Nelson model is applicable only when the ratio of thickness of adherents to the thickness of the epoxy is greater than 10 and less than 30. This validity of Chen and Nelson model over a very small range of epoxy modulus can be explained by the fact that the Chen and Nelson model considers only the normal stresses on the adherent layers and ignores the bending moment caused by the differential thermal expansion between the layers.

2.8.1. Results

The plot for normalized displacement of VCSEL is shown in Figure 2.32. It can be seen that a VCSEL located at the end of the array will undergo exponentially large displacement relative to the device in the center of the array. This displacement of the VCSEL is next used to calculate the coupling efficiency loss [60]. The optical model developed earlier was used to determine the coupling efficiency loss. The extent of coupling efficiency loss when a beam shifts with respect to the optical fiber is shown in Figure 2.33. The coupling efficiency loss is plotted with respect to the normalized beam shift in Figure 2.33. It can be seen that the coupling efficiency does not change initially; however, it drops sharply beyond a certain value of normalized beam shift. This is because the optical beam when it enters the optical fiber has a minimum spot size. The minimum spot size for single-mode operation is $\sim$7 μm. Thus the coupling efficiency drops only when the displacement of the VCSEL causes the minimum spot size to shift beyond the diameter of the optical fiber. The coupling loss due to the displacement along the VCSEL array is shown in Figure 2.34. The coupling loss is plotted at four different temperatures. The multiple curves at each temperature represent the increase in loss with increase in time at that temperature. Thus, it can be seen that the coupling loss would be maximum for a VCSEL placed at the end of an array and would further increase with time. The coupling loss would also increase with increase in temperature. Figure 2.34 can be used as a design chart to determine the maxi-

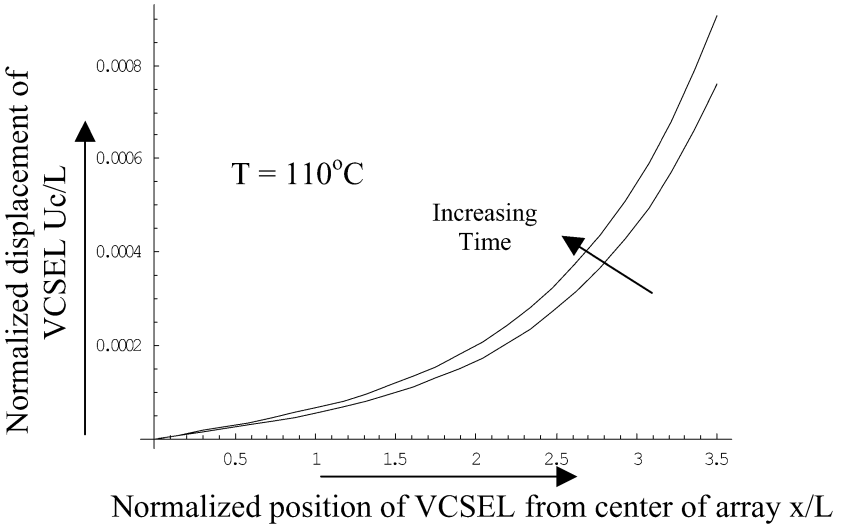


FIGURE 2.32. Normalized displacement as a function of position along the VCSEL array.

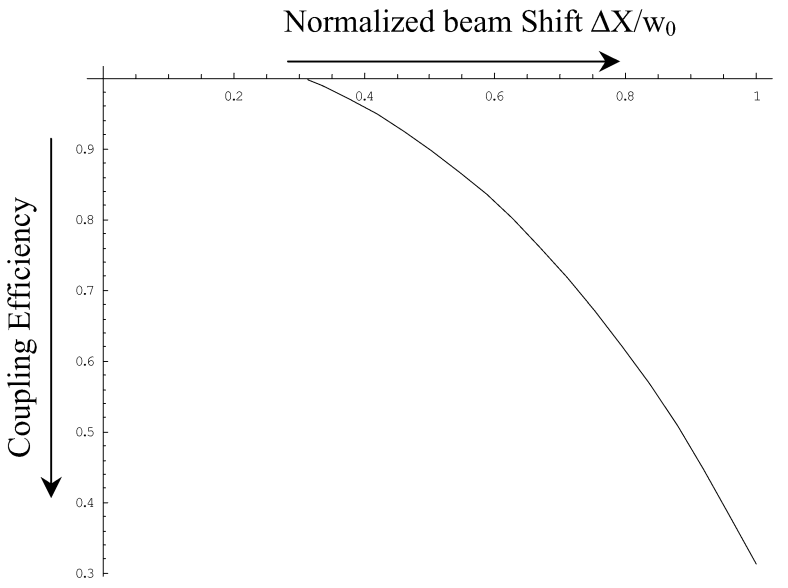


FIGURE 2.33. Coupling efficiency change due to normalized beam shift.

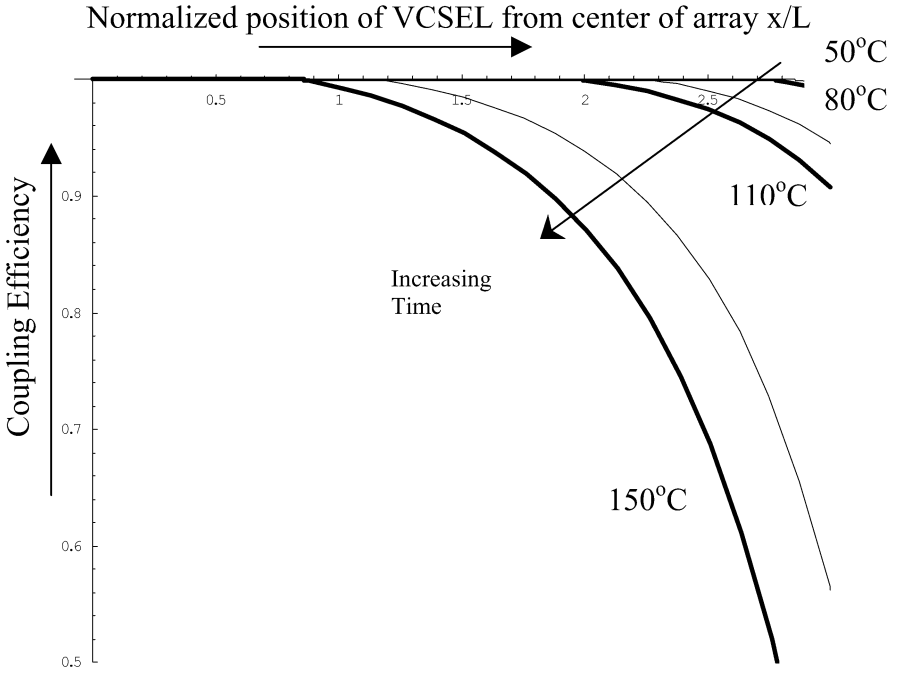


FIGURE 2.34. Coupling efficiency for VCSELs located at various positions along the length of the array at different temperatures.

mum allowable length of a VCSEL array operating at a particular temperature to achieve an acceptable coupling loss.

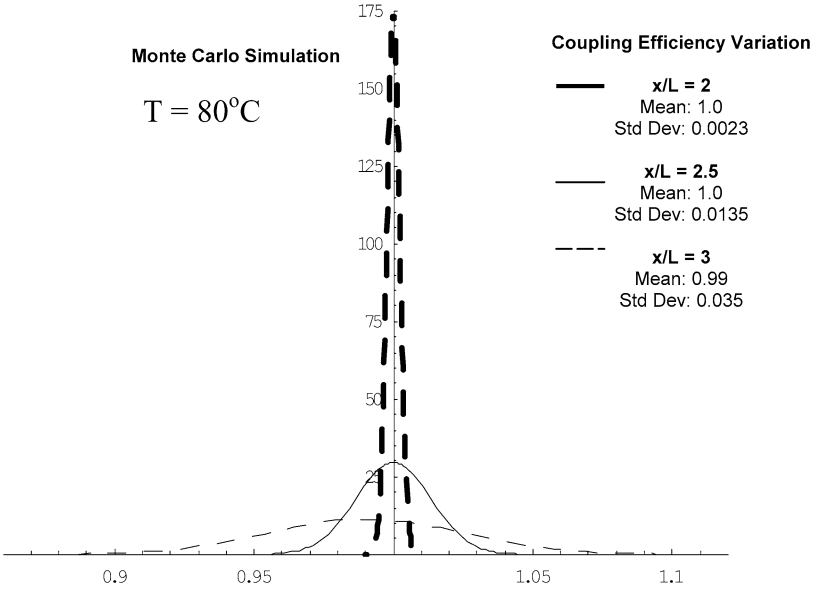


FIGURE 2.35. Stochastic variation of coupling efficiency as a result of material uncertainty for VCSELs located at various positions along the length of the array.

The stochastic variation of the material parameters E_1, E_2 and η_2 can be used to determine the stochastic variation in the coupling efficiency. We performed Monte Carlo simulation to determine the stochastic variation in coupling efficiency caused by the stochastic variation in material behavior. A total of 2000 simulations were conducted within Mathematica [12]. The input variation in material parameters E_1, E_2 and η_2 were propagated through Equation (2.48) to determine the stochastic variation in the VCSEL displacement. The stochastic variation in displacement was then used as input to the optical model to determine the variation in the coupling efficiency. The variation in the coupling efficiency is plotted for three values of normalized position (Figure 2.35) along the VCSEL array. It can be seen from Figure 2.35 that for the end VCSEL of a longer array, the mean coupling efficiency decreases marginally and the standard deviation increases exponentially relative to that for a shorter array. Thus the uncertainty in coupling efficiency is larger at the end of the VCSEL as compared to the center of the array. This again enables design decisions since the probability that a certain prescribed coupling efficiency will be achieved can now be determined.

As explained earlier, the uncertainty in the material behavior is not included in the optimization procedure. The uncertainty in coupling efficiency resulting from material behavior is combined with the geometric uncertainty and evaluated for initial and probabilistic design (Figure 2.36). The effect of material behavior and its uncertainty is a function of temperature and has significant effect on the coupling efficiency value and its uncertainty at temperature above 80°C and when the VCSEL array effective length x/L is larger than 2.5. Thus the material processing parameters of the bonding epoxies must be tightly controlled to reduce the resulting uncertainty in its behavior.

The Monte Carlo Simulation results are compared with the results from the approximation techniques FOSM and SOSM (Figure 2.37). It can be seen that the mean values of coupling efficiency for both FOSM and SOSM are very close with respect to the results

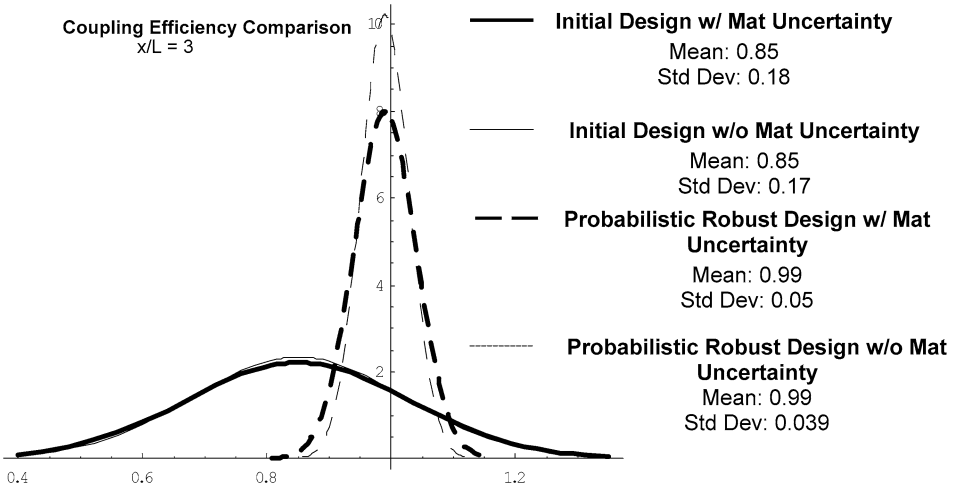


FIGURE 2.36. A comparison of uncertainty in coupling efficiency arising out of material and geometric variations.

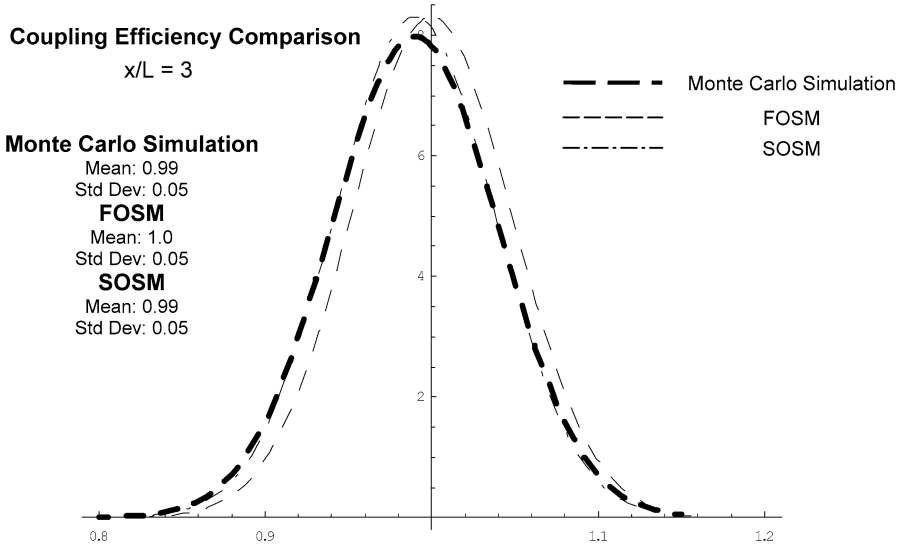


FIGURE 2.37. Comparison of Mote Carlo Simulation, FOSM and SOSM analysis for the mean and variance values of coupling efficiency resulting from material and geometric variations for single-mode operation.

from Monte Carlo Simulation. The standard deviation value for FOSM and SOSM agree reasonably with respect to the values obtained through Monte Carlo Simulation. Thus the first order and second order second moment methods can be used as reasonable approximations to determine the system uncertainty emerging from the variables affecting the system. The use of the approximation methods greatly reduces the computational expense as compared to the more expensive Monte Carlo simulation method.

The probabilistic design procedure demonstrated above reflects the physics and the variability of the behavior and performance of the photonic system much better than the

deterministic procedure. The variability and uncertainty are inherent in nature of the material behavior, manufacturing and application conditions and cannot be ignored in the design procedure when their effect on the performance is large as can be seen in Figure 2.36. The probabilistic models provide a goal oriented utilization of the experimental data in the optimizing the performance of the system. It should however be noted that the probabilistic methods though powerful are limited to the availability of the uncertainty quantification through experiments. In cases where the experimental data is not available, the deterministic robust design procedures must be used to optimize the performance of the system.

2.9. SUMMARY

The procedures described in this chapter enable deterministic and probabilistic design of fiber-optic systems. The developed techniques are demonstrated for a 12×1 fiber-optic package using a curved reflector concept. The systems approaches such as the graph partitioning procedure and system decomposition through simulated annealing are very useful in identifying the critical variables that influence the uncertainty in fiber-optic systems. The novel deterministic robust design formulation developed in this chapter minimizes the sensitivity of the objective function to changes in the design variables. The probabilistic robust design technique includes the uncertainty in design variables in determining the design with least probability of failure.

The effect of material behavior on the system performance was demonstrated in the chapter. The analytical model used to determine the displacement of the VCSEL and shear stresses on the interface are applicable for any tri-layered bonded system. The operating temperature of the VCSEL has a strong effect on the coupling loss. The displacement of the VCSEL caused by the viscoelastic behavior of the epoxy increases with operating temperature of the VCSEL and thus results in increased coupling loss. The effect of material uncertainty on the coupling efficiency is dependent on the length of the VCSEL array. The effect of geometric uncertainty on the coupling efficiency is however, dependent on the dimensions and position of the curved reflector and the laser array with respect to the optical fiber. The effect of geometric uncertainty is therefore the same for each VCSEL in the array. For the example system considered in this chapter, the geometric uncertainty initially dominates the coupling loss, however, as the length of the array increases the material uncertainty dominates the coupling loss and this effect is dependent on time due to the viscous behavior of bond epoxy. The material behavior would therefore result in increased coupling loss with time. Overall, the procedures described in this chapter enable stochastic design of photonic systems in general and fiber-optic systems in particular. Such analyses are critical in systems where extreme accuracy in assessing performance is needed or where wide uncertainty is the norm.

REFERENCES

1. T. Bierhoff, A. Wallarabenstein, A. Himmler, E. Griese, and G. Mrozynski, An approach to model wave propagation in highly multimodal optical waveguides with rough surfaces, Proceedings of Xth-International Symposium on Theoretical Electrical Engineering (ISTET), Magdeburg, Germany, 1999, pp. 515–520.
2. T. Bierhoff, A. Wallarabenstein, A. Himmler, E. Griese, and G. Mrozynski, Ray tracing technique and its verification for the analysis of highly multimode optical waveguides with rough surfaces, IEEE Transaction on Magnetics, 37(5), Part 1, pp. 3307–3310 (2001).

3. T.P. Kurzweg, J.A. Martinez, S.P. Levitan, P.J. Marchand, and D.M. Chiarulli, New models for optical MEMS, Proceedings of the SPIE—The International Society for Optical Engineering, 4198, pp. 63–74 (2001).
4. B.E.A. Saleh and M.C. Teich, Fundamentals of Photonics, Wiley-Interscience, New York, 1991.
5. P. Accordi, C.P. Basola, G.P. Bava, G. Chiaretti, and I. Montroseet, Coupling efficiency evaluation of multimode fiber devices using GRIN rod lenses, Applied Optics, 29(1), pp. 37–46 (1990).
6. K.Y. Lee and W.S. Wang, Ray optics analysis of the coupling efficiency from a gaussian beam to a rectangular multimode embedded strip waveguide, Fiber and Integrated Optics, 13(3), pp. 321–330 (1994).
7. K. Xue and Q. Zhang, A novel low-cost high performance micro pseudo-spherical lens optimized for fiber collimators, Proceedings of SPIE—The International Soceity for Optical Engineering, 4906, pp. 249–252 (2002).
8. W.H. Cheng, The optimal power coupling from GaAs lasers into spherical-ended fibers, Proceedings of the IEEE, 69(3), pp. 396–397 (1981).
9. B.V. Hunter, Algorithm determines efficiency of fiber-coupling calculation, Laser Focus World, 39(7), pp. 63–66 (2003).
10. K.J. Garcia, Calculating component coupling coefficients, Laser Focus World, (August), pp. 51–54 (2000).
11. K. Kawano, Coupling of laser light to optical fibers, Review of Laser Engineering, 22(4), pp. 226–234 (1994).
12. MATHEMATICA, Version 5.0, Copyright 1988–2005 Wolfram Research Inc.
13. G. Khanarian and H. Celanese, Optical properties of cyclic olefin copolymers, Optical Engineering, 40(6), pp. 1024–1029 (2001).
14. D. Vez, S. Hunziker, S. Eitel, U. Lott, M. Moser, R. Hoevel, H.-P. Gauggel, A. Hold, and K. Hulden, Packaged 850 nm vertical-cavity surface-emitting lasers as low-cost optical sources for transparent fiber-optic links, 33rd European Microwave Conference, 2002, pp. 611–615.
15. 1550 nm, Vertical Cavity Surface Emitting Laser for Single Mode Operation, Product Data Sheet, Vertilas GmbH, Germany, 2003.
16. Corning Incorporated (Private Communication) (2003).
17. R.T. Haftka and A. Gurdal, Elements of Structural Optimization, Kluwer Academic Publications, Dordtecht, Netherlands, 1992.
18. A. Kusiak and N. Larson, Decomposition and representation methods in mechanical design, Transactions of the ASME, Special 50th Anniversary Design Issue, 117, pp. 17–24 (1995).
19. N.F. Michelena and P.Y. Papalambros, Optimal model-based partitioning for power train system design, ASME Design Automation Conference, Boston, 1995.
20. B.W. Kernighan and S. Lin, An efficient heuristic procedure for partitioning graphs, The Bell System Technical Journal, 49, pp. 291–307 (1970).
21. S. Radhakrishnan, G. Subbarayan, L. Nguyen, and W. Mazotti, Optimization and stochastic procedures for robust design of photonic packages with applications to a generic package, Proceedings of the 53rd Electronics Components and Technology Conference, IEEE, 2003, pp. 720–726.
22. T. Bi, Y. Ni, C.M. Shen, and F.F. Wu, An on-line distributed intelligent fault section estimation system for large-scale power networks, Electrical Power Systems Research, 62, pp. 173–182 (2002).
23. C.U. Saraydar and A. Yener, Adaptive cell sectorization for CDMA systems, IEEE Journal of Selected Areas in Communications, 19(6), pp. 1041–1051 (2001).
24. C. Walshaw and M. Cross, Mesh partitioning: a multilevel balancing and refinement algorithm, Society of Industrial and Applied Mathematics, 22(1), pp. 63–80 (2000).
25. M. Ouyang, M. Toulouse, K. Thulasiraman, F. Glover, and J.S. Deogun, Multilevel cooperative search for the circuit/hypergraph partitioning problem, IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 21(6), pp. 685–693 (2002).
26. G. Subbarayan and R. Raj, A methodology for integrating materials science with system engineering, Materials and Design, 20, pp. 1–12 (1999).
27. C.M. Fiduccia and R.M. Mattheyses, A linear time heuristic for improving network partitions, Proceedings of 19th IEEE Design Automation Conference, IEEE, 1982, pp. 175–181.
28. B. Hendrickson and R. Leland, The Chaco User's Guide, Version 2.0, Technical Report SAND95-2344, Sandia National Laboratories, Albuquerque, NM, 1995.
29. M.K. Goldberg and M. Burstein, Heuristic improvement technique for bisection of VLSI networks, Proceedings IEEE International conference on Computer Design: VLSI in Computers, 1995, pp. 122–125.
30. D.S. Johnson, C.R. Aragon, L.A. McGeouch, and C. Schevon, Optimization by simulated annealing: an experimental evaluation; Part I, Graph partitioning, Operations Research, 37(6), pp. 865–892 (1989).
31. J. Sheild, Partitioning concurrent VLSI simulation programs onto a multiprocessor by simulated annealing, IEEE Proceedings—E Computers and Digital Techniques, 134(1), pp. 24–30 (1987).

32. S. Kirkpatrick, Optimization by simulated annealing: quantitative studies, Journal of Statistical Physics, 34(5-6), pp. 975–86 (1984).
33. N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller, Equation of state calculations by fast computing machines, Journal of Chemical Physics, 21(6), pp. 1087–1092 (1953).
34. S. Kirkpatrick, C.D. Gelatt, Jr., and M.P. Vecchi, Optimization by simulated annealing, Science, 220(4598), pp. 671–380 (1983).
35. V. Cerny, Thermodynamical approach to the traveling Salesman problem: an efficient simulation algorithm, Journal of Optimization Theory & Applications, 45(1), pp. 41–51 (1985).
36. E. Aarts and L. Krost, Simulated Annealing and Boltzmann Machines, A Stochastic Approach to Combinatorial Optimization and Neural Computing, John Wiley and Sons, 1989.
37. R.T. Haftka and A. Gurdal, Elements of Structural Optimization, Kluwer Academic Publications, Dordtecht, Netherlands, 1992.
38. G. Taguchi, Introduction to Quality Engineering, Krauss International Publications, White Plains, NY, 1986.
39. G. Taguchi, System of Experimental Designs, Krauss International Publications, Vol. 1 and 2, White Plains, NY, 1986.
40. J.N. Otto and E. Antonsson, Extensions to the Taguchi method of product design, Design Theory and Methodology, ASME, DE-31, pp. 21–30 (1991).
41. B. Ramakrishna and S.S. Rao, A robust optimization approach using Taguchi's loss function for solving nonlinear optimization problems, Advances in Design Automation, 32, pp. 241–248 (1991).
42. T. Chang, A.C. Ward, J. Lee, and E.H. Jacox, Distributed design with conceptual robustness, a procedure based on Taguchi's parameter design, Concurrent Product Design, ASME, DE-74, pp. 19–29 (1994).
43. S. Sunderasan, K. Ishi, and D.R. Houser, A robust optimization procedure with variaton on design variables and constraints, Engineering Optimization, 20, pp. 163–179 (1992).
44. J.J. Dennis and R. Schnabel, Numerical Methods for Unconstrained Optimization and Nonlinear Equations, Prentice-Hall Inc., Englewood Cliffs, New Jersey, 1983.
45. S. Mahadevan and A. Haldar, Reliability-based optimization using SFEM, Lecture Notes in Engineering, 61, pp. 241–250 (1991).
46. E. Suhir, Applied Probability for Engineers and Scientists, McGraw Hill, New York, 1997.
47. S.V.L. Chandu and R.V. Grandhi, General purpose procedure for reliability based structural optimization under parametric uncertainties, Advances in Engineering Software, 23(1), pp. 7–14 (1995).
48. J.O. Lee, Y.-S. Yang, and W.-S. Ruy, A comparative study on reliability-index and target-performance based probabilistic structural design optimization, Computers and Structures, 80(3-4), pp. 257–269 (2002).
49. H.A. Jensen, Reliability-based optimization of uncertain systems in structural dynamics, AIAA Journal, 40(4), pp. 731–738 (2002).
50. M. Allen, M. Raulli, K. Maute, and D.M. Frangopol, Reliability-based analysis and design optimization of electrostatically actuated MEMS, Computers and Structures, 82(13–14), pp. 1007–1020 (2004).
51. L.L. Howell, S.S. Rao, and A. Midha, Reliability-based optimal design of a bistable compliant mechanism, Proceedings of the 19th Annual ASME Design Automation Conference, Part I, 1998, pp. 441–448.
52. E. Ponslet, G. Maglaras, R.T. Haftka, E. Nikolaidis, and H.H. Cudney, Comparison of probabilistic and deterministic optimizations using genetic algorithms, Structural Optimization, 10(3-4) (1995).
53. E. Suhir, Thermal stress modeling in microelectronics and photonic structures, and the application of the probabilistic approach: review and extension, The International Journal of Microcircuits and Electronic Packaging, 23(2), pp. 215–223 (2000).
54. E. Suhir, Probabilistic approach to evaluate improvements in the reliability of chip-substrate (chip-card) assembly, IEEE CPMT Transactions, Part A, 20(1), pp. 60–63 (1997).
55. E. Suhir and B. Poborets, Solder glass attachment in cerdip/cerquad packages: thermally induced stresses and mechanical reliability, ASME Journal of Electronic Packaging, 112(3), pp. 204–209 (1990).
56. W.T. Chen and C.W. Nelson, Thermal stress in bonded joints, IBM Journal of Research and Development, 23(2), pp. 179–188 (1979).
57. E. Suhir, Stresses in bi-metal thermostats, Journal of Applied Mechanics, 53, pp. 657–660 (1986).
58. E. Suhir, Calculated thermal induced stresses in adhesively bonded and soldered assemblies, Proceedings of the ISHM International Symposium on Microelectronics, Atlanta, Georgia, 1986, pp. 383–392.
59. E. Suhir, Die attachment design and its influence on thermal stresses in the die and the attachment, Proceedings of the IEEE 37th Electronics Components Conference, 1987, pp. 508–517.
60. N. Barbes and P. Walsh, Loss of Gaussian beams through off-axis circular apertures, Applied Optics, 27(7), pp. 1230–1232 (1988).

3

The Wirebonded Interconnect: A Mainstay for Electronics

Harry K. Charles, Jr.

The Johns Hopkins University, Applied Physics Laboratory, 11100 Johns Hopkins Road, Laurel, Maryland 20723-6099, USA

3.1. INTRODUCTION

3.1.1. Integrated Circuit Revolution

3.1.1.1. Device Trends Semiconductor device technology has had an unparalleled rise in density, functionality, and complexity in its history since the invention of the bipolar transistor in 1947 [4] and the birth of the integrated circuit (IC) in 1958 [46]. In fact, IC technology has followed a path of doubling its complexity (or number of devices per single piece of silicon or chip) every 18 months to two years since its birth [57,71]. Electronic fabrication technology has the ability to put over 100 million transistors on a single piece of silicon (chip) less than 2 cm^2 in area. A billion transistors on the same size chip have already demonstrated with 10^{11} devices (transistors) predicted on a similar size semiconductor slice by 2010.

3.1.1.2. Input/Output Trends With this extremely rapid rise in chip density and functionality, the requirement for increased inputs/outputs (I/O) has also risen dramatically. Transistors required three to four interconnects and were the mainstay semiconductor product during the decade of the fifties. Early ICs required a dozen or so interconnect wires; but as the IC revolution continued, I/O requirements increased rapidly. Today, ICs routinely have several hundred I/O pads with some approaching the 1000 mark (random logic and microprocessors). A few devices even have higher I/O numbers, usually in the 1000 to 1500 range. The complex, increased functionality ICs of the future will have I/O requirements in the thousands. It should be remembered, however, that systems will still contain a wide variety of chip types ranging from memory with I/O counts less than 100 to special-purpose microprocessors and random logic with I/Os in the thousands. Thus, an effective interconnection system must be able to transcend the full range of I/O number and density requirements.

3.1.2. Interconnection Types

3.1.2.1. Overview There are three major forms of electrical interconnection for integrated circuits and related packaging applications: (1) wirebonding, (2) flip chip attachment, and (3) tape automated bonding (TAB) as shown schematically in Figures 3.1–3.4, respectively. Many other forms of interconnect exist to meet special needs or performance

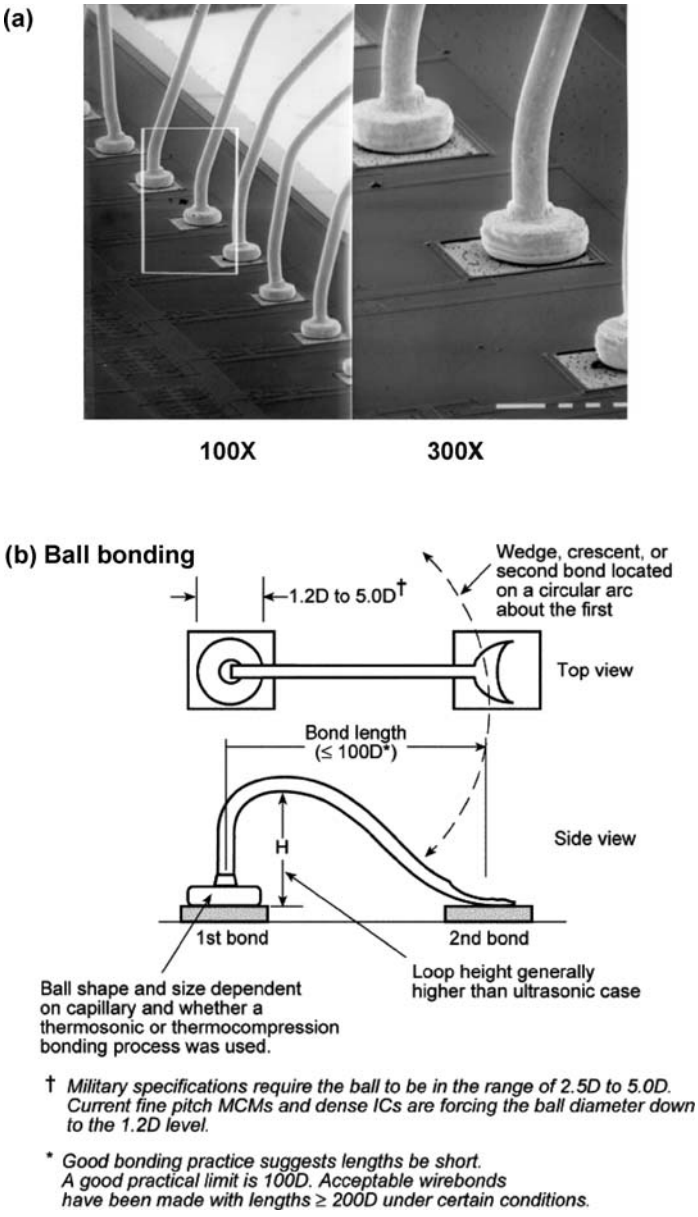


FIGURE 3.1. Ball bonds (thermocompression or thermosonic). (a) Scanning electron microscope photomicrograph of typical ball bonds; (b) schematic representation of ball bonds with important parameters indicated.

requirements. These range from completely deposited multilayer thin film interconnection schemes such as high density interconnect (HDI) [50] to interconnection techniques involving "G-shaped" springs [53], laser written conductors [25,52,79], and elastomeric pressure contacts. Detail description of these techniques is beyond the scope of this work.

Wirebonding, by far, is the most dominant form of electrical chip interconnection. Over four trillion wirebonds are made annually. This staggering number of wirebonds ac-

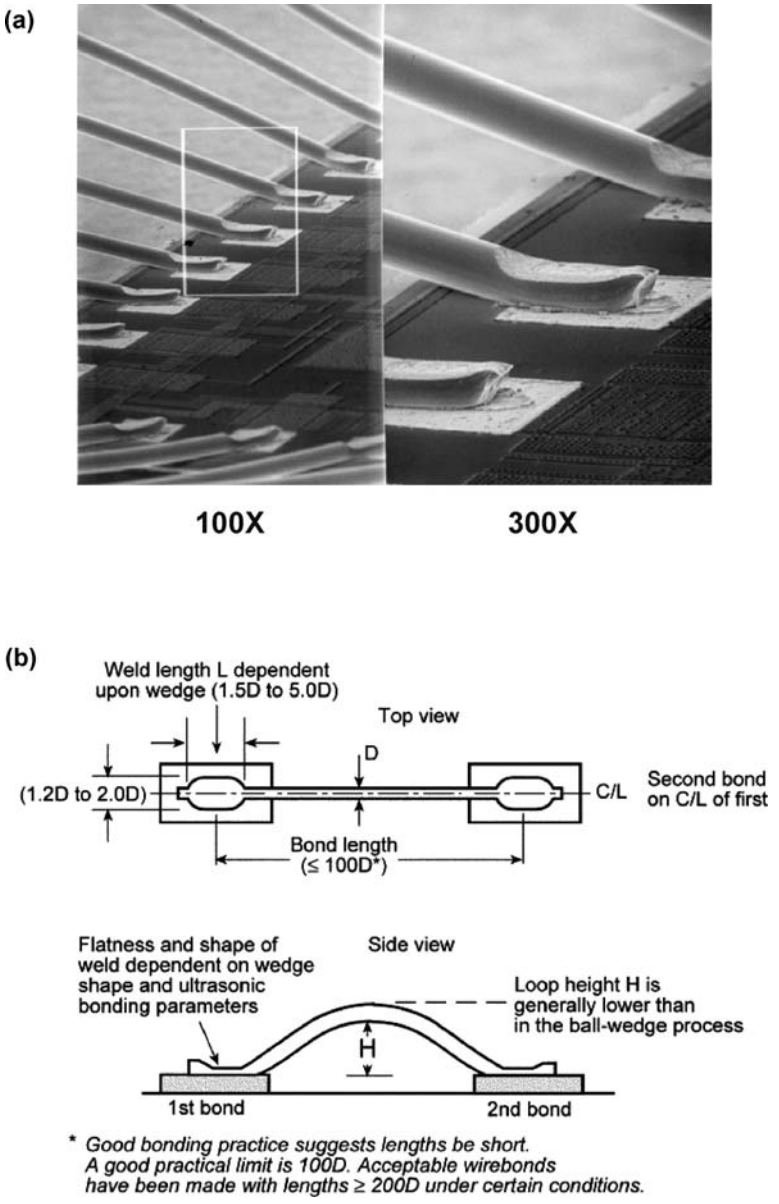


FIGURE 3.2. Ultrasonic bonds (wedge bonds). (a) Scanning electron microscope photomicrographs of typical ultrasonic wedge bonds; (b) schematic representation of ultrasonic bonds with important parameters indicated.

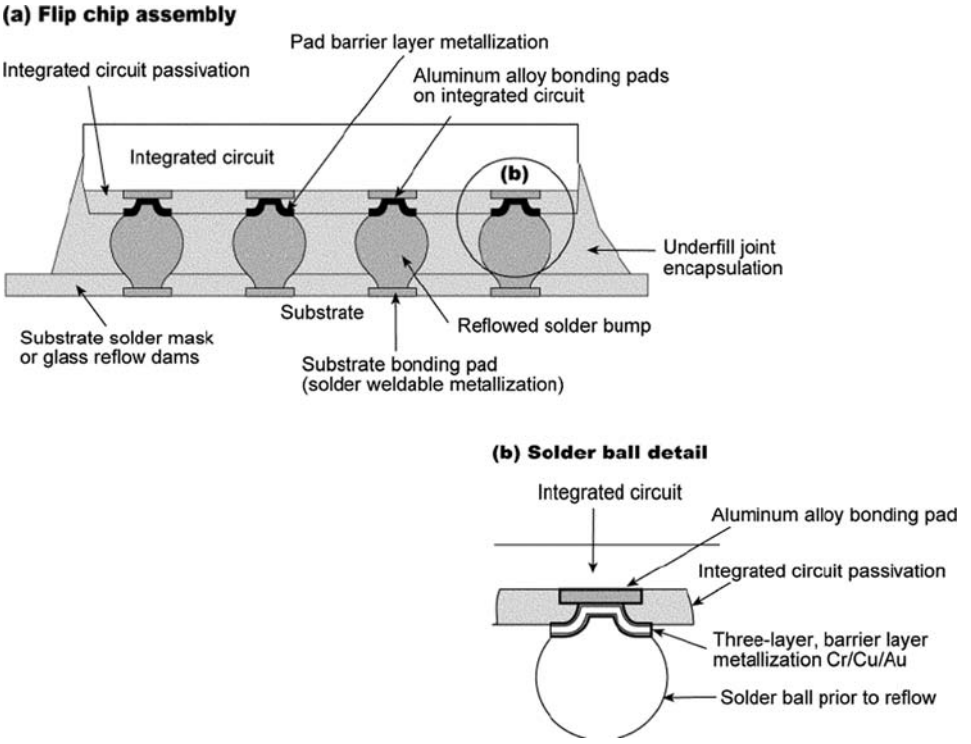


FIGURE 3.3. Schematic representation of the flip chip bonding process. (a) Cross-section of a flip chip assembly; (b) detail of the solder ball and barrier layer metalization prior to reflow.

counts for well over 90 percent of all the first-level interconnects (chip to package or chip to board) produced. The details of the formation, application, and the future of wirebonding in relationship to electronic products are the focus of this article. Prior to describing wirebonding in detail, it is necessary to put wirebonding in perspective with the other forms of chip electrical interconnection.

3.1.2.2. Flip Chip The basic flip chip process was developed by IBM in the early 1960s [56]. In their process called controlled collapse chip connection (C4 for short), solder bumps are formed on wettable chip bonding pads. A mating solder-wettable metalization pattern is created on the package or substrate. The chip of IC is placed upside down (flip chip), and all joints are formed simultaneously by reflow soldering. Figure 3.3 is a schematic representation of an IC attached by the flip chip process.

In the original C4 process, copper spheres were embedded in the solder bumps to keep the edges of unpassivated silicon chips from electrically shorting to the solder-coated substrate metalizations (usually thick films). With the rapid growth of IC technology, including effective die passivations, the copper spheres were removed, and the current flip chip or C4 processes were developed. In the current process, the solder bump is constrained from completely collapsing or flowing out over the entire bonding site by surface tension and the use of solder masks or dams. In thick film circuitry, the dams are glass, while for organic boards, the dams or masks are organic resins. The flow on the chip is constrained by a special bonding pad metallurgy that consists of a pad (circular or hexagonal in shape) of evaporated chromium, copper, and gold. The pad metallurgy is then coated by evaporation

with, for example, Sn5 (5 weight percent tin, 95 weight percent lead) or Sn10 (10 weight percent Sn, 90 weight percent Pb) to a thickness of 100–125 μm. Thickness of this plating is, of course, ultimately determined by the pad size (pitch). The high lead content tin-lead solders have excellent strength and fatigue resistance but the reflow temperatures are high (approximately 315°C). This high reflow temperature typically limits these materials to inorganic substrates (ceramics, silicon, etc.) The use of lower melting point alloys such as Sn 63 (63 weight percent Sn, 37 weight percent Pb), and those based on indium have allowed organic boards to be used. Because of the reduced fatigue resistance of the lower melting temperature solder alloys, the larger coefficient of the thermal expansion mismatches for the materials involved (i.e., silicon chip and organic substrate), and the increased size of the IC chips themselves, flip chipping on organic boards requires the use of underfills [41] to achieve the required product lifetimes under most operational scenarios.

Solder pads can be placed over active areas on the ICs, because the bonding process (solder reflow) involves little or no force that could damage sensitive structures. Thus, high density area arrays (of solder bumps) can be formed over the entire IC surface (providing a very high number of I/Os). For example, on a square chip with a side length of 10 mm and 25 μm (in diameter) solder bumps placed on 75 μm centers, an array (133 × 133) of over 17,000 interconnection sites can be formed. For lower I/O number requirements, both the ball diameter and pitch (center-to-center spacing of solder balls) can be increased, thus, improving overall ease of attachment and increased fatigue resistance. With 100 μm diameter bumps on 250 μm centers, a 40 × 40 array of solder bumps (1600 I/O) can be formed on a 10 mm square chip.

Besides providing the highest I/O density of the major interconnect types, the flip chip solder joint with its associated low inductance and low capacitance behavior is a very high performance interconnect. A typical flip chip solder joint exhibits low insertion loss even for signals at frequencies greater than 100 GHz. Although wirebonding is the most widely used and least expensive first level interconnection scheme as described below, it is still a relatively slow process (even with automatic wire bonders producing at rates of 10–15 bonds per second) when compared to the mass reflow associated with flip chipping.

3.1.2.3. Tape Automated Bonding (TAB) Tape automated bonding is also a "gang" bonding method in which bonds (on the chip, lead frame, or substrate) are formed simultaneously. Separate processes are necessary, however, for the connection of the tape to the chip (first or inner lead bond) and then for the connection of the chip and its now attached lead structure to the package or substrate (second or outer lead bond). This contrasts to the flip chip case where all bonds (first and second) are made simultaneously.

The initial TAB process involves the bonding of ICs to the tape (prefabricated metallic interconnection pattern, either freestanding or supported on an organic carrier film, usually in a format of one pattern wide by several hundred patterns long) using thermocompression bonding or solder reflow. The choice of thermocompression bonding or solder reflow depends upon the type of interconnection bump used on the chip (bumped chip) or on the tape (bumped tape). Bumping of the chip or the tape finger is necessary not only to effect the interconnect but also to prevent damage of the passivation surrounding the chip bonding pad as the bonding operation is performed. This initial or first bonding operation is called inner lead bonding (ILB). After the completion of the inner lead bonding process, the chip, which is now attached to a lead frame (single layer tape) or a lead pattern on an organic carrier (double layer or multi-layer tape), can be tested, encapsulated, and/or environmentally screened. The strip format of the tape facilitates the use of automated equipment. Subsequently, the individual pre-tested, encapsulated, and/or environmentally

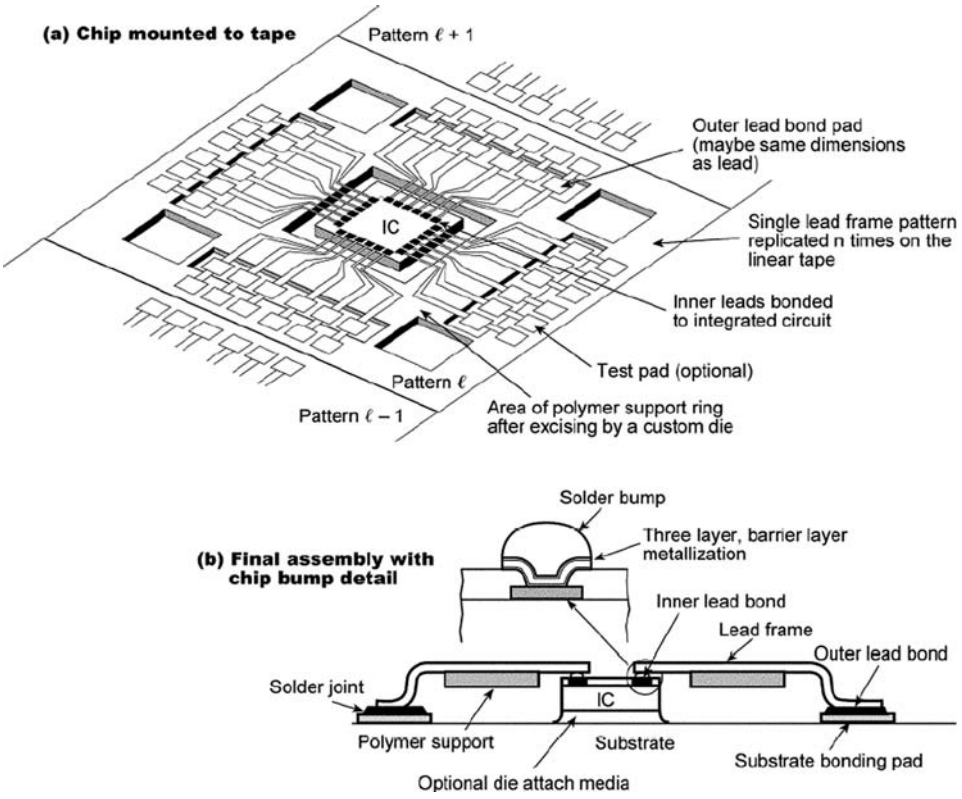


FIGURE 3.4. Schematic representation of the tape automated bonding (TAB) process. (a) Isometric view of chip mounted on carrier tape; (b) cross-sectional view of final chip to package or substrate assembly with inner lead solder bump detail.

screened parts can be excised from the tape and attached to a package, substrates, or board by a process called outer lead bonding (OLB). Basic schematic representations of TAB processes are shown in Figure 3.4.

The pre-patterned tape comes in many forms, widths, and materials depending upon circuit requirements, tape fabrication process, bonding equipment, and the metallurgy(s) involved. The tape can have either single or multiple conductor layers. The multiple conductor layers are separated by intervening dielectric layers, which are typically a form of polyimide. More detailed descriptions of tape construction and materials can be found elsewhere [12].

Bonding the chip to the tape lead frame is usually accomplished with a pre-deposited gold bump (solder bumps are also used). This interconnection bump is either placed on the chip with appropriate interface metallurgy or on the tape (as mentioned above). The bumps are needed to reach the recessed bonding sites (below the top level of the passivation layer) and minimize the TAB lead forces on the passivation surrounding the bonding pad. The interface metallurgy usually consists of several metal layers designed to provide low contact resistance, improved bump adhesion, and a hermetic seal of the pad with its surrounding passivation. Typically, these layers consist of an adhesion layer (chromium, titanium, etc.), a barrier layer (copper, nickel, platinum, or palladium), and, finally, the bump metal (gold, gold/copper, or solder-plated copper). Most mating lead frames consist of gold- or

tin-plated copper. Generic interface metallurgies for various tape systems are presented by Tummula et al. ([80], p. II-225).

The ILB process can be accomplished for gold bumps by using thermocompression bonding with a heated thermode [80]. Thermodes exist in many forms, including solid and bladed. Just as in thermocompression wirebonding, the inner lead bond strength is strongly dependent on the temperature, dwell time (length of time the thermode is in contact with the lead), and the loading (applied force) during the bonding process. The bond termination material and the cleanliness of the bond interface also have an effect. If a solder bump is used, then the ILB process is solder reflow. Either a high lead content solder (e.g., Sn5 or Sn10) or a tin-gold eutectic alloy solder (80 weight percent Sn, 20 weight percent Au) is used. In the tin-gold process, tin-plated leads are bonded to gold bumps (or vice versa). The tin-gold eutectic attach process produces low stress, uses a relatively low temperature (280°C) when compared to other ILB process, and is generally applicable to most tapes and bonding situations.

Outer lead bonding is typically performed with a heated blade-type thermode, which forces the TAB leads against the bonding pads on the package, substrate, or board. Typically, the OLB thermodes are larger than their ILB counterparts because of lead fan out and the larger OLB bonding sites. The bonding sites or pads are usually coated with a solder on solder paste, and once the heated thermode causes the solder to reflow, its temperature is reduced and the solder is allowed to solidify prior to the removal of the thermode. Other simultaneous or gang soldering techniques such as vapor phase, infrared, and hot air soldering may be used with appropriate fixturing. Thermocompression bonding between two compatible metallurgies also has been used for OLB.

Prior to OLB, it is necessary to remove any interconnection lead support bars or common plating connections used in the tape fabrication process that may cause lead shorting in the final TAB assembly. Such removal is done by punching or cutting. If further chip testing is required prior to final assembly, it is typically done at this time. In order to test the chips thoroughly it may be necessary to separate the lead frame cells with chips attached and place them in a tape carrier that is compatible with the testing apparatus. With proper design of the tape carrier, testing at speed and full function verification are possible. Once testing is complete, the chip with its TAB lead frame structure is cut from the carrier tape with a metal die. The leads are then bent to shape (e.g., full wing) to provide the proper mechanical compliance.

3.1.2.4. Interconnection Requirements Wirebonded interconnects are usually applied to perimeter bonding pads on ICs. Perimeter bonding pads are located over non-active areas of the chip, thus preventing any damage to the IC, due to forces associated with the bonding process. Using special processes coupled with precision bonding machine control, several researchers and a few manufacturers have wirebonded successfully over active regions; but this is not a widely accepted or recommended practice. Flip chip reflow soldering, on the other hand, can be used over active regions, because it exerts little or no force in the attachment process. TAB, depending upon the tape form (area or perimeter), and the type of inner lead bonding (e.g., solder reflow), can be used in either mode, although perimeter TAB is by far the most common.

Figure 3.5 illustrates current and projected I/O requirements for various types of electronic products. As can be seen, the I/O requirement range from less than 100 to almost 5000 depending upon product type and the time period considered. To gain some understanding of the implications of these large and increasing I/O numbers, let's consider how they might be supported from an interconnection point of view. Figure 3.6 plots

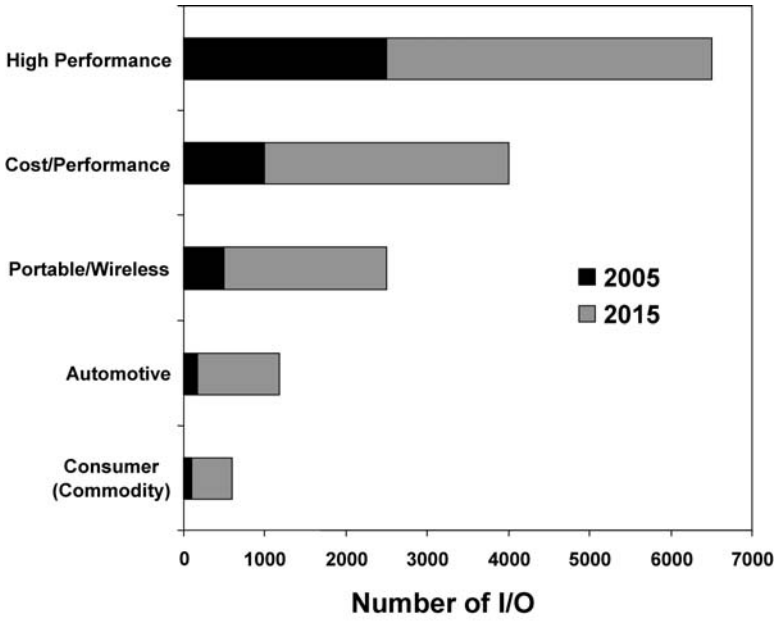


FIGURE 3.5. Maximum expected I/O for different classes of electronic products both now and ten years in the future.

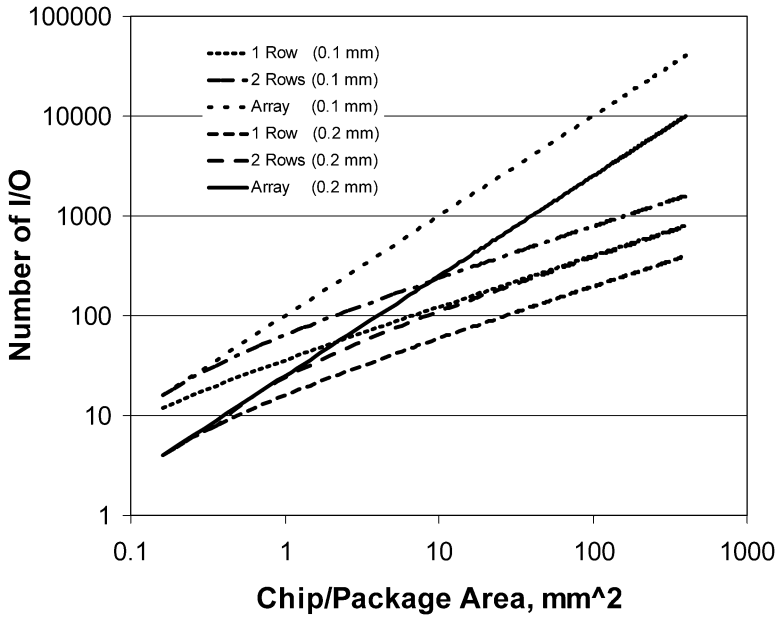


FIGURE 3.6. Number of I/O's as a function of chip package area for both perimeter (1 row and 2 rows) and area array interconnection points (bonding pads).

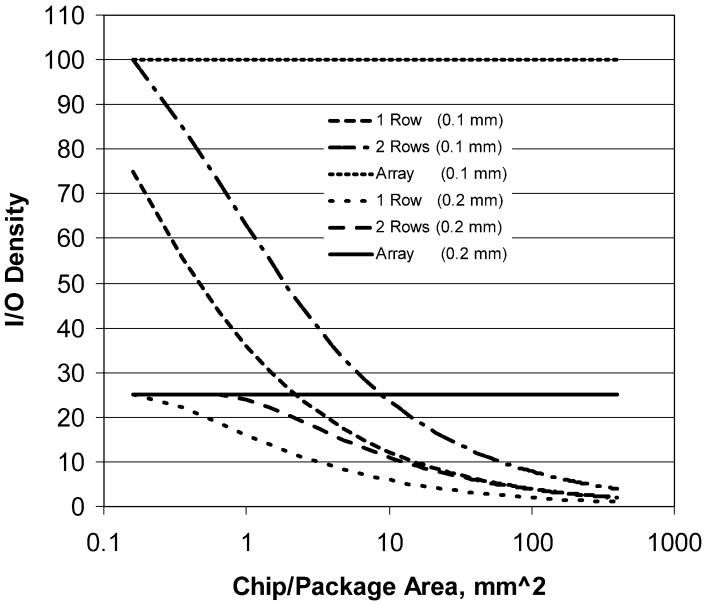


FIGURE 3.7. Package or chip I/O density (I/O per unit area) as a function of the area for both perimeter (1 row and 2 rows) and area array interconnection points (bonding pads).

the number of I/O versus chip area for the two major interconnection types: wirebonding (perimeter attachment) and flip chip (area attachment). Wirebonding, even with two rows of bonds at an extremely fine pitch (e.g., 75 μm), requires a relatively large chip size (350 mm^2) to reach 1000 I/O while a chip of that size can support almost 16,000 I/O using flip chipping. The curves for wirebonding are also applicable for predicting I/O count using perimeter TAB. As mentioned above, area TAB exists but is not widely used. Area array interconnects can easily exceed 1000 I/O even on small-size chips with a relaxed bond-to-bond spacing or pitch. It is common to normalize the I/O numbers with respect to the chip area, thus forming the I/O density (number of I/O per unit area). Figure 3.7 plots I/O density versus chip area for various pitches of the interconnect. For an area array the I/O density is constant for a given pitch regardless of the chip size, while for a perimeter bonded chip, even with double rows, the I/O density falls off exponentially with increasing chip area.

Similarly, many other IC design, process, and material parameters affect IC bondability in addition to increased active device density and rising interconnection requirements. The aluminum-silicon alloy system (Al + 1% Si), which was standard on many early integrated circuits, has been changed by adding copper (up to 4%) to prevent electromigration as the spacing between adjacent lines has decreased. The addition of copper has produced bondability problems. Research has shown [36] that copper content above 2% prevents effective wirebonding. Another manifestation of shrinking line size is that the lines are becoming much more resistive, forcing the replacement of the aluminum-silicon alloy system with a metal having higher electrical conductivity, such as copper. Copper requires trenching encapsulation with either chromium or titanium adhesion layers [29]. The rigid organic dielectric layers on the IC will be replaced by organic materials with lower dielectric constants, such as polyimide, benzocyclobutone, or Teflon®-based materials (polytetrafluoroethylene). The ultimate goal, if interconnect topologies and copper passivation

processes can be developed, would be to use air as the dielectric. Using copper as the IC metalization with soft organics as the intervening dielectric layers, present challenges to the first-level (on-chip) interconnection processes, especially wirebonding. Copper metalization pads will necessitate copper wirebonding or a suitable barrier layer metalization cap to allow bonding with gold or aluminum wires. A metal flash on the copper pad to prevent oxidation is also necessary prior to firming flip chip solder balls.

3.1.3. Wirebond Importance

As mentioned above, wirebonding is the dominant form of first-level interconnect because of its flexibility, low interconnect cost, ease of use, and relatively low capitalization costs. Wirebonding is extremely flexible. It can bond various chip types, metallurgies, pad sizes and locations, and configurations, etc., by changing machine parameters, software programs, and, perhaps, bonding tools (capillaries and wedges), bonding wire, wire size, and bond type (e.g., ball bonding to wedge bonding). Such changes are straightforward, can be performed very quickly, and usually at low cost, except in the case of the bond type that will require moving to a different bonding machine with all of its associated acquisition and set-up costs. In addition, the cost of a fully automatic wirebonding machine and a suitable wirebond testing machine along with an organic die attach system can be acquired for less than $250,000 (2005 U.S. dollars). Wirebond interconnections in volume production cost between $0.001 and $0.002 (U.S. dollars) per interconnect.

The other major interconnect types, flip chip and TAB, require major tooling and capital equipment to produce the on-chip solder bumps (flip chip) or the custom pre-patterned tape (TAB). In either technique, a minor change in chip pad geometry will require a costly photo tooling change in addition to a new acquisition cycle for mating substrates or tape. This lack of flexibility and increased complexity (additional processes) produce a per interconnect cost of $0.05–0.10 (U.S. dollars). Given that the cost of interconnect is much greater than wirebonding and the number of I/Os possible is rapidly increasing (fine pitch wirebondings), one might ask, why are the other interconnection methods considered important?

Flip chip has four major advantages: (1) it can produce the highest number of interconnections per unit area (and the interconnection density is constant); (2) all the bonds are contained within the chip area, i.e., there is no second bond or outer lead bond location beyond the chip perimeter; (3) it has extremely low inductance and capacitance per joint, thus allowing operation at very high frequency (up to 100 GHz); and (4) it has the most robust replacement process (of any of the major interconnection schemes) for preserving the under laying board, substrate, or the chip [65]. It is interesting to note that in addition to flip chip's large interconnect potential at fine pitch it can satisfy most practical I/O number requirements at a much larger pitch, making large robust solder joints possible.

While flip chip for performance and wirebonding for cost and flexibility exceed the capabilities of TAB, TAB does have one interesting advantage over the other interconnection types. In the TAB process, the IC is attached to its final lead structure (via inner lead bonding) prior to placement in a package or directly on a substrate. Such attachment allows for testing both at speed and temperature in a lead configuration that is closely representative of its final use state; thus in principle, solving the known good die (KGD) problem [5,7,62]. Die testing techniques for wirebonded interconnected chips (prior to placement at final chip location) have been developed, but they lack the utility of TAB and always

involve some form of wirebond lifting (removal) and replacement, which carries with it an inherently greater risk of good die loss. Flip chip test scenarios have also been developed and usually involve reflow to a test substrate or some form of flexible pressure type contact structure or interposer.

Given the discussion above, it is clear that wirebonding will be the dominant form of first-level interconnect (chip to package, substrate, or board) for some time to come in all applications that: (1) can afford the size of the perimeter extension (beyond the chip) required for the second bond; (2) allow wirebonding to affect the required number of I/O interconnections with perimeter bonding pads; and (3) have a frequency of operation low enough (i.e., <10 GHz) so that wire loss is manageable.

If the chip packaging requirements exceed the wirebonding restrictions listed above, then flip chip should be the interconnect of choice. Flip chip usage is increasing rapidly worldwide. It is particularly effective in applications where its special advantages of small size and high performance are required. Flip chipping is clearly the second most important first-level interconnect. Special testing needs may, under certain circumstances, require the use of TAB.

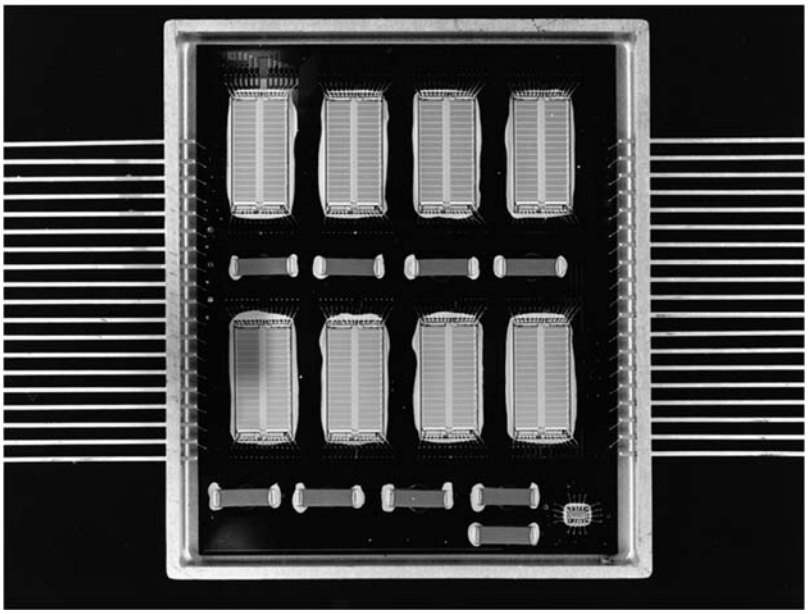
3.2. WIREBONDING BASICS

Figure 3.8 illustrates an example of a modern wirebonded circuits. The wirebonding process begins by firmly attaching the backside of the integrated circuit or wirebondable component to the appropriate substrate location or package bottom by using an organic adhesive, a low melting point glass, the reflow of a metal alloy, or a gold-silicon eutectic alloy process [30]. Once bonded in place (the process is called die or chip attach), wires are attached to the chip bonding pads using special tools (capillaries or wedges) and various combinations of heat, pressure (force), and ultrasonic energy. Depending upon tool type and choice of welding energy (direct heat or ultrasonic heating or both), three major techniques for wirebonding have emerged over the years since microelectronic wirebonding was developed in the mid-1940s to the mid-1950s timeframe [20,36]: thermocompression bonding, ultrasonic bonding, and thermosonic bonding.

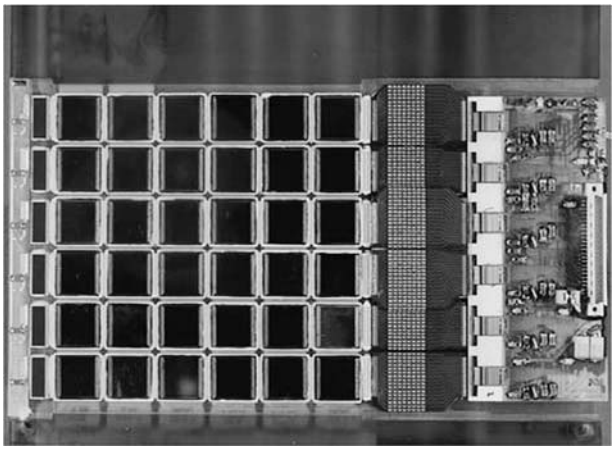
Thermocompression bonding and thermosonic bonding methods produce a ball-wedge (first bond-second bond) type bond [Figure 3.1(a)], where the wedge (tail, crescent, or second) bond lies on an arc about the first bond or ball bond as shown in Figure 3.1(b). Ultrasonic bonding produces a symmetric wedge-wedge (first bond-second bond) style bond as shown in Figure 3.2(a). In ultrasonic bonding, the second bond lies along the center line of the first [see Figure 3.2(b)].

3.2.1. Thermocompression Bonding

A thermocompression bond (or weld) is the result of bringing two metal surfaces (bonding wire and the substrate or pad metalization, for example) together in intimate contact during a controlled time, temperature, and pressure (or force) cycle. During this "bonding" cycle, the wire and, to some extent, the underlying metalization undergo plastic deformation and interdiffusion on the atomic scale. This atomic interdiffusion can result in a uniform gold welded interface, if both gold wire and gold pad or substrate metalization are used. Gold-aluminum intermetallics [63] are formed when gold wire and aluminum pads (or vice versa) are used. Regardless, the plastic deformation that occurs at the bonding interface ensures: intimate surface contact between the wire and the pad, provides an



(a)



(b)

FIGURE 3.8. Examples of wirebonded circuitry. (a) Static RAM module using MCM-D technology. Unit contains 300 gold thermosonic wirebonds. (b) Experimental X-ray detector for use in space. 36-detector chips with bond pads on both sides of the chip. Each chip has over 200 wirebonds per side. Total wirebonds on assembly exceed 18,000. Chips mounted on open frame to allow wirebonder access to both sides.

increase in the interfacial bonding area, and breaks down any interfacial film layer (oxide, contamination, etc.). Surface roughness, voids, oxides, and absorbed chemical species or moisture layers can all impede the intimate metal-to-metal contact and limit the extent and strength of the interfacial weld; thus, causing a poor bond. In some cases, this interfacial contamination (usually on the pad) is so extensive that it prevents bonding altogether.

The interfacial bonding temperatures are typically in the range of 300–400°C [43] for bonds made by thermocompression bonding. The bonding cycle, exclusive of bond po-

sitioning, takes a fraction of a second. In thermocompression bonding, the required heat for interface formation is applied by either a heated capillary (the bonding tool through which the wire feeds) or by mounting the substrate and/or package on a heated stage (column). With stage or column heat, the die and package combination must come into thermal equilibrium with the stage, which can take seconds to minutes depending upon mass. Because of the high stage or column temperatures ($>200°C$) involved in thermocompression bonding, IC or device die attachment is usually limited to the gold-silicon eutectic or certain metal alloy attaches. Also, long times on heated stages can cause reliability problems with previously placed wirebonds, such as uncontrolled intermetallic growth. Most modern thermocompression bonders use a combination of both capillary and column heat. The capillary is made of ceramic, ruby, tungsten carbide, or other refractory material. A typical ball bonding cycle is illustrated in Figure 3.9. There are five major steps in the ball bonding process: (1) ball formation (views a and b, Figure 3.9); (2) ball attachment to IC or substrate pad (first bond) (view c, Figure 3.9); (3) traverse to second bond location (view d, Figure 3.9); (4) wire attachment to package or board pad (second bond) (view e, Figure 3.9); and (5) wire separations (view f, Figure 3.9). The initial ball formation step is accomplished by cutting the wire end as it extends through the capillary with an electronic discharge. This cutting is called flame-off due to the fact that in the early days of wirebonding an open flame hydrogen (or forming gas) torch was used to cut the wire. Once cut, the ends of the wire ball up due to surface tension and capillary action. Figure 3.10 illustrates free air balls produced with gold wire by a negative electronic flame-off system. Heat, time, and pressure or force are the major determining factors in the formation of thermocompression bonds. Typically, the forces used in thermocompression bonding are higher than in other ball bonding methods (i.e., thermosonic ball bonding), resulting in a much more flattened ball. Thus, the first bond is "nail head" shaped rather than just a slightly flattened ball as obtained with standard pitch thermosonic ball bonding [e.g., see Figure 3.1(a)].

Gold wire is used in most thermocompression wirebonding processes because it is easily deformed under pressure at elevated temperature and very resistant to oxide growth that can inhibit proper ball formation. Aluminum wire, because of its rapid oxide growth, has difficulty in forming properly shaped balls on standard bonding machines. Successful aluminum wire ball bonds have been formed using an inert atmosphere around the bonding head to minimize oxide formation [28,60]. Copper and other materials (e.g., palladium and platinum) have also been ball bonded [48] in both thermocompression and thermosonic applications. Also, wedge style thermocompression bonding with many different materials has been performed [6,51]. Wedge style thermocompression bonding forms the basis of the thermode ILB attachment used in TAB.

3.2.2. Ultrasonic Bonding

Ultrasonic bonding (or wedge bonding) is a lower-temperature process in which the source of energy for the metal welding is ultrasonic energy produced by a transducer vibrating the bonding tool (wedge) in the frequency range of 20 to 300 kHz. The most common frequency is 60 kHz ([36], pp. 23–26), although higher frequency ultrasonics are in use or being considered for difficult bonding situations. Thermosonic bonding at higher frequencies will be discussed in Section 3.4.4 below. The ultrasonic wedge bonding process is illustrated in Figure 3.11. In ultrasonic bonding, the wedge tip vibrates parallel to the bonding pad. Ultrasonic bonds are typically formed with aluminum or aluminum alloy wire on either aluminum or gold pads. Gold wire ultrasonic bonding has been performed

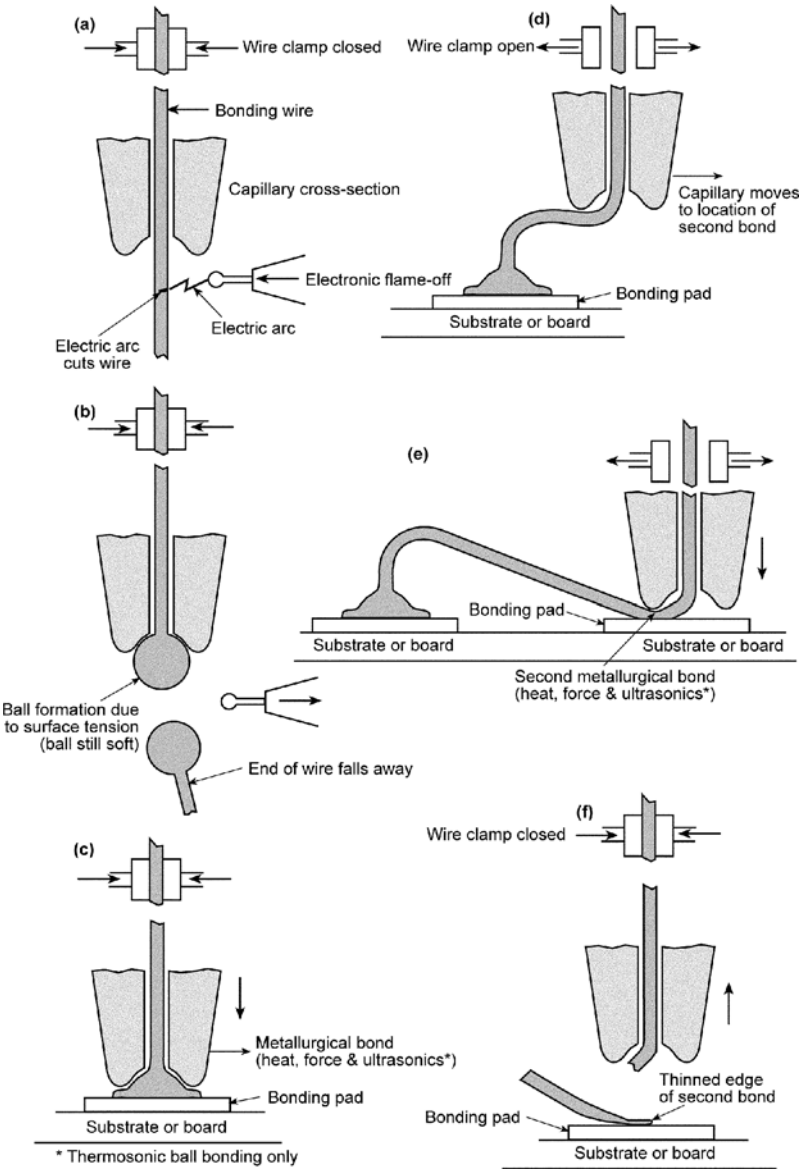


FIGURE 3.9. Schematic representation of the ball bonding cycle: (a) flame-off; (b) ball formation; (c) first bond; (d) transition to second bond; (e) second bond; and (f) separation of wire after second bond.

with both round wire and flat ribbon, although it is not widely used because of cost. Gold ribbon, because of its rectangular cross-section, provides a lower inductance interconnect (compared to a round wire of equivalent cross-sectional area) useful in radio frequency and microwave chip interconnect. In special applications, copper and palladium have been bonded by the ultrasonic process ([30], pp. 409–410). The major advantages of ultrasonic bonding include the ability to effect strong bonds with little or no applied substrate heat (implying the use of low temperature die attachment methods); and it typically can be per-

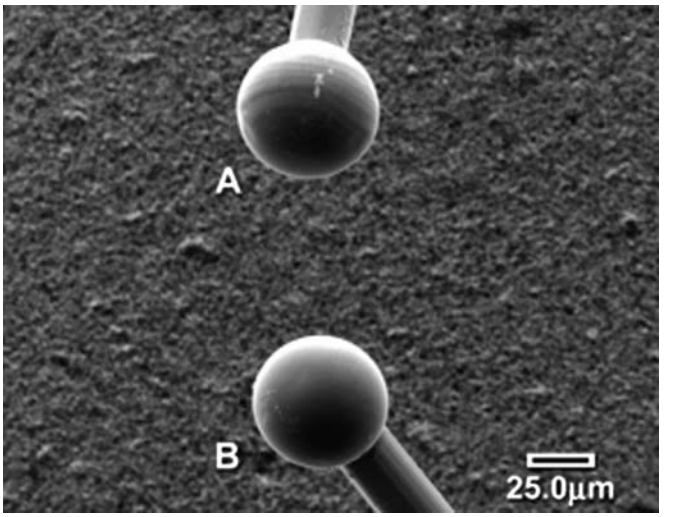


FIGURE 3.10. Scanning electron photomicrograph of free air balls produced by negative electronic flame-off. The gold wire diameter is 25.4 µm. (A) Free air ball made on 100 kHz bonder (62.2 µm diameter); (B) free air ball made on 60 kHz bonder (59.7 µm diameter). Magnification approximately 350×.

formed at finer pitches (because of the elongated, narrow shape of the bond compared to the round ball diameter) than ball bonding methods. Automated wedge or ultrasonic bonders are typically slower than ball bonders due to the requirement that the second bond must be in line with the first bond; i.e., follow the centerline of the wedge. Thus, either the entire package (substrate) or the bonding head must be rotated to bond in different directions. This slows down the bonding process when compared to ball bonding, which can place the second bond anywhere on a circle surrounding the first bond with only transversal movement of the head (or stage). [See Figure 3.1(b).]

3.2.3. Thermosonic Bonding

In thermosonic wirebonding, ultrasonic energy is combined with the ball bonding capillary technique employed in thermocompression bonding. Typically, the thermosonic bonding process is performed in a manner analogous to the thermocompression bonding process, except the capillary is not heated (or held at a lower temperature when compared to the capillary temperature in thermocompression bonding); and the stage or column temperatures are typically 150°C or less. To generate the required interfacial heat for welding at the interface of the wire and the pad, short bursts (tens of milliseconds) of ultrasonic energy are applied to the capillary when the wire and the pad are in contact. Because of the addition of ultrasonic energy (causing localized heat generation at the wire–pad interface), the requirements on stage and capillary heat (as mentioned above) and pressure (force) can be relaxed. The applied forces in thermosonic bonding are typically much less than those encountered in thermocompression bonding, thus allowing bonding over delicate or force sensitive chip or substrate regions. Since interconnections are made with the ICs (and substrates) held at temperatures of 150°C or less, they can be attached with epoxy or other organic adhesives without fear of degradation (i.e., prolonged exposures at temperatures above their glass transition temperature) due to excessive bonder stage or column temperature. Because the temperatures are lower, there is also significantly less risk of uncontrolled

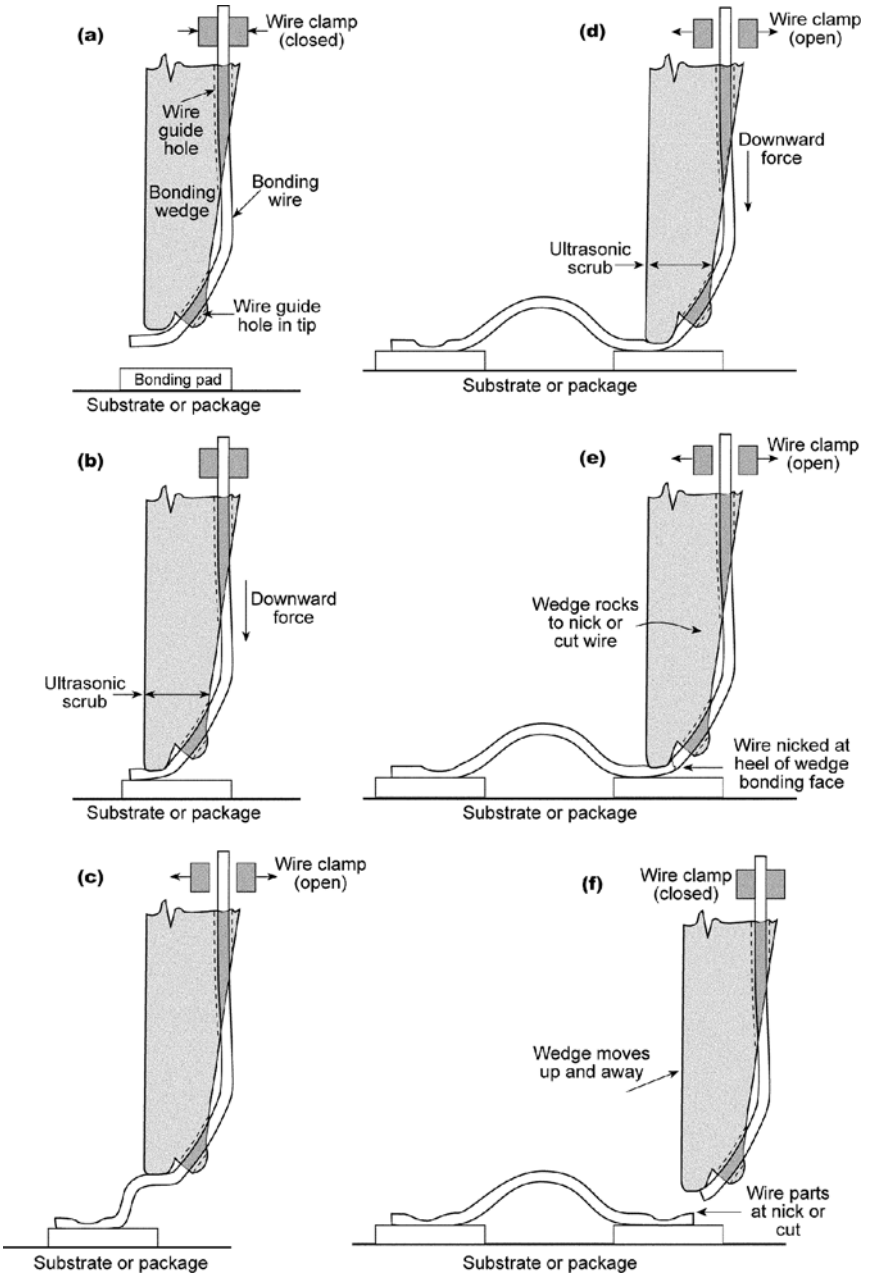


FIGURE 3.11. Schematic representation of the ultrasonic (wedge) bonding cycle: (a) initial wire-wedge configuration; (b) first bond; (c) transition to second bond; (d) second bond; (e) wire nicking or cutting operation; and (f) wire separation after second bond.

intermetallic growth. Thermosonic wirebonding is conducted primarily with gold wire, but aluminum [60], copper [48], and palladium [6] wires have been bonded successfully by the thermosonic process. As the metalization on high performance ICs migrates from alu-

minum alloys to copper [29], new pad stack configurations (e.g., copper-nickel-gold or, perhaps, just copper) will emerge. These new pad stacks will require the reevaluation of the thermosonic bonding process and, perhaps, the full consideration of the use of copper wire. Copper thermosonic wirebonding has been successfully used in the connection of the ICs to copper alloy lead frames in dual-in-line packages [39].

Such thermosonic bonding evaluations are already underway for the new substrate and pad structures encountered in the development of multichip modules (MCMs) [14]. Some of the results from these new structures and material evaluations are discussed in Section 3.4.2 below.

3.2.4. Wirebond Reliability

Wirebonding has evolved into an extremely reliable first-level microelectronic interconnection technique due to the introduction of fully automated wirebonding in the late 1970s. Coupled with the precision control of automatic wirebonders, such things as improved bonding, pad metallurgy, controlled bonding wire impurity content, effective pad cleaning processes, high purity and stable die attach adhesives, and reduced temperature bonding processes (ultrasonic and thermosonic) have all contributed to the widespread use and reliability of wirebonds. In fact, wirebond defect rates for chips bonded in single chip packages have reached the low parts per million level. Despite these improvements and high wire yields (defects less than 30 parts per million and in several instances as low as 3 parts per million (6σ) [34]), many problems still can occur in wirebonded systems.

These problems can include: mechanical fatigue due to conditions of thermal or power cycling; interactions both chemical and mechanical with encapsulation materials during molding and after cure; corrosion induced by die attach media, the atmosphere, and other process-related conditions; and wire structural changes due to the bonding parameters, such as the uncontrolled grain growth associated with the heat-affected zone. The most widely studied and publicized wirebond reliability probability is associated with the alloying reactions that occur at the gold wire–aluminum alloy bonding pad interface (and, to a much lesser degree, aluminum wire-gold bonding pad interface). Aluminum-gold intermetallic formation occurs naturally during the bonding process and contributes significantly to the integrity of the gold–aluminum interface. Intermetallics (in particular, $AuAl_2$ or purple plague and Au_5Al_2 or white plague) are generally brittle; and, under conditions of vibration or flexing (either mechanically or thermally induced due to coefficient of thermal expansion mismatches), may break due to metal fatigue or stress cracking, resulting in bond failure [63].

At elevated temperatures, aluminum rapidly diffuses into the gold forming the $AuAl_2$ phase, leaving behind Kirkendall voids [63] at the aluminum–$AuAl_2$ interface. Figure 3.12 shows views of extensive intermetallic growth around and under various thermosonic wirebonds (both ball and tail bonds). Kirkendall voiding has also been observed at gold–Au_5Al_2 interfaces. Excessive intermetallic growth can lead to the coalescence of voids, which can lead to a bond crack or lift and an open circuit. Impurities in the bonding wire, on the pad metalization, or at the wirebond–pad interface have been shown to cause rapid intermetallic growth and Kirkendall voiding at temperatures below those associated with normal intermetallic formation [8]. Table 3.1 gives the formation temperature, activation energies, and some notes for the five aluminum-gold intermetallics. The deleterious effects of intermetallics can be controlled if the time of exposure to high temperature is minimized and if proper materials and cleaning procedures are used [82]. Design rules have been developed

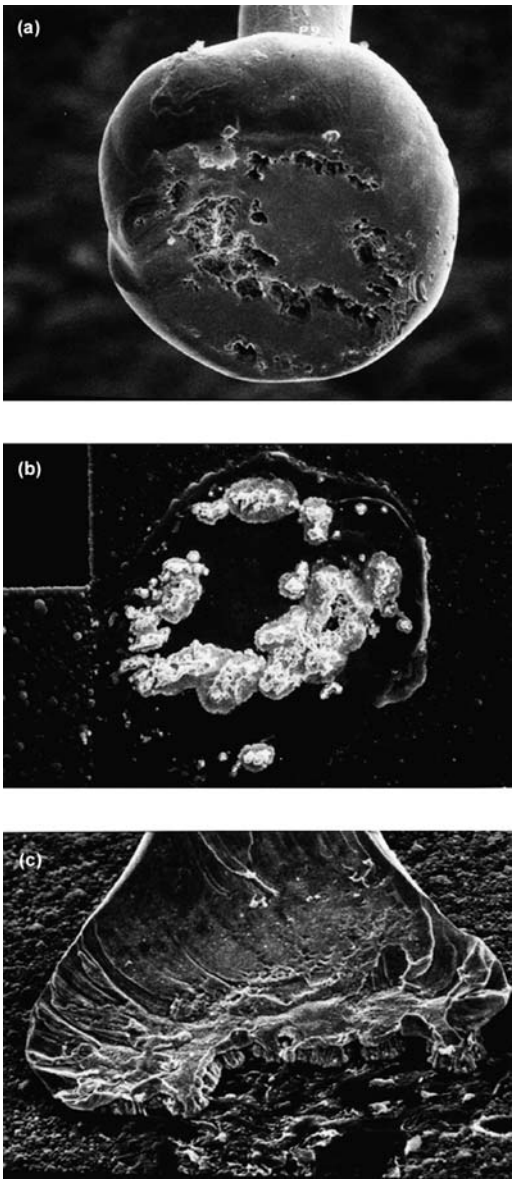


FIGURE 3.12. Scanning electron photomicrographs of advanced intermetallic growth: (a) underside of ball bond with regions of intermetallic voiding (Kirkendall); (b) residual intermetallic left on bonding pad corresponding to the voided regions of the ball in view (a); and (c) tail bond with extensive intermetallic formation under the bond edge and consuming part of the flattened bond region. Magnification approximately 75×.

for minimizing intermetallic void failures by controlling film layer composition and thickness [22]. In addition, proper optimization of the wirebonding process has a significant influence on intermetallic growth.

TABLE 3.1.
Aluminum-gold intermetallic alloy properties.

Alloy[a]	Formation temperature, °C	Activation energy[b]			Comments
		eV	kJ/mol^{-1}	kcal/mol^{-1}	
Au_5Al_2	23−100	0.62	59.4	14.3	Tan in color
Au_2Al	50−80	1.02	98.3	23.5	Metallic gray in color (orthorhombic, randomly oriented monocrystals)
$AuAl_2$	150	1.20	115.8	27.7	Deep purple in color (purple plague-resistivity 8 μΩ cm)
Au_4Al	∼150				Tan in color
AuAl	∼250				White in color

[a] The intermetallic alloys typically form in the order listed (Au_5Al_2, ... , AuAl) consistent with their temperature of formation.
[b] A range of activation energies from 0.2 eV to 1.2 eV, have been observed for the aluminum-gold system depending upon growth, testing, and contamination conditions.

3.2.5. Wirebond Testing

Since its introduction in the 1970s, the destructive wirebond pull test [37] is the most widely accepted technique for the evaluation and control of both wirebond quality and the associated setup of bonding machine parameters. Despite its widespread use, due to low cost and ease of use, the destructive wirebond pull test has some significant disadvantages. First, since it is destructive, it can only provide information on a lot sample basis for production product. It can be used for pre- and post-lot qualifiers to help setup the bonding machine and, of course, as a post mortem diagnostic tool in failure analysis or as part of routine destructive physical analysis. Thus, it does not provide a measure of quality for each bond. Second, in fine pitch wirebonded circuitry, it is difficult to insert the hook between adjacent wires without touching bonds (wires) other than the one of interest. Third, the destructive pull test provides very little information on the strength or overall quality of the bond interfaces as long as the chief failure mode is a wire break.

Only in the case of catastrophic interface failure, such as those encountered with impurity-driven intermetallic growth [8], will the destructive wirebond pull test yield information other than the relative breaking strength of the wire assuming appropriate correction is made for both the wire and test geometries [9]. This phenomenon is especially true in standard ball bonding situations where the ball of relatively large diameter (nominally 2.5–5.0 times the wire diameter) forms an effective bonding pad attachment that is many times stronger than the breaking strength of the wire. Although usually much stronger than the nominal wire breaking strength, except in the case of very fine pitch ball bonding where the diameter of the ball is approaching that of the wire (e.g., $1.2D$ where D is the diameter of the wire) ([36], pp. 255–260), the strength of the ball-to-bonding pad attachment can vary significantly owing to the influence of bonding machining parameters, composition of the interfacial material, and environmental stresses.

These factors have led to the development of two complementary tests: (1) the 100% nondestructive pull test (NDPT) [1], and (2) the ball shear test [2]. The 100% NDPT provides a degree of confidence that each bond is strong (at least to the nondestructive preset force limit [1]. The ball shear test can be used to investigate not only the interface between the wirebond ball and the bonding pad, but also the influence of both pre- and post-bonding factors. Table 3.2 summarizes the areas of application for both the wirebond pull test and the ball bond shear test. A careful review of Table 3.2 illustrates the complementary nature

TABLE 3.2.
A comparison of areas of applicability between the wirebond pull test (ASTM Standard Test Method F458-84), and the ball bond shear test (ASTM Standard Test Method F1269-89).

Area of applicability	Wirebond pull test	Ball bond shear test
Module geometry	Yes	No
Wirebond geometry	Yes	No
Wire quality, defects, etc.	Yes	No[a]
Second bond	Yes[b]	No
Bonding machine set-up, optimization, etc.	No[c]	Yes
Process development	No[c]	Yes
Substrate, bonding pad quality	No[c]	Yes

[a] Sensitivity to contamination, insensitive to mechanical defects.
[b] Extremely dependent on geometry.
[c] Insensitive unless the effect is catastrophic.

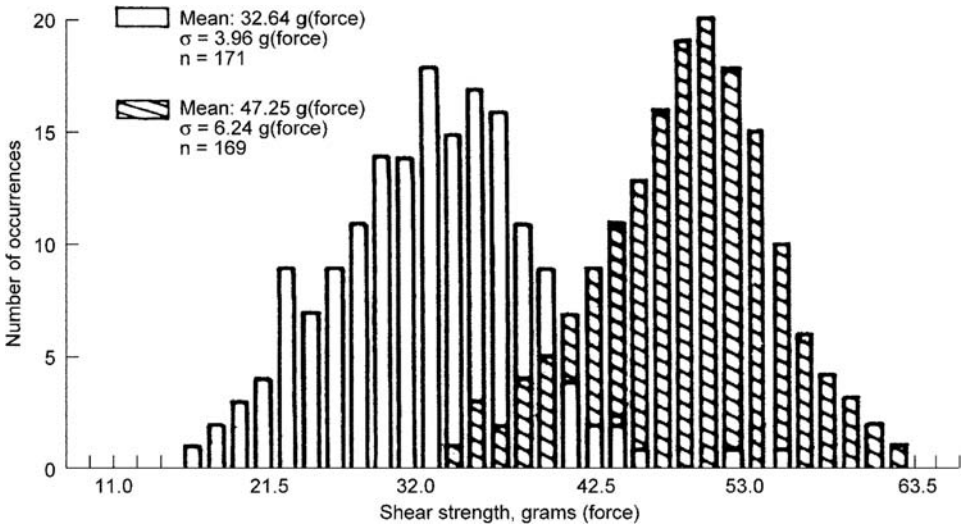


FIGURE 3.13. Histograms of gold thermosonic ball bond shear strengths for bonds placed on aluminum metallization (over silicon). Histogram A (open bars) are the shear test results after the bonding machine was set up using the wirebond pull test ($n = 171$, $\mu = 32.64$ grams (force), $\sigma = 6.24$ grams (force)), Histogram B (shaded bars) are the shear test results after the bonding machine was optimized using the ball shear test ($n = 169$, $\mu = 47.25$ grams (force), $\sigma = 3.96$ grams (force)).

of the destructive wirebond pull test and the ball-bond shear test. Figure 3.13 illustrates the improvement that can be achieved in the strength of the interface between the wirebond ball and the bonding pad by using the ball shear test (instead of the wirebond pull test) to optimize the bonding machine parameters [10].

As mentioned above the most common gauge of wirebond quality has been mechanical testing, i.e., the wirebond pull test and the ball bond shear test. Improvements in wirebond technology have caused both tests to have limitations. The pull test requires a book to be placed under a wire, which is very difficult in situations where the wires are closely spaced without damaging adjacent wires. There is also the difficulty of applying a

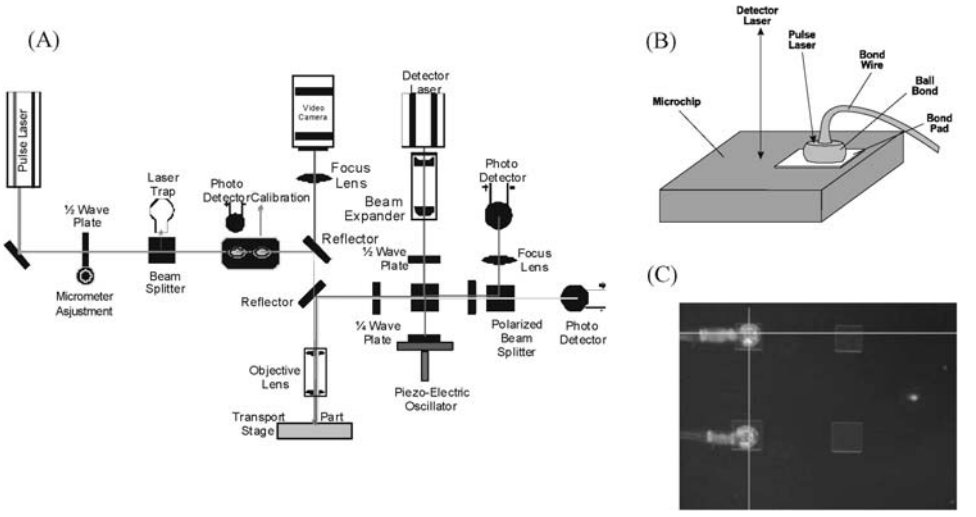


FIGURE 3.14. Laser-induced ultrasonic energy wirebond evaluation system. (A) Optical system schematic; (B) schematic representation of placement of the excitation and detection laser beams relative to the wirebond; (C) a photomicrograph showing the location of the excitation laser (cross-hairs on top of ball bond) and the detection laser (white dot on right).

consistent force to the bond interface, since the tensile and shear forces on the bond vary with the wire length and the hook position along the wire [36]. The ball shear test requires that a ram (wedge-shaped tool with a flat or slightly curved face) be placed on the major diameter of the ball. If the ball is low profile or flat such as those encountered in fine pitch wirebonding (Section 3.4.1), or thermocompression wirebonding, the ram can easily ride up over the ball. With closely spaced bonds (50–60 μm or less separation) the ram can run into adjacent bonds causing damage. Mechanical testing also tends to be time consuming and more importantly destructive. Even in non-destructive modes (see above) wires are deformed and ball edges flatten, thus giving rise to concerns about future product reliability. Hence most people recommend the mechanical testing of product on a lot sample basis only and, of course for the set-up of wirebonding machines.

A new method for wirebond testing is being developed to address the mechanical test limitations [70]. The technique uses a laser to generate an ultrasonic pulse which is passed through the bond interface and detected nearby. The test is non-destructive, fast, and appears to detect bond interface anomalies. The ultrasonic wave train is thermoelastically generated by a sub-nanosecond laser pulse hitting the top of the ball or wedge bond. It next travels through the ball or wedge and the bond interface is then detected on the surface of the integrated circuit by a laser interferometer that measures changes in the surface height. This surface displacement versus time data is then numerically converted to power versus frequency data, or Power Spectral Density (PSD). The laser ultrasonic bond testing has several potential advantages over the standard mechanical tests: (1) it is non-contact and (2) it is non-destructive. All devices produced can be tested, so quality data does not have to be inferred from a lot sample. In addition, the equipment is controlled by computer so the potential exists to fully implement the test for high production rates when attached to a wirebonder for real-time bond assessment.

A schematic representation of the test configuration is shown in Figure 3.14. Figure 3.15 presents displacement versus time curves recorded by the interferometric detection

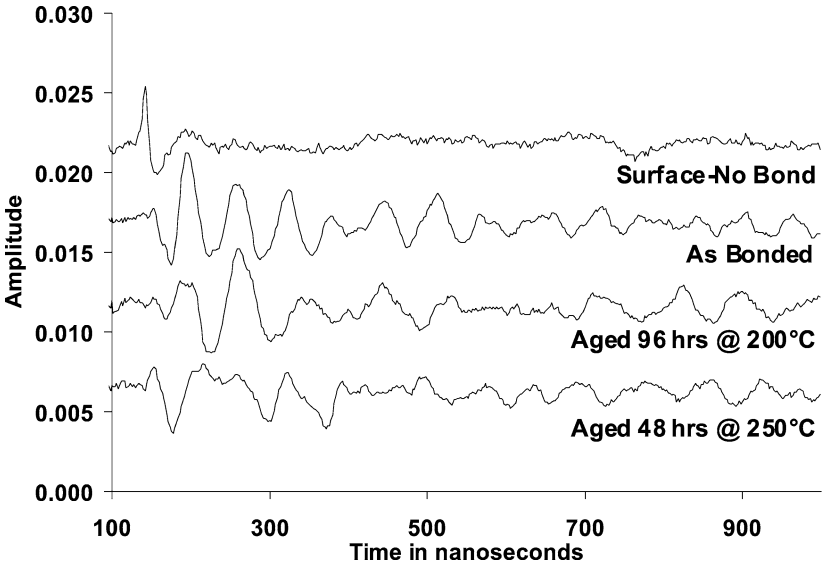


FIGURE 3.15. Displacement amplitude vs. time for bonds with different aging conditions. "No bond" illustrates noise level after bond pad surface is pulsed with a laser. Traces represent averages of at least seven individual trials and have been offset in amplitude for clarity.

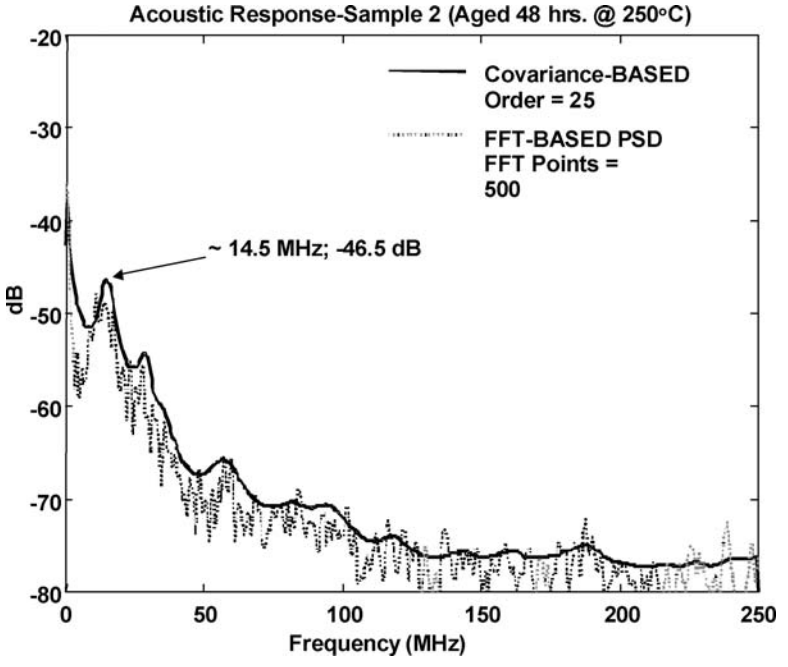


FIGURE 3.16. Comparison of the power spectral density (PSD) resulting from the fast Fourier transform (FFT) and the auto regressive (covariance based) numerical methods.

TABLE 3.3.
The effect of thermal aging on the power spectral density (PSD) behavior of typical thermosonic ball bonds made to various metalizations on silicon.

Sample	Frequency, MHz	Power, dB	Shear strength,[a] grams (force)
Sample 1: Al-1%Si			
Substrate	18.5	−56.0	
As bonded	16.5	−39.0	51.7 ± 1.8
Aged: 96 hrs @ 200°C	13.5	−44.5	60.7 ± 2.6
Sample 2: Al-1% + 0.5%Cu			
Substrate	19.5	−57.0	
As bonded	16.5	−45.5	54.3 ± 2.5
Aged: 48 hrs @ 250°C	14.5	−46.5	57.6 ± 2.2

[a] Shear strength obtained from other samples in the same sample population.

system. The numerical analysis results for a representative sample (bond aged 48 hours at 250°C) are shown in Figure 3.16. The dotted spectrum is the result of applying standard Fast Fourier Transform analysis methods to the displacement versus time curve to extract the PSD. Further analysis using an autoregressive covariance-based technique produced the sold line shown in Figure 3.16. The covariance method clearly shows a resonance response at 14.5 MHz. Applying this method to the other samples produced the data shown in Table 3.3. Table 3.3 presents the fundamental peak frequency and power levels for the aged samples along with shear strength data from bonds of the same population. Details of these results along with complete description of the method can be found in the papers by Romenesko et al. [70].

The laser ultrasonic bond evaluation has correlated a shift in the ultrasonic frequency spectrum with both bond aging and intermetallic growth. The ultrasonic wave detected was shown to be a true surface wave and thus, non-dispersive in nature. Results proving the ultrasonic wave is a surface wave are given in Figure 3.17. This means that the detected frequency shifts cannot be attributable to spectral changes due to dispersion as the detection point is moved farther away from the bond pad. In addition, no significant directional dependence of the spectrum was found—again indicating that the measurements are insensitive to the detector location relative the crystal axes of the semiconductor.

3.2.6. Bonding Automation and Optimization

Originally, wirebonding was done manually where the operator controlled every step of the bonding operation from flame-off to wire clamping and breaking (on large diameter wire even manual cutters were used). In manual bonding, operator skill was paramount to the fabrication of high-quality, reliable wire bonds. Even as the technology evolved and semi-automatic wirebonders appeared (flame-off and bonding cycle under machine control, but positioning or bond alignment was left to the operator), operator skill was key to producing highly successful (reliable) wirebonds [34]. Today, fully automated wirebonders dominate the scene. Both automatic thermosonic and ultrasonic wirebonders are in widespread use. Automatic wirebonders use pattern recognition to locate the bonding pads on both the chip and the package or substrate; and then, under complete computer control, the machines automatically bond all connections at rates exceeding 15 wirebonds (30 welds) per second. Position accuracies at those bonding rates are typically ± 2.5 to ± 3 μm. Us-

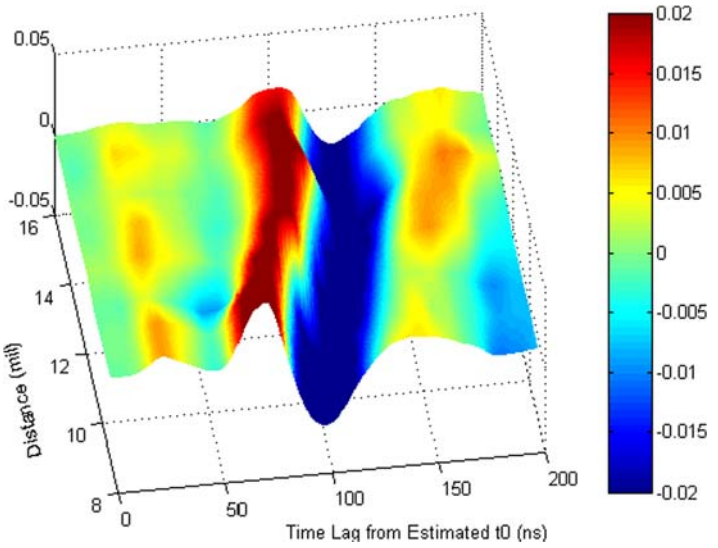


FIGURE 3.17. Results of arrival time measurement with distance. Waveforms are arranged on edge and spaced by the distance to the detector, showing arrival time to be linear with distance. Vertical axis is displacement amplitude.

ing automatic component handlers, automatic bonding machines can sustain such rates for hours. Such automation, with its concomitant accuracy and improved process control has dropped wirebond failure rates for individually packaged parts (single chip packages) into the low part per million range [34].

Failure rates associated with multichip modules, chip-on-board (or COB) and chip-on-flex are significantly higher as a result of the complex structures and new materials present in these advanced packaging structures. Some of the bonding issues for these complex circuits and structures are described in Section 3.4.2 below.

Bonding machine optimization can be accomplished in several ways depending upon the availability of test samples and trained personnel. The most straight forward way is to do a fractional factorial experimental design [10] which minimizes trials and eliminates inherent operator bias. Typically, the machine set up parameters of interest include the ultrasonic energy (P), the substrate temperature (T), and the duration of the ultrasonic energy or dwell time (D). The bonding force is usually not considered (once an initial set up has been done) since it is typically held constant for a given substrate, hybrid, or module configuration. The force is usually set to a level that promotes long capillary lifetime, thus eliminating the need to change capillaries during an experimental set (which helps minimize bond variations and improves reproducibility). For the three variables mentioned above, the bonding parameter experiments would involve a simple 2^3 factorial design with each of the variables in turn being set to expected low (-1) and high ($+1$) range limits as shown in Table 3.4. The experimental design can be unreplicated provided sufficient number of samples (>35) exist for each treatment. Random execution order should be established for all the experimental treatments to eliminate any potential memory effects. The responses denoted as S_i can be the mean shear strengths for first bond analysis (recommended) or the wirebond pull strength for each treatment. The second and third order effects are also shown in Table 3.4. The calculation of any one of these effects is simply the

TABLE 3.4.
2^3 factorial experimental design (unreplicated).

P^a	T^b	D^c	$P \times T$	$P \times D$	$T \times D$	$P \times T \times D$	Response^d
−1	−1	−1	+1	+1	+1	−1	S_1
−1	−1	+1	+1	−1	−1	+1	S_2
−1	+1	−1	−1	+1	−1	+1	S_3
−1	+1	+1	−1	−1	+1	−1	S_4
+1	−1	−1	−1	−1	+1	+1	S_5
+1	−1	+1	−1	+1	−1	−1	S_6
+1	+1	−1	+1	−1	−1	−1	S_7
+1	+1	+1	+1	+1	+1	+1	S_8

[a] P = bond power (e.g., first bond power setting), where −1 represents the low power value and +1 represents the high power value.
[b] T = temperature (substrate), °C. Again, −1 represents the low temperature setting and +1 represents the high temperature value.
[c] D = dwell time, ms. As above, −1 represents the shortest dwell time and +1 the longest.
[d] S_i = response function, typically the shear strength.

sum of the products for each level with the corresponding response all divided by $2^{(n-1)}$ where $n = 3$. For example, the effect of bond power is

$$P = (-S_1 - S_2 - S_3 - S_4 + S_5 + S_6 + S_7 + S_8)/4.$$

In order to determine the statistical significance of a particular effect with an unreplicated experimental design, an estimate of the sample variance is needed. A method for estimating the variance and confidence intervals at various significance levels has been described previously [18].

Using the same 2^3 factorial design concept with replicated center points, a linear model for ball bond shear strength in terms of P, T, and D can be constructed. The resultant ball shear equation simplifies the understanding of how the bonding parameters influence bond strength without the need for complex three dimensional plots, although with widespread availability of high performance computers, even on the shop floor, three dimensional contour plots may be preferred. In addition, the linear factorial design provides an efficient means for generating new models should different substrates and substrate metalization be required.

3.3. MATERIALS

3.3.1. Bonding Wire

Microelectronic bonding wire comes in a variety of pure and alloy materials. In addition to round wire, flat-ribbon material is available in some materials for special applications such as radio frequency and microwave circuits. Round wire is by far the most common, and fine round wires with diameters as small as 5 μm are produced commercially. Large diameter round wires up to 500 μm in diameter are used for power applications. Ribbons range from 50 μm to 1200 μm in width and come in various thicknesses.

The major materials used for these wires (and ribbons) are gold (pure and alloys), aluminum (pure), aluminum with 1% silicon, aluminum with magnesium, and, more recently, copper. Typical properties for these wires are given in Tables 3.5 and 3.6. Other wires, such

TABLE 3.5.
Mechanical properties of bonding wire.

Material	Wire diameter,[a] μm	Temper[b]	Elongation, %	Tensile strength, MPa	Comments
Aluminum (99.99% pure)	18–75 (small diameter)	H	2−6	1.9–2.5	Softer than other wire. Sags more than other wires for equivalent diameters. Difficult to handle in small diameters.
		M	6−12	1.7–1.9	
		S	12−18	1.5–1.9	
	75–500 (large diameter)		5−10	1.4–1.5	
			10−20	1.0–1.4	
Aluminum + 1% silicon	25–250	H	1−5	2.9–3.5	Standard integrated circuit bonding wire (wedge bonding). Since 1% silicon greatly exceeds the room temperature solubility of silicon in aluminum, there is a tendency for Si to precipitate at bonding temperatures—unless the alloy is homogeneous at the nanometer level.
		M	5−10	2.2–2.6	
		S	10−20	1.5–1.9	
Aluminum + 0.5–1% magnesium	25–250	H	1−5	2.9–3.5	Does not form a precipitative phase since room temperature solubility in silicon is 2%. Excellent fatigue resistance—mitigates low cycle fatigue in power devices. Sometimes small amounts of palladium (0.1–0.15%) are added.
		M	5−15	2.2–2.6	
		S	10−20	1.5–1.9	
Gold (99.99% pure)	18–50	H	1−3	3.0–4.7	Mainstay ball bonding wire. Sometimes very hard gold wire (>7 MPa tensile strength, <1% elongation) is used for wedge bonding.
		SR	3−6	3.6–4.1	
		A	4−8	3.2–3.8	
Gold (98.5% pure) + 1% palladium	18–37		0.5−3	8.7–10.4	Formulated for stud bumping. Produces consistent uniform sized balls.

[a] Typical wire sizes available from various manufacturers.
[b] Temper: H = hard, M = medium, S = soft, SR = stress relieved, and A = annealed.

as palladium and silver have been bonded in the past as described above. Gold has been the dominant material used for the ball bonding process, while aluminum and its alloys predominate in the wedge (ultrasonic) bonding process. The gold used is extremely pure (99.99%) with total impurities typically less than 100 ppm. Beryllium is the key impurity used to stabilize the wire and control some of its mechanical properties. The gold wire used for stud bumping (single ended ball bonds) is not as pure, with a significant amount of palladium (∼1%) added to ensure the formation of uniform balls with minimum tails (residual wire remaining on the ball after wire is broken). Aluminum with 1% silicon matches the common alloy used for semiconductor device metalization and offers improved strength and stiffness over pure aluminum in small diameter applications. Pure aluminum is used in most large-wire applications, while aluminum and magnesium are used in cases where the interconnect is subject to conditions of low-cycle fatigue or on-off power cycling [67].

Because microelectronic bond wires are drawn through a series of dies, the as-drawn wire has significant residual strain and, while strong, is often brittle (low elongation). To

TABLE 3.6.
Thermal and electrical properties of bonding wire materials.

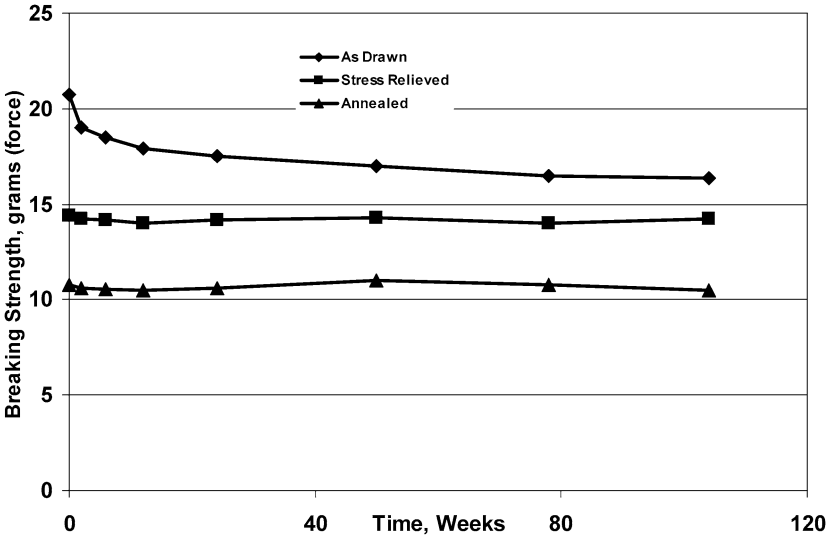
Material	Melting point, °C	Thermal conductivity, w/m K	Coefficient of thermal expansion $\times 10^{-6}$/°C	Electrical resistivity $\times 10^{-6}$ Ω cm	Electrical conductivity % IACS[a]
Aluminum (99.99% pure)	660	230	23–24	2.49−2.77	69−62
Aluminum + 1% silicon	600–630	195	22–23	2.96−3.18	58−54
Aluminum + 0.5–1% magnesium	654	180–195	22–24	3.01	57
Gold (99.99% pure)	1063	312	14–15	2.20−2.29	78−75
Copper (99.99% pure)	1083	395	16–17	1.72−1.81	100−95
Palladium (99.99%)	1552	75	10–12	10.75−15.63	16−11

[a] IACS = International Annealed Copper Standard. 100% IACS = 5.81×10^5/Ω cm.

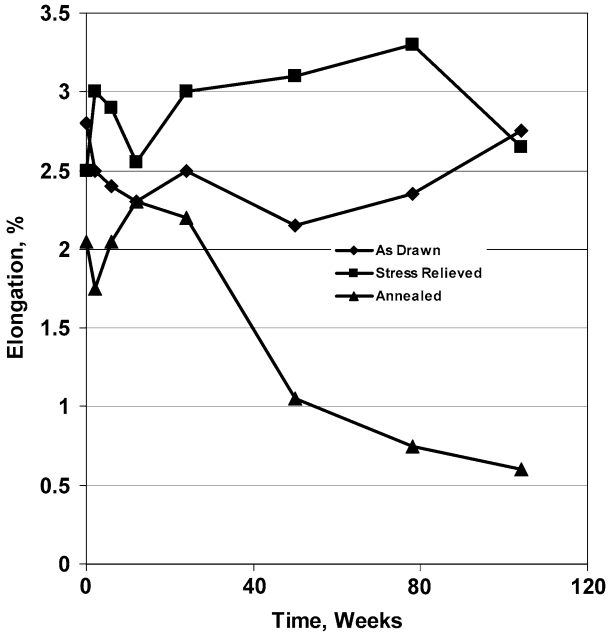
overcome these factors, the wire is typically strain relieved and sometimes annealed to achieve more desirable properties for the bonding process. Some of the effects of these post drawing processes can be seen in Tables 3.5 and 3.6. Figures 3.18 to 3.20 show the effects of time after manufacture (in controlled storage) on bonding wire properties for a few wire types. It is clear that depending upon the temper of the wire, storage time can have a significant affect on wire properties and hence on the quality of the bonds themselves.

There is great interest in replacing gold bonding wire with copper wire both for reduced cost, and, as the integrated circuit metalizations migrate to copper, for direct bonding to the copper pads. Thus, precluding the need for the copper pads to have barrier layer metalization (nickel-gold or titanium-tungsten gold), necessary with gold wire. Copper wire also has a high electrical conductivity and because of its strength, it resists wire sweep during the injection molding integrated circuit encapsulation processes. Since copper rapidly oxidizes in air, the ball formation process must be done in an inert atmosphere requiring significant bonding machine modifications. Copper has higher shear modules than gold (48 GPa versus 26 GPa) and Cu balls are significantly harder than gold balls (e.g., 50 compared to 35 on the Knoop Hardness Scale), thus, creating the potential for damage to delicate chips and substrates in the bonding process. Copper ball bonding produces a significant increase in cratering [21]. Several changes to bonding machine operation have been proposed as possible solutions to the copper hardness problem including increased substrate and capillary heat, reduced ultrasonic energy and a rapid first bond touchdown (to keep the ball hot and hence softer).

Bonding to copper pads, unless barriered as described above, could require significantly more ultrasonic energy due to the formation of copper oxides. In a similar vein, copper ball bonds made to conventional aluminum alloy pads seems to be viable. Copper-aluminum intermetallics exist ($CuAl_2$ and CuAl) and some studies have indicated rapid increases in joint resistance during thermal aging [45]. Most studies report the reliability of the copper-aluminum system to be equal to that of the gold-aluminum system. The bondability is probably more of an issue than the reliability, even with the mitigating measures described above, because the hard copper ball is likely to "push" the soft aluminum metalization aside during the bonding process, especially with today's thin IC metalizations (∼0.5 μm), resulting in a weak bond or a no stick situation. Similarly, cratering and the

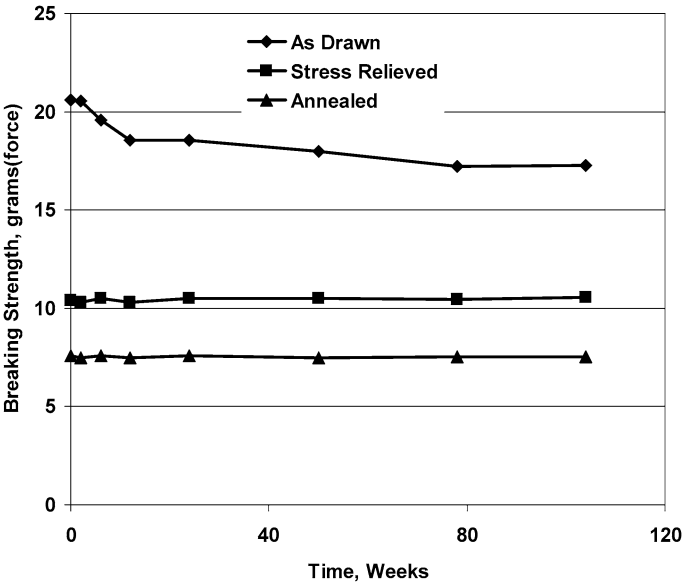


(a)

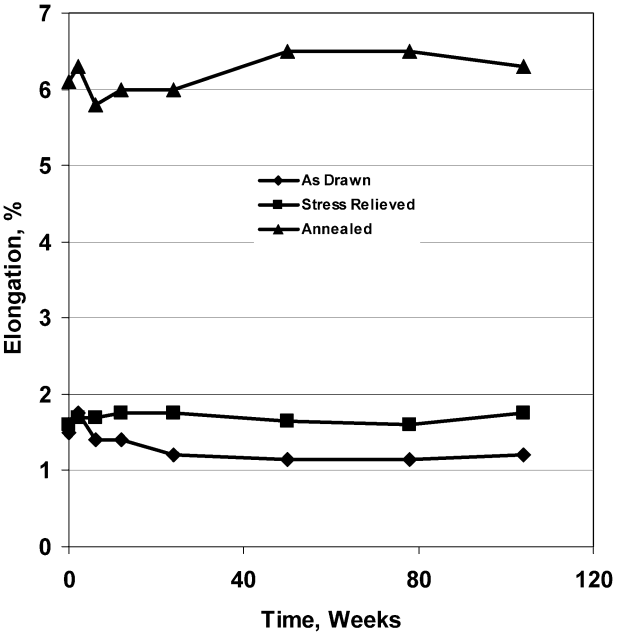


(b)

FIGURE 3.18. Breaking strength and elongation of aluminum bonding wire (Al + 1% Si) as a function of storage time for various wire tempers: (a) breaking strength grams (force); (b) elongation in %.



(a)



(b)

FIGURE 3.19. Breaking strength and elongation of gold bonding wire (99.99% Au + Be) as a function of storage time for various wire tempers: (a) breaking strength in grams (force); (b) elongation in %.

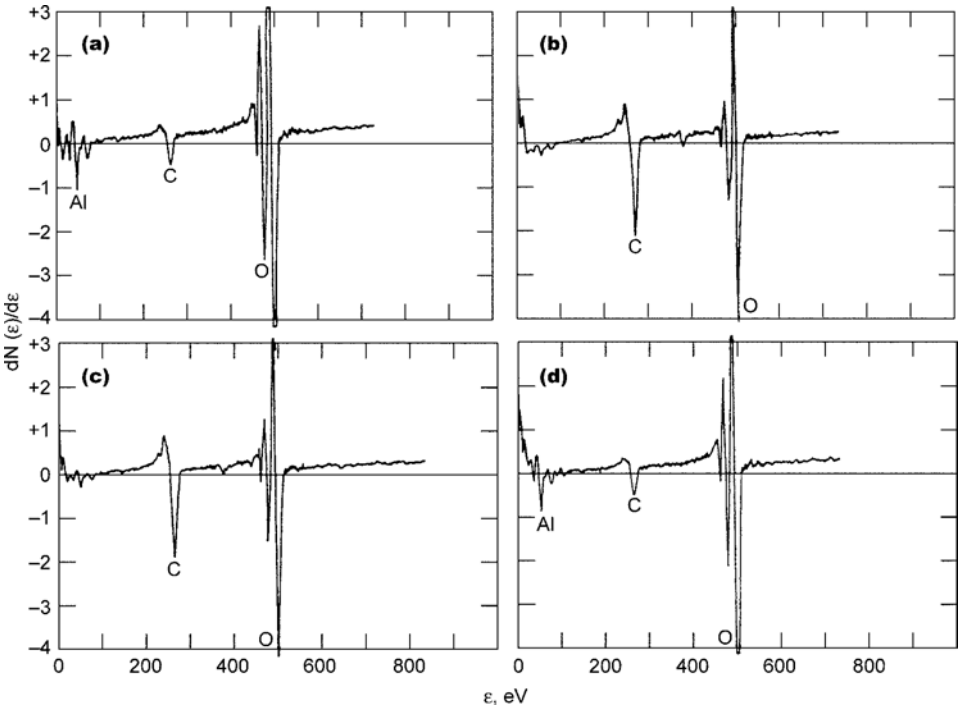


FIGURE 3.20. Auger electron spectra of aluminum metalized silicon substrates both pre- and post-cleaning with solvents and UV-ozone: (a) as processed substrate (uncontaminated); (b) substrate contaminated with photo resist; (c) substrate cleaned with solvent; and (d) substrate cleaned with UV-ozone.

susceptibility of Cu to corrosion (sulfur, halogens, etc.) could inhibit the widespread use of copper ball bonding. It is also very difficult to make small balls such as required for fine pitch wirebonding (see Section 3.4.1 below).

3.3.2. Bond Pad Metallurgy

Various metalization schemes have been used to ensure the bondability of chips, packages, and substrates. Unfortunately, most chip metalizations are selected for reasons other than the ability to form wirebonds. Historically, the typical chip metalization is aluminum containing a small percentage of silicon (typically 1%). The presence of silicon prevents the rapid diffusion of the underlying silicon (in the contact window) into the aluminum and thus reducing the pit formation in the silicon. Such pitting allows aluminum to migrate into the pits, creating aluminum conductive spikes which can damage performance or destroy device operation. Too much silicon in the aluminum can cause silicon precipitation during heat treatment and form silicon crystallites or nodules on the bonding pad surface and in contact with the underlying silicon. Such effects can cause both bonding and electrical problems.

To ensure adequate electromigration resistance as device geometrics shrink [59], alloying elements are also incorporated into or sometimes placed under the standard chip metalization. For example, copper is often added to aluminum and aluminum with silicon in concentrations between 0.5 and over 5% by weight to prevent electromigration. Above about 2 weight percent copper, the wire bondability of the aluminum-copper alloy has

been shown to decrease, while lower amounts have exhibited excellent bondability [77]. Aluminum with small amounts of copper, however, is subject to Al_2Cu hillock formation during thermal processing. These hillocks can cause interlayer shorts, etc. Thus, to prevent Al_2Cu hillocks, process engineers add more copper (>4 weight percent), which causes widespread hillock growth; but the hillocks are very limited in height, thus reducing the shorting potential at the expense of bondability. Higher copper levels also increases the susceptibility of aluminum to corrosion and may lead to surface oxide formation, which can further reduce bondability.

Titanium-tungsten or titanium nitride layers are sometimes added under pads to improve adhesion and to stiffen the pads on soft or flexible substrates. If process conditions are improperly controlled, these under layers can reduce bondability. Titanium also has been alloyed with aluminum metalization on chips to reduce electromigration. Again, potential titanium migration to the surface can cause bonding problems. The titanium also increases the hardness of metalizations, which in general requires more aggressive bonding parameters to effect high quality bonds. To achieve the highest bondability in the presence of titanium, bonding temperatures must be substantially increased ($\geq 180°C$), which requires the use of high-temperature die attach (e.g., the gold-silicon eutectic). As recommended by Harman ([36], pp. 243–246), capping with a thin layer of pure aluminum (0.25–0.5 μm in thickness) would allow various metalizations to be used and still provide the best metallurgy for high-yield bonding. Care must be exercised to keep the pure aluminum cap metalization thin, because it has been shown that bond strength decreases with increasing aluminum layer thickness [82].

Gold metalization also can be an effective cap to ensure bondability. Gold was originally used on some semiconductor devices. The pad stack typically was titanium-palladium-gold. Such pad stacks produced excellent bondability providing the gold thickness hardness and morphology were carefully controlled. Today, gold is rarely used on integrated circuits, but is widely used on package bonding pads and substrates to provide a wire bondable surface. The search for bondable gold has been the subject of many articles over the last decade or two. Gold deposited by thin film deposition is inherently bondable due to its purity and fine grain structure. Most gold bonding problems have been associated with either screen printed inks used in thick film or low temperature cofired processes or with plated gold.

The bondability of thick-film metalizations, particularly gold-based films, has been of concern for many years in the microelectronics industry. Statements such as "bondable" gold still appear in various forms in the commercial advertising literature without any quantification. The implication is that if you use the particular company's bondable gold that wirebound performance should approach the ideal, i.e., wirebond pull and ball bond shear strengths close to those obtainable with thin films. Historically, authors such as Jellison and Wagner [44], have shown that with clean substrates and thermocompression bonding, thick film gold substrates yielded similar ball bond shear strengths as comparable bonds made to thin-film gold. Some studies actually showed that bonding to thick film gold was less sensitive than bonding to thin films in the presence of surface contamination. The role of surface cleanliness prior to bonding on both thick and thin films cannot be over emphasized and it has been studied in great detail by several authors including Jellison [43], and Weiner et al. [82].

In the past, the role of surface composition, surface morphology and actual conductor or bonding pad geometry has not been addressed in detail to the same levels as the cleanliness problem. From the studies that have been performed, Spencer [73], Golfarb [31],

Prather et al. [64], it is clear that conductor composition, morphology, and geometry are extremely important factors in thick film bondability. Certain manufacturers over the years have "flattened" or "coined" the thick film at the bonding site by using special tools placed in bonding machines. Such processes are very expensive and time consuming.

Our studies (e.g., [69]), have shown slightly different but not necessarily conflicting results. In our studies we compared different metalization ink types: pure gold, lightly alloyed gold, and heavily alloyed gold. The pure gold was an oxide bonded gold made for wirebonding using gold wire. The lightly alloyed gold was oxide bonded and especially formulated to retard strength loss (due to intermetallic diffusion/formation) that occurred when making aluminum wirebonds. The heavily alloyed gold (which contained significant amounts of platinum and palladium) was primarily made for solder reflow operations. These metalizations were screen printed and fired using a test pattern consisting of various line and bonding pad sizes, ranging from 125 μm to 500 μm. Thin film vacuum deposited pure gold (3 μm in thickness) was also patterned and used as a reference in these bondability studies. Pad surface and line morphology and shape were measured using a scanning electronic microscope and a stylus profilometer, respectively. Surface impurities were analyzed by Auger electron spectroscopy and wirebond quality was assessed by both the ball bond shear test and the wirebond pull test (See Section 3.2.5).

The metalization type had the greatest effect on both the ball bond shear strength and wirebond pull strength. Pure thin film gold demonstrated the best bondability and had the highest average shear strength. The lightly alloyed thick film gold (made for aluminum wirebonding) gave results similar to the thin film gold. The pure thick film gold and the heavily alloyed gold produced significantly poorer results (e.g., 35 grams (force) shear strength compared to 48 grams (force) shear strength on average) for comparably sized and placed bonds. Surface morphologies were different between all four metalizations with the thin film surface being extremely smooth, small grained with no pores. The heavily alloyed gold surface was extremely porous and very rough compared to the other metalizations. The pure thick film gold and the lightly alloyed gold had similar morphology, although the lightly alloyed gold was slightly rougher and more porous.

In a design of experiments study, parameters such as surface porosity, surface curvature and pad or line width size were determined to be secondary effects. Ball location on bonding pads or lines seemed to have little effect on the thin film and pure thick film bonding results. As surface porosity and roughness increased effects associated with ball location became slightly more dominant. Tail bonds seemed to be more affected than the ball bonds. Mechanical operations such as burnishing (scrubbing with an abrasive) or coining appeared to have little effect and in the case of the heavily alloyed gold burnishing significantly reduced the bondability.

Results of the study indicated that surface composition was the key factor in bondability. This result is consistent with findings of Harman [36] in his Chapter 6 on plated golds. He further correlates bondability or lack there of with film hardness, i.e., soft gold is preferred. In our studies the hardness of the thick film layers increased with increasing impurity concentration, based on gold ball deformation, at given force level. No quantitative measurements of hardness were made.

3.3.3. Gold Plating

3.3.3.1. Electroplated Gold Impurities in electroplated gold layers have long been a source of bonding problems. Impurities have caused both low bonding yields and premature failures during accelerated testing or real life operational use. Horsting [40] presented

fundamental studies that related gold purity to the formation of "purple plague" and hence bond failures. Horsting believed that the accelerated diffusion of the impurities into bond intermetallic regions caused precipitates to form which acted as nucleation points for vacancies causing more rapid void formation during the normal interdiffusion of gold and aluminum. The actual impurities in the gold were not precisely determined by Horsting due to equipment limitations, but qualitatively he principally found nickel, iron, cobalt, and boron. Later, researchers confirmed Horsting's rapid impurity diffusion theories Newsome et al. [58].

Gold electroplating bathes typically consist of potassium-gold-cyanide solutions plus additives such as buffers, citrates, phosphates, carbonates, and lactates. Impurities such as thallium, lead, and arsenic are added to improve plating deposition rate and as modifiers to reduce grain size—hence changing surface morphology. Thallium has been the impurity most often linked to wirebonding problems [26,27], but work by Wakabayashi [81] identified lead as another significant cause. He also indicated that under certain plating conditions, arsenic could improve bond strength. Impurities such as lead and thallium can cause the gold crystal structure to change on the bonding surface. Surface morphology can also be changed by varying the plating parameters. To date there is not conclusive proof that the subtle changes in surface morphology in plated gold layers have a correlatable effect on bondability and bond strength, unlike the experiences with thick films above.

Other plated gold phenomena such as hydrogen entrapment and film hardness can also cause bonding problems. Hydrogen entrapment can be mitigated by annealing, providing the assembly can withstand the annealing environment (minimum of 2 days at 150°C). Such annealing, while removing hydrogen, also reduces its hardness. Hardness thus becomes a key bonding indicator, if not the root cause, of bondability problems.

3.3.3.2. Electroless Autocatalytic Gold The key to wirebonding on laminate technology for MCMs and COB implementations is the ability to do electroless gold plating on the pre-patterned copper metalization. In working with commercial plating vendors, electroless gold (autocatalytic) plating solutions can be found or developed with standard or modified chemistry that meet the deposition needs (99.99% pure gold up to 1 μm in thickness) for a variety of substrates and applications. Typical laminate processes require a nickel barrier layer over the copper. It is necessary that these autocatalytic gold processes be able to plate on nickel as well as on copper. Two major types of autocatalytic gold plating chemistries exist: (1) high deposition rate strongly basic systems containing cyanide; and (2) neutral pH systems without cyanide. The high deposition rate systems have a pH of about 12 and can erode certain circuit board materials such as polyimide during long plating runs. Several variants of these high deposition rate systems exist including ones which plate gold directly on copper and others which will plate gold onto nickel coated surfaces. Typical plating bath temperatures range from 70°C to 100°C. Such systems have been used to produce bondable gold, but the high bath temperatures, the difficulty in plating on nickel (requires exacting bath chemistry at all times), and the erosion of the substrate material has made these chemistries unsuitable for most organic-based MCMs and COB assemblies. Such chemistries are useful for plating circuits built on ceramic substrates.

The issues associated with the high deposition rate systems caused the development of neutral pH (nominally 7.5) autocatalytic gold processes. These baths contain no cyanide and can operate at 70°C or less and do not erode polyimide. With these systems bondable gold up to 1 μm in thickness can be deposited over nickel barrier layers. Compatible electroless nickel plating solutions exist for copper metalizations. The copper metalization

TABLE 3.7.
Wirebond pull strength for various thicknesses of autocatalytic gold plating over a nickel barrier (2.5 μm thick) on a copper metalized printed wiring board.

Gold plating thickness, μm	Number of bonds	NDPT[a] failures	Pull strength[b], grams (force)	
			As bonded	After 150°C aging[c]
0.40	129	1	10.6	9.8
0.65	149	0	10.0	10.1
0.90	138	0	9.4	10.6

[a] NDPT = non destructive pull test (at a 2.5 grams (force) limit).
[b] Sample sizes approximately 70 bonds. Standard deviations within ±10%.
[c] 160 hours (polyimide-glass board material).

must first be sensitized with a palladium-based activator. Table 3.7 presents some wirebond reliability data for gold bonds made to various thicknesses of autocatalytically plated gold (neutral pH). The data indicates that bonds remain strong even after extensive thermal aging at 150°C provided the gold is at least 0.65 μm in thickness. Other experiments have shown that a minimum of 0.5 μm is necessary to achieve uniform bonding and reliability after thermal testing.

3.3.4. Pad Cleaning

In order to make high quality, reliable wirebonds, the bonding pads must be clean. Many techniques have been tried over the years, but of all the methods, UV-ozone [82] and oxygen plasma [47] have proved to be the most effective in removing organic contamination. They are also effective against certain inorganic materials that form either a volatile oxide or, if not volatile, one that can be easily removed. While these techniques have been shown to remove a wide variety of contamination types, care must be exercised in their use. Because of the strong oxidizing environments present in O_2 plasma and UV-ozone reactors, metals such as silver, copper, and nickel may oxidize, and thus reduce their bondability. To reduce such effects in plasma reactors, argon is sometimes mixed with the oxygen. These oxygen-argon plasma cleaners are quite effective, combining reactive ion cleaning with physical sputter etching. With any kind of plasma environment, there is a possibility of active circuit radiation damage. Based on this author's experience, this probability is extremely low for oxygen-argon plasma cleaners and should not be viewed as a deterrent to their use. Similarly, because UV radiation can excite impurity states (color centers) in alumina-based ceramics, there is a tendency for white alumina ceramic substrates to appear yellow after UV-ozone treatment. The induced color change can be reversed by a subsequent thermal treatment. Table 3.8 and Figure 3.20 show the effectiveness of UV-ozone cleaning (over solvent cleaning) in removing intentional surface contamination.

Before leaving cleaning, a few comments about ultrasonic cleaning should be made. Historically, there have been several published reports (e.g., [68]), and much anecdotal conversation describing wirebond degradation or failure due to ultrasonic cleaning. Most of the reported incidents center on wirebonds in cavity type packages, such as those used for hybrids or hermetic single chip applications.

As with all mechanical structures, a wirebond has a resonant frequency which if excited will cause the wire to vibrate and in turn may cause fatigue and ultimate failure. The

TABLE 3.8.
Average ball bond shear strength (grams (force)) for various cleaning treatments and thermal aging conditions for thermosonically bonded 25.4 μm gold wire on 1 μm thickness aluminum (on silicon). Average ball diameter was 90 μm (±3 μm).

Sample set	Cleaning conditions	As bonded	Thermally aged
A	No clean[a]	50.9 (±7.1)	47.8 (±7.9)[b]
	Plasma clean[c]	52.2 (±6.5)	52.1 (±6.7)
B	No clean	50.0 (±6.2)	48.6 (±7.1)[d]
	Contaminated[e]	38.9 (±4.1)	40.3 (±5.8)
	Solvent clean	37.3 (±6.1)	37.9 (±7.3)
	Plasma clean[f]	47.5 (±6.0)	47.9 (±6.7)
	UV-Ozone clean	53.0 (±5.1)	54.2 (±5.8)

[a] No clean as received from substrate fabrication.
[b] Sample set A aged for 96 hours at 150°C.
[c] Argon-oxygen plasma (90% Ar, 10% O_2).
[d] Sample set B aged for 168 hours at 125°C.
[e] Contamination agents were photoresist and outgassing products of epoxy cure.
[f] Argon-oxygen plasma (50% Ar, 50% O_2).

resonant frequency of a given diameter bonded wire is dependent on the length and height of the loop. For reasonable geometries and relatively short lengths (<2.5 mm) the resonant frequency of a typical wirebond is quite high (>30 kHz). Historically ultrasonic cleaners operated in the 20 kHz regime, and most of the reported damage occurred with long wire bonds (>2.5 mm) placed in large industrial cleaners (high energy). Thus, the ultrasonic cleaning of cavity type devices with short wires should be safe. Today, ultrasonic cleaners span a broad frequency range from 20 to over 100 kHz. According to Harman ([36], p. 230), it is unlikely that high frequency ultrasonic cleaners (>50–60 kHz) will damage wirebonds.

With pin or ribbon leaded packages in which the pin or ribbon feeds directly inside the package to form the wirebond attachment point, special care needs to be taken to ensure that the external lead structure does not resonant. Resonance in these external leads can set up vibration on the pin or ribbon end inside the package and can cause wire or wirebond failure, especially if the wire is relatively stiff. This would be especially important when parts in quad flat packages are cleaned prior to board attachment.

With today's fully encapsulated microcircuits, the cleaning of parts ultrasonically poses little risk, especially for leadless or short leaded components. The potential danger occurs when cleaning exposed wirebonds in open packages or in COB or flex applications. Another potential danger could be associated with microelectromechanical systems (MEMS) where ultrasonic resonance could cause mechanical failure of the MEMS structures themselves in addition to the potential damage to wirebonds. Again, it is a question of the resonance frequency of the structure compared to the ultrasonic agitation frequency. In all cases with exposed wires and structures, if ultrasonic cleaning methods are employed, cavitation should be avoided [36].

3.4. ADVANCED BONDING METHODS

3.4.1. Fine Pitch Bonding

Fine pitch ball and wedge bonding is continuing to evolve rapidly. While most ball-bonded products are still in a pitch range of 100 μm and above, production quantities of



FIGURE 3.21. Scanning electron photomicrograph of ultrafine pitch (55 μm) thermosonic ball bonding. The bonds were made on a K&S Model 8020 automatic ball bonder using 23 μm (0.9 mil) diameter gold alloy wire. Pad metalization was Al + 1% Si + 2% Cu on SiO_2 with a nominal 1 μm thickness (photomicrograph courtesy of L. Levine, K&S).

90 μm pitch are being manufactured. Pitches in the 60 to 90 μm range have begun the transition to volume production, while pitches of 60 μm and below have been demonstrated on a limited scale (see Figure 3.20). Such bonds must be made with bottleneck or stepped-neck capillaries [24]. Today most bonding machines are limited to minimum pitches between 35 μm and 70 μm. An example of fine pitch ball bonding is shown in Figure 3.21. The bond is quite different from a traditional ball bond. It is quite low, almost nail head-like with a "ball" diameter in the range of 1.2–1.5 times the wire diameter. The low height of the nail head (typically 5 to 15 μm) makes the fine pitch ball bond difficult to shear. Most fine pitch ball bonds are still done with 25 μm diameter gold wire, although 18 to 20 μm wire is gaining popularity. Very fine pitches ($\leq$70 μm) and wires smaller than 25 μm in diameter are subject to greater damage in handling and molding operations than their larger more robust counterparts.

Wedge bonding leads the fine pitch parade. Wedge bonds at pitches of 40 μm have been demonstrated using 10 μm diameter gold wire. Wedge bonds at 60 μm and above are made in high volume production using 25 μm diameter gold or aluminum wires. To achieve such fine pitches, the wedge bonds typically have low deformation (1.2 wire diameters). Narrow, cutaway wedge tools are necessary to prevent adjacent wire damage during bonding. An example of 40 μm pitch wedge bonding is shown in Figure 3.22.

Fine pitch bonding is limited by the lack of chips with appropriately sized and spaced bonding pads that can take full advantage of the reduced size, high density wirebonding technology. Shrinking bonding pad size and pitch on chips is further hampered by limitations in test probe placement and movement. High frequency bonding (>60 kHz) has

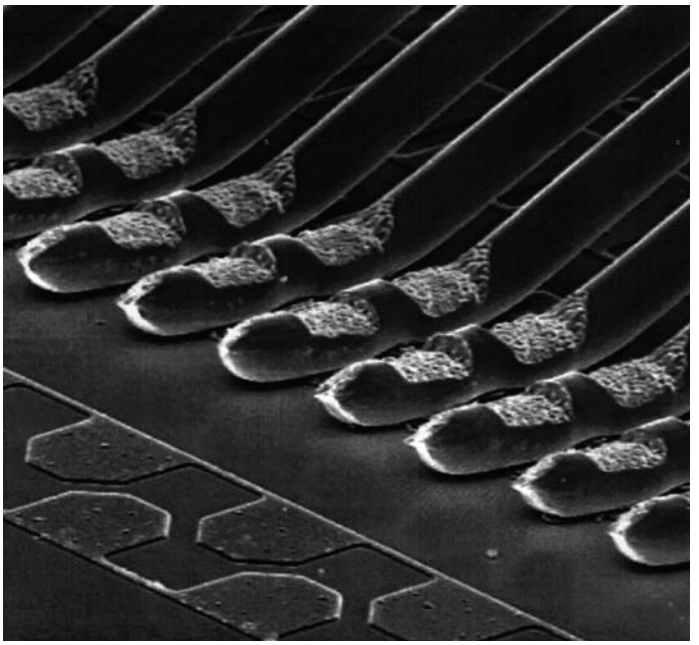


FIGURE 3.22. Scanning electron photomicrograph of ultrafine pitch (40 μm) wedge bonds. The bonds were made on a K&S Model 8060 automatic wedge bonder using 20 μm (0.8 mil) diameter gold alloy wire. Pad metalization was Al + 1% Si + 2% Cu on SiO_2 with nominal 1 μm thickness (photomicrograph courtesy of L. Levine, K&S).

been shown to be beneficial in bonding fine pitch circuitry [32]. More details on higher frequency wirebonding will be given in Section 3.4.4 below.

There are many issues associated with the implementation or use of fine pitch wirebonding. Fine pitch bonding can only be accomplished successfully if the entire process (chip, package or substrate, bonding machine, and bonding practice) is designed from the beginning with fine pitch in mind. The size, placement, and shape of the bonding pads must be coordinated with the selection of the wirebonding machine, the die attach machine and process, and the package or substrate (board) layout. Square bonding pads (hexagonal or round, also) are optimal for ball bonding but pose some limitations for wedge bonding. Ideal wedge bonding pads would be long and narrow [61]; but these are seldom used because of the need to be flexible in bonding method choice and that automatic wedge bonders are, at best, a factor of two slower than automatic ball bonders (due to the need to index either the bonding head or the sample table to maintain wire alignment under the wedge). Thus, a high volume wedge bonded product will cost more than a product interconnected by ball bonding methods, even given the difference in wire cost (aluminum vs. gold, respectively).

It also should be recognized that extremely fine pitch, with any bonding technology, can result in higher costs due to added constraints, reduced throughput (generally lower bonding speeds), and typically a more fragile product. Both equipment and workers associated with the fine pitch process are typically more expensive than those associated with a conventional (low pitch) process. Automatic bonders need the latest in precision pattern recognition coupled with the most accurate placement control. Programming time is greater, and workers must be better trained to master the art of fine pitch. Die attachment machines also must have greater accuracy in the placement process than machines used

in conventional pitch processes. Packages, as the wirebonding pitch declines (especially with multiple tiers of bonding pads), must be carefully designed to give maximum bonding tool access, while minimizing chances of wires touching or wire misplacement causing shorting. The ultimate design practice will force package and substrate pitches to those of the chip, thus minimizing fan out and keeping wire lengths short, which will reduce lead inductance and minimize injection molding wire sweep [76]. Copper wire could have an advantage in both electrical performance and mechanical integrity, but the ability to form minimal size balls (necessary for fine pitch) is still in question.

The solution to wires touching and shorting in fine pitch wirebonding could be the use of insulated wire. Insulated wire with appropriate bonding pad capping metalization could also allow COB assemblies to be made without insulating glob tops on overcoats. Insulated bonding wire has been around for over 20 years, but has never received widespread attention, mainly due to a host of implementation/reliability problems including wire coating contamination of the capillary, flame off inconsistencies, and low second bond strength. Recent advances [55] in wire coating technology appear to have made the specter of coated wire viable. Coated bonding wire has obvious advantages including allowing wires to be close together, cross, and even touch. Such ability could solve wire sweep issues and die/wire shorting problems encountered in stacked die or in high density wirebonding in general. The newer coatings appear to be about 0.5 μm in the thickness on 25 μm gold wire with breakdown strengths approaching 200 volts and the ability to survive baking temperatures of 300°C. Wire strength and bonding ability appear not to be reduced by the coating.

3.4.2. Soft Substrates

Deformable or soft substrates in modern wirebonding applications are usually associated with organic-based boards or layers as follows: thin-film, multilayer structures on inorganic carriers such as encountered in multichip modules (MCM-Ds); laminate-type organic constructs such as encountered in printed wiring boards (PWBs), MCM-Ls and COB structures [16]; and chips mounted to unreinforced laminates and/or flexible film layers.

MCM-D modules are made using deposited dielectric and thin-film metal layers. The carrier for these deposited films is usually silicon, although polished ceramics have been used in the past [12]. The dielectric materials are typically spun-on layers of polyimide. Benzocyclobutene (BCB), and several lesser-known polymers [74] also have been used. These dielectric layers usually range in thickness from 5 to 25 μm (or more), with as many as six layers being reported. Metalization schemes have been gold (with suitable adhesion layers such as chromium and tungsten), copper (again with suitable adhesion layers), and aluminum. In addition to organic dielectric layer softness, metal adhesion has been a challenge and requires careful processing to ensure metal layer integrity and inner layer adhesion.

In bonding to MCM-D structures, both thermosonic ball bonding and ultrasonic wedge bonding have been used [54]. In bonding to MCM-Ds, two major issues arise: (1) the size of the bonding pad and (2) the number and thickness of the soft layers (polyimide, BCB, etc.) under the pad. It has been shown [15] that the pad bends or cups under the application of the bonding force. This cupping is due to the compliant nature of the organic material. Elevated temperatures exacerbate the issue, effectively softening the polymer even more. Small bonding pads have less area over which to distribute the load and are thus more susceptible to this cupping or bending phenomenon. Pad deformations un-

der bonding forces and the application of ultrasonic energy have been studied by Takeda et al. [75].

Their results show that normal sized gold pads on copper traces (on polyimide flex boards) can deform as much as 20 μm under normal (but high end of the range) force and ultrasonic energy bonding conditions. They also verified that the use of a nickel under layer (under the gold pad) can significantly reduce the deformation below 10 μm for all bonding conditions. Others have noted similar deformations but the amount of deformation was smaller. In our work, for example, we have observed that for a given bonding force, the deformation increases with organic layer thickness. Pad reinforcement structures and interlayer metalization tend to mitigate deformation. Similarly, a marked decrease in deformation was observed as the bonding force was reduced in all samples, with little or no correlation to changes in sample thickness.

In addition to unreinforced substrate materials, MCM-L and COB implementations can use fiber reinforced organic matrix material such as polyimide or epoxy. The reinforcing fibers are typically glass, although materials such as Kevlar®, quartz, and Aramid® have been used. Sometimes high-frequency circuitry is built on non-fiber reinforced substrates with very low dielectric constants such as Teflon® (polytetrafluoroethylene). Most of these "laminate" technologies use copper metalization protected by thin layers of plated gold (usually with a nickel barrier layer under the gold). The thicknesses of both the metal and dielectric layers are larger than those of the MCM-D technology by factors of 5 for the metals and at least an order of magnitude or more for the dielectrics.

Other MCM-L implementations use fiber reinforced cores with non-reinforced resin layers on their surfaces [35]. Such structures can employ a variety of metalization schemes put in place and patterned by a combination of thin-film deposition (MCM-D) and PWB techniques. Via fills can be plated or actually filled with conductive organic resins [33].

Wirebonding to most MCM-L substrates including those in ball-grid arrays (BGA) and chip-scale packages (CSP) is similar to bonding to PWBs provided the substrates are made with fiber-reinforced resin laminates (e.g., polyimide-glass, epoxy-glass). Direct bonding to PWBs has been done for some time in COB applications. Many problems still exist with bonding to standard PWB fiber reinforced laminates, let alone the new problems associated with reduced pad sizes, unreinforced organic layers, and different via construction techniques found in today's MCM-Ls, BGA and CSP substrates, and integrated circuit redistribution layers. Both aluminum wedge bonds and thermosonic gold ball bonds have been used in COB applications. Wedge bonding is often preferred because it can be done without added substrate heat. Large COB assemblies will tend to warp and possibly soften if heated to or near their glass transition temperature (T_g). FR-4 (epoxy-glass) circuit boards have a T_g around 120°C, while T_g of various polyimide boards exceeds 200°C. Such high-temperature resins can be thermosonically bonded provided proper substrate clamping and backside support is available for large area assemblies. Successful thermosonic bonds have been made at temperatures below 100–110°C so that even FR-4 can be bonded. Even with the thick metalizations typically encountered in the COB arena (e.g., nominally 17–35 μm), anomalies can exist in wirebonding, especially as pads shrink in size. Bonding to BGA and CSP flexible substrates is typically done with gold-ball bonding because of the need for controlled shape bonds and bonds that are very close to the chip edges to keep the package footprint as small as possible. Because of the small area and reduced thickness of the substrate, special care has to be exercised in the bonding process.

In addition to flexible and software substrates, two other difficult bonding situations exist in both: thinned-die and stacked-die (either thinned or not). Thinned die have been around for some time, especially in microwave applications where gallium arsenide (GaAs)

microwave devices have been thinned to 100 μm or less to provide better thermal performance. Gallium arsenide is more susceptible to bond cratering and to mechanically induced electrical defects than silicon. For a detailed study of cratering on silicon die, see the paper by Clatterbaugh and Charles [23]. GaAs is weaker than silicon by a factor of 2. The two major material characteristics or parameters that are most relevant to cratering have been shown to be hardness and fracture toughness. Hardness is a measure of the material resistance to deformation while fracture toughness is a measure of the energy (or stress) required to propagate an existing microcrack. The Vicker's hardness for GaAs is 6.9 ($\pm$0.6) GPa while silicon is 11.7 ($\pm$1.5) GPa. In a similar vein, the fracture toughness of GaAs and silicon are 1.0 J/m^2 and 2.1 J/m^2, respectively. Thinned silicon die are now being mounted to flexible circuit boards. Silicon die as thin as 25 μm have demonstrated electrical integrity. Wirebonding, because of the thinness of the die and the softness of the flexible substrate, has proven difficult and most of these assemblies have been flip chipped (i.e., attachment by solder reflow [3]).

Stacked die present their own set of issues, but in general, the problems involve multiple geometries in a given component package with closely spaced wirebonds that can overlap. In addition, sometimes the bonding must be done to chips that are cantilevered over another chip without a means of mechanical support under the bonding pad areas. Fixturing and very careful control of bonding parameters (reduced force and power, higher frequency, and temperature) has allowed successful wirebonding to stacked geometries with as many as six chips. A full discussion of the details of wirebonding to stacked chips is not possible in this work, but some insight can be gained by reading Yao et al. [84].

3.4.3. Machine Improvements

Many bonding machine improvements have been made. While some are manufacturer specific, most are commonly available throughout the industry. These improvements are typically aimed at improving the speed of bonding; increasing the accuracy of the bond placement for fine pitch; improving the bondability of difficult-to-bond metalizations and substrate structures; and controlling the complete bond geometry for greater repeatability, reliability, and, of course, electrical impedance control for high-frequency applications. Such improvements include air bearings to increase bonding speed and reduce machine down-time owing to wear, laser interferometry for precise head positioning, and improved pattern recognition software to enhance learning and bonding speeds as well as encompassing larger chip libraries.

Other software improvements allow complete control of bond shape and length. Such control allows the repeatable fabrication of bonds with a given impedance for microwave and wireless circuitry. Another ramification of bonding machine improvement is the potential use of higher frequency ultrasonics (up to 300 kHz). Research suggests that higher frequencies can reduce bond dwell time and still achieve high quality bonds. Similarly, the application of higher frequency ultrasonics has been reported to enhance the bondability of difficult substrates such as soft ones encountered in MCM-D and COB applications. See Section 3.4.4 below. The introduction of a delay (after force application) prior to the onset of the ultrasonic energy burst has also been shown to be effective in difficult bonding situations.

3.4.4. Higher Frequency Wirebonding

Most of the world's current wirebonding machines have ultrasonic generators and transducers that operate at nominally 60 kHz. The choice of 60 kHz was made several

decades ago based on transducer (bonding head) dimensions for microelectronic assemblies and stability during the bonding (transducer loading) operation [36]. Other frequencies from 25 to 300 kHz have been used to attach wires. Ultrasonic welding and material softening have been reported in the range between 0.1 Hz [83] and 1 MHz [49]. Today's interest in higher frequency bonding stems from reports by various authors [32,38, 42,66,72,78] that using higher ultrasonic frequencies produces better welding at lower temperatures in shorter bonding times (dwell times). It has also been indicated that higher frequency wirebonding improves bonding to pads on soft polymer layers such as Teflon® or unreinforced polyimide. While all these improvements were real for the particular situations in hand, few if any controlled studies (systematic, side-by-side experiments on the same substrates with an attempt to control all variables except frequency) have been performed. The following material presents excerpts from the one such study [18,19].

Three metalization schemes were used in this study: (1) aluminum (99.99% pure) with a titanium-titanium nitride (Ti/TiN) adhesion layer; (2) aluminum plus one percent silicon alloy (Al + 1% Si) again with a Ti/TiN adhesion layer; and (3) gold metalization with a titanium-tungsten adhesion layer (TiW). The metal bonding pad formation layers were sputter deposited to thicknesses between 1 μm and 2 μm on silicon base layers. The silicon wafers were p-type with a nominal resistivity of 30–50 Ω·cm. The wafers were thermally oxidized to achieve an SiO_2 thickness of 1 μm prior to metal deposition or spin coating with polyimide. The polyimide layers were between 5 and 20 μm in thickness. The gold metalization was also deposited on highly polished ceramic (99.6% pure alumina) substrates.

Various test structures were photolitographically patterned on each of the metal layers [15,17]. The patterns included: arrays of bonding parts of varying sizes (150 μm to 25 μm square), a daisy chain pattern consisting of almost 650 wirebonds with the resistance of the wirebonds accounting for over 60% of the total resistance of the circuit, and a radially distributed wirebond pattern for shock and vibration testing.

All wirebonding for the study was performed with two semiautomatic thermosonic ball bonders (Marpet Enterprises, Inc., Model 827) equipped with negative electronic flame off (Uthe Technology, Inc., Model 228-1) for uniform control of free air ball size. The flame offs were adjusted to produce 60 ± 2 μm diameter free air balls as shown in Figure 3.10. One of the MEI Model 827 wirebonding machines was equipped with a UTI Model 25ST (64.1 kHz) transducer driven by a standard UTI Model 10G ultrasonic generator. The other Model 827 wirebonding machine was equipped with a UTI Model 4ST (99.5 kHz) transducer which was driven by a UTI 10G generator tuned for 100 kHz. In order to make both transducer waveforms similar since the Model 25ST transducer is much larger than the Model 4ST a short 60 kHz transducer Model 17STL (63.1 kHz) was also used. A comparison of the transducer dimensions is given by Charles et al. [17]. The uniformity of the as-bonded product (both 60 kHz and 100 kHz) is shown in Figure 3.23.

This study has yielded a large amount of data. Key observations and findings include the following. It is clear that significant differences exist between bonding at nominally 60 kHz and bonding at 100 kHz. In addition to differences in transducer electronic waveforms between the standard 60 kHz (long) and the 100 kHz transducer, there exist differences in bonding machine optimization behavior. The 60 kHz system appeared to have a larger bonding window (i.e., for a given force and substrate temperature), and a wider range of ultrasonic power and dwell times produced acceptable bonds (strong, yet not over bonded or with wire damage) when compared to the bonds produced by the 100 kHz system. The 100 kHz bonding window, in addition to being smaller than the 60 kHz window,

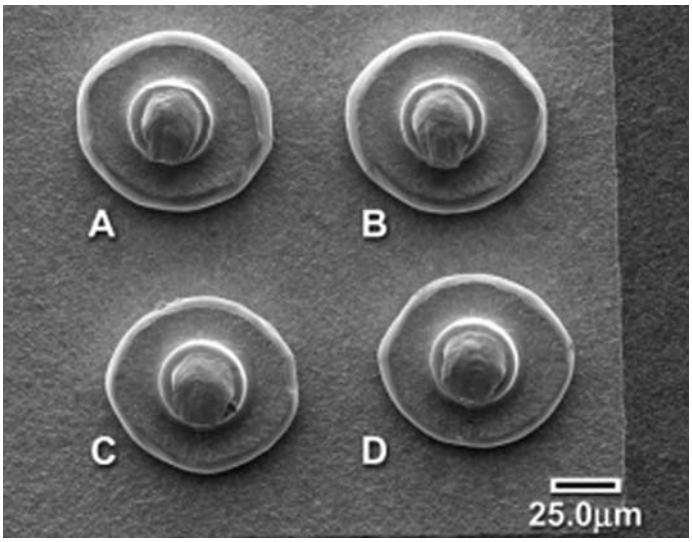


FIGURE 3.23. Scanning electron photomicrograph of single ended ball bonds bonded to substrate pad metalization. Balls A and B were bonded at 100 kHz. Balls C and D were bonded at 60 kHz. The average diameter for balls A and B is 80.5 µm. The average diameter for balls C and D is 77.2 µm. The magnification is approximately 300×.

TABLE 3.9.
Gold thermosonic ball bond shear strength (grams (force)) on gold and aluminum (1% Si) metalizations at both 60 and 100 kHz[a].

Metal	60-kHz	100-kHz	|Δ means|	Significant[b]
Au (on ceramic)	68.4 ± 3.7	84.8 ± 6.5	16.4	Yes (highly)
Al + 1% Si (on silicon)	54.0 ± 3.2	50.6 ± 2.9	3.4	Yes
|Δ means|	14	34.2		
Significant[b]	Yes (highly)	Yes (highly)		

[a] Nominal sample size at each frequency was 100.
[b] 99% confidence that the difference in the means are significant using analysis of variance with the F-test.

was also sharper (i.e., a smaller change in ultrasonic power and/or dwell in relationship to the window edge was required to go from either a no-bond condition or to an over-bonded condition when compared to the 60 kHz system). Despite the smaller, sharper fall-off of the bonding window, the 100 kHz system has one obvious advantage. It formed strong bonds in times that are 30 to 60% shorter than comparable dwells for the 60 kHz system. Comparison of both bonding systems and their transducer waveforms indicate that the 100 kHz system has much faster bonding pulse rise and fall times, along with a more stable voltage (or current) amplitude envelope that the 60 kHz system. Switching to a short 60 kHz transducer with dimensions comparable to those of the 100 kHz transducer produced ultrasonic drive parameters (voltage and current) similar to those of the 100 kHz transducer.

Shear test data on gold substrate metalizations showed that an optimized 100 kHz system produced much stronger bonds than the 60 kHz system (see Table 3.9). As can be seen from Figures 3.10 and 3.23 and Table 3.10, this difference cannot be accounted for by

TABLE 3.10.
Gold thermosonic ball bond average diameters[a] (μm) on gold and aluminum (1% Si) metalizations at both 60 and 100 kHz.[b]

Metal	60 kHz	100 kHz	|Δ means|	Significant[c]
Gold	89.1 ± 4.0	88.3 ± 2.9	0.8	No
Al + 1% Si (on silicon)	91.3 ± 2.3	92.0 ± 2.0	0.7	No
|Δ means|	2.2	3.7		
Significant[c]	Yes	Yes		

[a] Average diameter $= \frac{1}{n}\sum_i^n [(X_i + Y_i)/2]$.
[b] Nominal sample size at each frequency was 100.
[c] 99% confidence that the difference in the means are significant using analysis of variance with the F-test.

TABLE 3.11.
Gold thermosonic ball bond diameters (in μm) in directions perpendicular (X-direction) and parallel (Y-direction) to the direction of the ultrasonic scrub on gold and aluminum (1% Si) metalizations at both 60 and 100 kHz.[a]

Metal	Frequency	X-direction	Y-direction	|Δ means|	Significant[b]
Gold (on ceramic)	60 kHz	84.9 ± 4.8	93.2 ± 5.2	8.3	Yes (highly)
	100 kHz	82.8 ± 3.4	93.8 ± 4.0	11.0	Yes (highly)
Al + 1% Si (on silicon)	60 kHz	93.9 ± 2.9	88.7 ± 3.4	5.2	Yes (highly)
	100 kHz	98.7 ± 2.8	85.4 ± 2.6	13.3	Yes (highly)

[a] Nominal sample size at each frequency was 100.
[b] 99% confidence that the difference in the means are significant using analysis of variance with the F-test.

ball diameters (either pre- or post-bonding), which were essentially the same for both the 60 and 100 kHz systems. When the data was analyzed for the Al + 1% Si metalization (on oxidized silicon), the 60 kHz bonds appeared stronger. Although the difference between the 60 kHz and 100 kHz test results was relatively small (less than 7%). However, when analysis of variance techniques were applied, the difference was significant at the 99% confidence level. Similar results were observed on full thermosonic ball bonds attached to an integrated circuit chip (Al + 1% Si metalization), on which both the ball shear test and the wirebond pull test gave a small edge to the 60 kHz system. Although this data set was relatively small, the student's t-test indicated that the results were significant at the 99% confidence level. Independent of frequency, the differences in ball bond shear strengths between metalization types, were relatively large and highly significant. Bonds on gold were always stronger than bonds on Al + 1% Si metalization consistent with the results shown in many previous studies [11–19].

Other differences were observed such as asymmetry of ball shape with metalization type. No differences in average ball diameters [(X-diameter + Y-diameter)/2] were observed with frequency. Any variations in average ball diameters even those between metalizations (Table 3.10) could be accounted for by differences in the free-air ball size. On the other hand the differences in the X and Y diameter measurements are highly significant and appear to depend on metalization type (Table 3.11). On gold metalization, the as bonded ball diameter in the Y-direction or the direction of the ultrasonic scrub is larger than the orthogonal non-scrub diameter (X-direction) with consistent measurements for

TABLE 3.12.
Gold thermosonic ball bond shear strength (grams (force)) on gold and aluminum (1% Si) metalizations at both 60 and 100 kHz under conditions of thermal aging.[a]

Metal	Aged[b]	60 kHz	100 kHz	|Δ means|	Significant[c]
Gold (on silicon)	No	81.4 ± 4.6	97.4 ± 3.7	16.0	Yes (highly)
	Yes	82.1 ± 3.3	96.4 ± 4.6	14.3	Yes (highly)
|Δ means|		0.7	1.0		
Significant[c]		No	No		
Al + 1% Si (on silicon)	No	47.0 ± 3.7	46.5 ± 4.3	0.5	No
	Yes	57.8 ± 3.3	56.1 ± 4.1	1.7	Yes (slightly)
|Δ means|		10.8	9.6		
Significant[c]		Yes (highly)	Yes (highly)		

[a] Nominal sample size at each frequency was 100.
[b] 120 hours at 150°C.
[c] 99% confidence that the difference in the means are significant using analysis of variance with the F-test.

both 60 and 100 kHz. On Al + 1% Si, the non-scrub direction (X-direction) is larger than the Y-direction by a significant amount for both the 60 kHz and 100 kHz bonding systems. Similar behavior was also observed for pure aluminum metalization. The cause of these phenomena is not well understood but is believed to be associated with the dynamics of the weld formation process. On gold there is the single interdiffusion of the gold wire and gold pad materials. On aluminum and aluminum alloys the formation of gold-aluminum intermetallics is key to the bonding process. The formation of the relatively hard intermetallics may tend to lock the developing bond in the direction of the scrub on the aluminum and aluminum alloy metalizations while on the gold (being relatively ductile) the bond may be able to fully expand in the scrub direction.

Table 3.12 shows results for both 60 and 100 kHz bonded samples under conditions of thermal aging (120 hrs at 150°C). Aging at 150°C has been shown to be very effective [15] for assessing wirebond (ball bond) quality and reliability without introducing unwanted effects caused by substrate interactions and other heat-related phenomena. Table 3.12 again illustrates the significant improvement in shear strength using 100 kHz bonding on gold metalization, this time for gold on a silicon substrate as compared to the gold on ceramic data given in Table 3.9. The small observed differences on the Al + 1% Si metalization for 60 kHz versus 100 kHz is also consistent with the results in Table 3.9, although in this case the difference is statistically insignificant at the 99% confidence level. Again, large and significant differences were observed in the shear strengths between the two metalizations with bonds to gold being much stronger than bonds on Al + 1% Si metalization. These results are consistent regardless of the bonding frequency. After aging, the shear strength of the bonds on gold, at both frequencies, remained essentially unchanged. On the Al + 1% Si metalization the strength of the bonds increased significantly for both frequencies. Again, 100 kHz bonding produced stronger bonds on gold metalization, while 60 kHz bonding appeared to have a slight edge on Al + 1% Si. The increase strength for the aged bonds on the Al + 1% Si metalization is consistent with similar increases reported previously under aging [13], but the timeframe for the existence of the increased strength above the as-bonded condition appears to be longer in this particular experimental series.

3.4.5. Stud Bumping

Alternative forms of flip chip technology, make use of wirebonder-produced bumps. These alternatives include the "stud bump and glue technique" using standard or anisotropically conductive adhesive as shown in Figure 3.24.

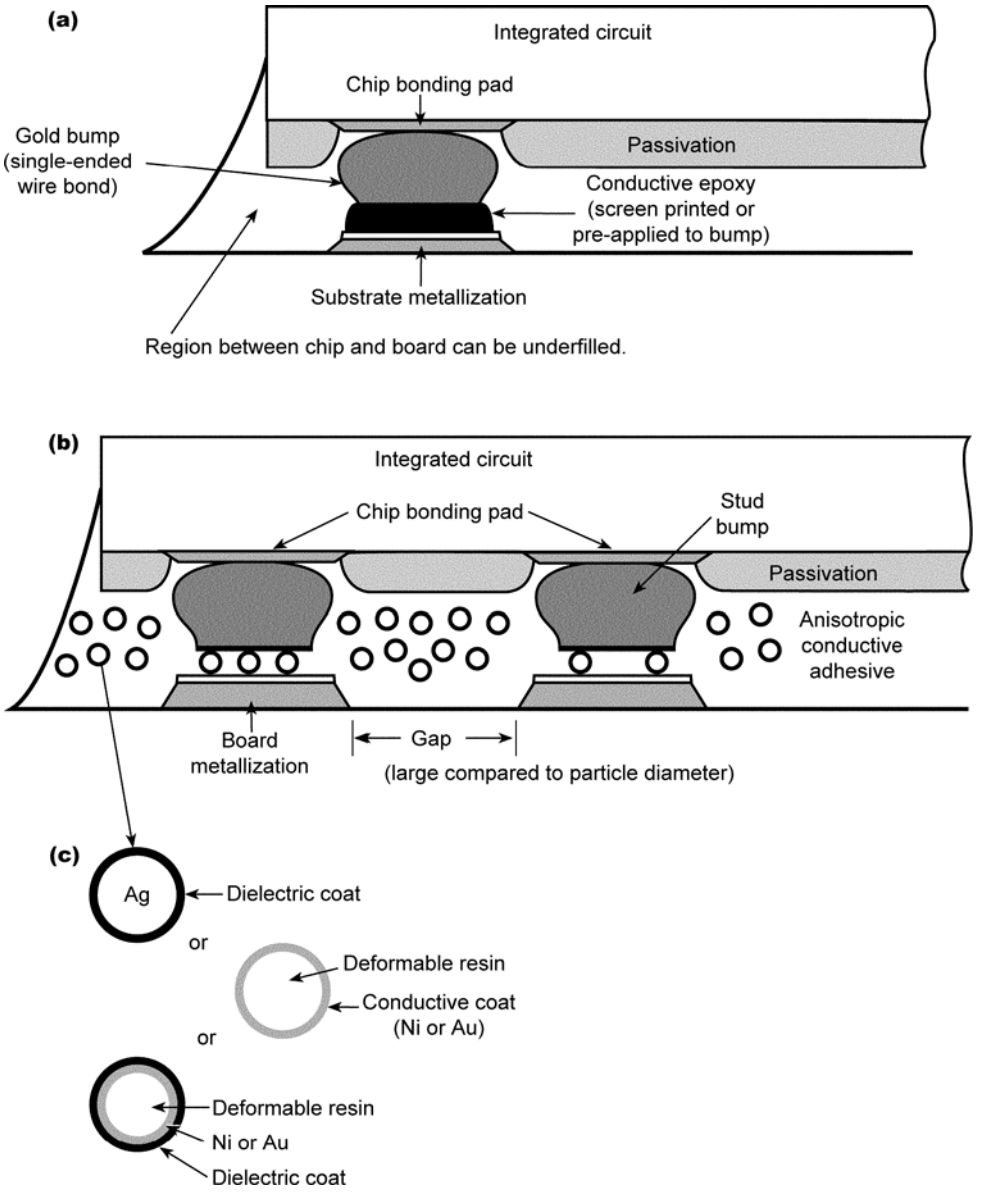


FIGURE 3.24. Schematic representation of attachment of a gold stud-bumped chip to a mating substrate using adhesive processes: (a) flip chip attachment using conventional conductive epoxy (screen printed or pre applied to the bump). Following adhesive cure the region between the chip and the board could be underfilled; (b) flip chip attachment using anisotropic conducting film. The anisotropic film not only makes the electrical contact but also acts as an underfill; (c) anisotropic adhesive conductive particle detail.

In the stud bump and glue technique, single-ended, thermosonic wirebonds are placed on the chip bonding pads by means of an automatic wirebonder using special bonding wire (Table 3.4). The balls are then coined or tamped to a uniform height using a special tool placed in the wirebonder. The stud-bumped chip is then pressed on a plate containing a thin layer of conductive adhesive (epoxy). As the chip is lifted from the plate, a small amount of conductive adhesive adheres to each bump. The chip is then placed on the corresponding substrate pads and the adhesive is cured, resulting in the geometry shown in Figure 3.24. In an alternative method, the epoxy can be preapplied to the substrate pads by screen printing or automated dispensing.

An anisotropic adhesive is an adhesive that has small conductive particles embedded in its non-conducting organic matrix. A bumped chip is then pushed down into the adhesive, capturing a few conducting particles between the bump and the mating bonding pad on the package or substrate. When the adhesive is cured, an electrical interconnect is made. In addition, the region between the chip and the board becomes rigid, mechanically holding the chip to the board. Thus, the adhesive also serves as an underfill [41].

3.5. SUMMARY

Wirebonding continues to be the dominant form of first-level chip connection. Approximately 90% of the world's chip production is wirebonded. Because of its sheer volume, flexibility, and low cost, it will continue to dominate chip interconnect for decades to come. Wirebonding is accomplished by three basic techniques using a variety of wire and pad metallurgies. Wirebonding is robust and, on rigid substrates, has been shown to be extremely reliable (defects in the low part per million range). Bonding to softer substrates, small pads, unconventional metallurgies and stacked components has presented challenges—challenges that wirebonding has been successful in meeting. With appropriate care and understanding of the processes, wirebonding, even under these challenging conditions, can be performed reliably with high yield. Wirebonding is continually improving through advancements in automation, refinement of welding kinetics, improvements in wire and pad metallurgies, improved cleaning methods, and a better and widespread understanding of wirebonding science.

ACKNOWLEDGMENTS

The author greatly acknowledges the support of JHU/APL's Electronic and Mechanical Services Groups on sample preparation and testing. Special thanks is given to Ms. Nancy L. Pickett for manuscript preparation.

REFERENCES

1. ASTM Standard Test Method: F458-84(1995)el, Standard non-destructive pull testing of wire bonds, Annual Book of ASTM Standards, West Conshohocken, Pennsylvania, USA, 1995.
2. ASTM Standard Test Method: F1269-89(1995)el, Test method for destructive shear testing of ball bonds, Annual Book of ASTM Standards, West Conshohocken, Pennsylvania, USA, 1995.
3. C.V. Banda, D.J. Mountain, H.K. Charles, Jr., J.S. Lehtonen, A.C. Keeney, R.W. Johnson, T. Zhang, and Z. Hou, Development of ultra-thin flip chip assemblies for low profile SiP applications, Proc. 37th Int. Microelectronics Symposium, Long Beach, California, 2004, pp. 551–555.

4. J. Bardeen and W.H. Brattain, The transistor, a semiconductor triode, Physical Review, 74, p. 230 (1948).
5. W.D. Barnhart, J. Van Rij, J.M. Petek, and H.K. Charles, Jr., The impact of KGD and module repair on multichip module costs, Proc. 32nd International Microelectronics Symposium, Chicago, Illinois, October 26–28, 1999, pp. 373–387.
6. A. Bischoff, F. Aldinger, and W. Heraeus, Reliability criteria of new low cost materials for bonding wires and substrates, Proc. 34th Electronic Components Conference, New Orleans, Louisiana, USA, 1984, pp. 411–417.
7. H.K. Charles, Jr., Tradeoffs in multichip module yield and cost with KGD probability and repair, Microelectronics Reliability, 41(5), pp. 715–733 (2001).
8. H.K. Charles, Jr., B.M. Romenesko, O.M. Uy, A.G. Bush, and R. Von Briesen, Hybrid wirebond testing—variables influencing bond strength and reliability, The International Journal for Hybrid Microelectronics, 5(1), pp. 260–269 (1982).
9. H.K. Charles, Jr., B.M. Romenesko, G.D. Wagner, R.C. Benson, and O.M. Uy, The influence of contamination on aluminum-gold intermetallics, Proc. Int. Reliability Physics Symposium, San Diego, California, USA, 1982, pp. 128–139.
10. H.K. Charles, Jr., G.V. Clatterbaugh, and J.A. Weiner, The ball bond shear test: its methodology and application, in D.C. Gupta, Ed., Semiconductor Processing, ASTM STP 850, 1984, pp. 429–457.
11. H.K. Charles, Jr., Ball bond shearing: an interlaboratory comparison, Proc. International Microelectronics Symposium, Atlanta, Georgia, 1986, pp. 265–274.
12. H.K. Charles, Jr. and G.V. Clatterbaugh, Thin film hybrids, in M.L. Minges, Ed., Electronic Materials Handbook, Vol. 1, Packaging, ASM International, Materials Park, Ohio, USA, 1989, pp. 313–331.
13. H.K. Charles, Jr., K.J. Mach, and R.L. Edwards, Multichip module (MCM) wirebonding, Proc. International Symposium on Electronic Packaging Technology (ISEPT '96), Shanghai, Peoples Republic of China, 1996, pp. 336–341.
14. H.K. Charles, Jr., K.J. Mach, R.L. Edwards, S.J. Lehtonen, and D.M. Lee, Wirebonding on various multichip module substrates and metallurgies, Proc. 47th Electronic Components and Technology Conference, San Jose, California, USA, 1997, pp. 670–675.
15. H.K. Charles, Jr., K.J. Mach, R.L. Edwards, A.S. Francomacaro, S.J. Lehtonen, and J.S. DeBoy, Wirebonding: reinventing the process for MCMs, Proc. International Symposium on Microelectronics, San Diego, California, USA, 1998, pp. 645–655.
16. H.K. Charles, Jr., K.J. Mach, R.L. Edwards, A.S. Francomacaro, S.J. Lehtonen, and J.S. DeBoy, Multichip module and chip-on-board wirebonding, Proc. 12th European Microelectronics Conf., Harrogate, Yorkshire, England, 1999, pp. 525–532.
17. H.K. Charles, Jr., K.J. Mach, R.L. Edwards, A.S. Francomacaro, J.S. DeBoy, and S.J. Lehtonen, High frequency wirebonding: its impact on bonding machine parameters and MCM substrate bondability, Proc. 34th International Microelectronics Symposium, Baltimore, Maryland, 2001, pp. 350–360.
18. H.K. Charles, Jr., K.J. Mach, S.J. Lehtonen, A.S. Francomacaro, J.S. DeBoy, and R.L. Edwards, High-frequency wirebonding: process and reliability implications, Proc. 52nd IEEE Electronic Components and Technology Conference, San Diego, California, 2002, pp. 881–890.
19. H.K. Charles, Jr., K.J. Mach, S.J. Lehtonen, A.S. Francomacaro, J.S. DeBoy, and R.L. Edwards, Wirebonding at high ultrasonic frequencies: reliability and process implications, Microelectronics Reliability, 43, pp. 141–153 (2003).
20. G.K.C. Chen, The role of micro-slip in ultrasonic bonding of microelectronic dimensions, Proc. 1972 International Microelectronic Symposium, Washington, DC, October 30–November 1, 1972, pp. 5-A-1-1–5-A-1-9.
21. T.B. Ching and W.H. Schroen, Bond pad structure reliability, 24th Annual Proc. Reliability Physics Symposium, Monterey, California, 1988, pp. 64–70.
22. G.V. Clatterbaugh, J.A. Weiner, and H.K. Charles, Jr., Gold-aluminum intermetallics, ball bond shear testing and thin film reaction couples, IEEE Trans. Components, Hybrids Manufacturing Technology, CHMT-7(4), pp. 349–356 (1984).
23. G.V. Clatterbaugh and H.K. Charles, Jr., The effect of high temperature intermetallic growth on ball shear induced cratering, IEEE Trans. Components, Hybrids and Manufacturing Technology, CHMT-13(4), pp. 167–175 (1990).
24. J.C. Demmin, Ultrasonic bonding tools for fine pitch, high reliability interconnects, Proc. Int. Conference on Multichip Modules, Denver, Colorado, USA, 1996, pp. 397–402.
25. V.J. Ehrlich and J.Y. Tsao, Laster direct writing for VLSI, VLSI Electronics: Microstructure Science, Vol. 7, Academic Press, 1983, pp. 129–164.
26. H.W. Endicott, H.K. James, and F. Nobel, Effects of gold-plating additives on semiconducting wire bonding, Plating and Surface Finishing, V, pp. 58–61 (1981).

27. K.L. Evans, T.T. Guthrie, and R.G. Hayes, Investigations of the effect of thallium on gold/aluminum wire bond reliability, Proc. ISTFA, Los Angeles, California, 1984, pp. 1–10.
28. B.L. Gehman, Bonding wire for microelectronic interconnections, IEEE Trans. Components Hybrids and Manufacturing Technology, CHMT-3(8), pp. 375–380 (1980).
29. L. Geppert, Solid state, IEEE Spectrum, 35(1), pp. 23–28 (1998).
30. A.B. Glaser and G.E. Subak-Sharpe, Integrated Engineering: Design Fabrication and Applications, Addison-Wesley, Reading, West Virginia, USA, 1979.
31. S. Goldfarb, Wirebonds on thick film conductors, Proc. 21st IEEE Electronics Components Conference, 1971, pp. 295–299.
32. B. Gonzalez, S. Knecht, and H. Handy, The effect of ultrasonic frequency on fine pitch Al wedge wirebonds, Proc. 46th Electronic Components and Technology Conference, Orlando, Florida, USA, 1996, pp. 1078–1087.
33. C.G. Gonzalez, R.A. Wessel, and S.A. Padlewski, Epoxy-based aqueous-processable photodielectric dry film and conductive via plug for PCB build-up and IC packaging, Proc. 48th Electronic Components and Technology Conference, Seattle, Washington, USA, 1998, pp. 138–143.
34. G.G. Harman, Wirebonding—towards 6σ yield and fine pitch, Proc. 42nd Electronic Components and Technology Conference, San Diego, California, USA, 1992, pp. 903–910.
35. G.G. Harman, Wire bonding to multichip modules and other soft substrates, Proc 1999 International Conference and Exhibition on Multichip Modules, Denver, Colorado, USA, 1995, pp. 292–301.
36. G.G. Harman, Wire Bonding in Microelectronics: Materials Processes, Reliability and Yield, McGraw-Hill, New York, USA, 1997.
37. G.G. Harman and C.A. Canon, The microelectronic wire bond pull test, how to use it, how to abuse it, IEEE Trans. Components, Hybrids and Manufacturing Technology, CHMT-1(3), pp. 203–210 (1978).
38. G. Heinen, R.J. Stierman, D. Edwards, and L. Nye, Wire bond over active circuits, Proc. 44th Electronic Components and Technology Conference (ECTC), Washington, DC, 1994, pp. 922–928.
39. J. Hirota, K. Machinda, T. Okuda, M. Shimotomai, and R. Kawanaka, The development of copper wire-bonding for plastic molded semiconductor packages, Proc. 35th IEEE Electronics Component Conference, Washington, DC, 1985, pp. 116–121.
40. C. Horsting, Purple plaque and gold purity, 10th Annual Proc. IRPS, Las Vegas, Nevada, 5–7 April 1972, pp. 155–158.
41. S. Ito, M. Kuwamura, S. Akizuki, K. Ikemura, T. Fukushima, and S. Sudo, Solid type cavity fill and underfill materials for new IC packaging applications, Proc. 45th IEEE Electronic Components and Technology Conference, Las Vegas, Nevada, USA, 1995.
42. V.P. Jaecklin, Room temperature ball bonding using high ultrasonic frequencies, Proc. Semicon: Test, Assembly and Packaging, Singapore, 1995, pp. 208–214.
43. J.L. Jellison, Effect of surface contamination on the thermocompression bondability of gold, IEEE Trans. Parts, Hybrids and Packaging, PHP-11, pp. 206–211 (1975).
44. J.L. Jellison and J.A. Wagner, Role of surface contaminates in the deformation welding of gold to thick and thin films, Proc. 29th IEEE Electronic Components Conferences, 1979, pp. 336–345.
45. C.N. Johnston, R.A. Susko, J.V. Siciliano, and R.J. Murcko, Temperature dependent wear-out mechanism for aluminum/copper wire bonds, Proc. International Microelectronics Symposium, Orlando, Florida, 1991, pp. 292–296.
46. J.S. Kilby, Invention of the integrated circuit, IEEE Trans. Electronic Devices, ED-23, pp. 648–654 (1976).
47. H.P. Klein, U. Durmutz, H. Pauthner, and H. Rohrich, Aluminum bond pad requirements for reliable wire bonds, Proc. IEEE Int. Symposium on Physics and Failure Analysis of ICs, Singapore, 1989, pp. 44–49.
48. J. Kurtz, D. Cousens, and M. Defour, Copper wire ball bonding, Proc. Int. Electronic Packaging Society Conference, New Orleans, Louisiana, USA, 1984, pp. 1–5.
49. B. Langenecker, Effects of ultrasound on deformation characteristics of metals, IEEE Transactions on Sonics and Ultrasonics, SU-13, pp. 1–8 (1966).
50. L.M. Levinson, C. Eichelberger, W. Wognarowski, and R.O. Carlson, High-density interconnect using laser lithography, Proc. International Symposium on Microelectronics, Seattle, Washington, October 17–19, 1988, pp. 301-306.
51. J. Ling and C.E. Albright, The influence of atmospheric contamination in copper to copper ultrasonic welding, Proc. 34th Electronic Components Conference, New Orleans, Louisiana, USA, 1984, pp. 209–218.
52. D. Liu, C. Zhang, J. Graves, and T. Kegresse, Laser direct-write (LDW) technology and its applications in low temperature co-fired ceramic (LTTC) electronics, Proc. 2003 International Symposium on Microelectronics, Boston, Massachusetts, Nov. 18–20, 2003, pp. 298–303.

53. G. Lo and Sitaraman, G-Helix: Lithography-based, wafer-level compliant chip-to-substrate interconnect, Proc. 54th Electronic Components and Technology Conference, Las Vegas, Nevada, June 1–4, 2004, pp. 320–325.
54. C. Meisser, Bonding techniques for plastic MCMs, Semiconductor International 14, pp. 120–124 (1991).
55. Microbonds, Inc., 151 Amber Street, Unit 1 Markham, Ontario, Canada L3R3B3, www.microbonds.com.
56. L.F. Miller, Controlled collapse reflow chip joining, IBM J. Res. Dev., 13, pp. 239–250 (1969).
57. G.E. Moore, VLSI: some fundamental challenges, IEEE Spectrum, 16(4), pp. 30–37 (1979).
58. J.L. Newsome, R.G. Oswald, and W.R. Rodregues de Miranda, Metallurgical aspects of aluminum wirebonds to gold metallization, 14th Annual Proceedings Reliability Physics, 1976, pp. 63–74.
59. H. Onoda, K. Itashimoto, and K. Touchi, Analysis of electromigration-induced failures on high temperature sputtered Al-alloy metallization, J. Vacuum Science Technology, A(13), pp. 1546–1555 (1995).
60. J. Onuki, M. Suwa, T. Iizuka, and S. Okikawa, Study of aluminum ball bonding for semiconductors, Proc. 34th Electronic Components Conference, New Orleans, Louisiana, USA, 1984, pp. 7–12.
61. K. Otsuka and T. Tamutsa, Ultrasonic wire bonding technology for custom LSIC with large number of pins, Proc. 31st IEEE Electronic Components Conference, Atlanta, Georgia, USA, 1981, pp. 350–355.
62. J.M. Petek and H.K. Charles, Jr., Known good die, die replacement (rework) and their influences on multichip module costs, Proc. IEEE 48th Electronic Components and Technology Conference, Seattle, Washington, USA, 1998, pp. 909–915.
63. E. Philofsky, Intermetallic formation in gold-aluminum systems, Solid State Electronics, 13(10), pp. 1391–1399 (1970).
64. J.B. Prather, S.D. Robertson, and J.W. Slemmons, Aluminum wire bonding to gold thick-film conductors, Electronic Packaging and Production, (May), pp. 68–71 (1974).
65. K.J. Puttlitz, An overview of flip chip replacement technology on MLC multichip modules, Proc. International Electronic Packaging Conference, 1991, pp. 909–928.
66. T.H. Ramsey and C. Alfaro, The effect of ultrasonic frequency on intermetallic reactivity of Au-Al bonds, Solid State Technology, 34, pp. 37–38 (1991).
67. K.V. Raui and E.M. Philofsky, Reliability improvement of wire bonds subjected to fatigue stresses, Proc. 10th IEEE Reliability Physics Symposium, Las Vegas, Nevada, USA, 1972, pp. 143–149.
68. J. Riddle, High cycle fatigue (ultrasonic) not corrosion in fine microelectronic bonding wire, Proc. 3rd ASM Conference on Electronics Packaging, Materials, Processes, and Corrosion in Microelectronics, Minneapolis, Minnesota, 1987, pp. 185–191.
69. B.M. Romensko, H.K. Charles, Jr., G.V. Clatterbaugh, and J.A. Weiner, Thick-film bondability: geometrical and morphological influences, The Int. J. for Hybrid Microelectronics, 8, pp. 408–419 (1985).
70. B.M. Romenesko, H.K. Charles, Jr., J.A. Cristion, and B.K. Sui, Gold-aluminum wirebond inteface testing using laser-induced ultrasonic energy, Proc. 50th Electronic Components and Technology Conference, Las Vegas, Nevada, 2000, pp. 706–710.
71. R.R. Schaller, Moore's law: past, present, and future, IEEE Spectrum, 34(6), pp. 53–59 (1997).
72. Y. Shirai, K. Otsuka, T. Araki, I. Seki, K. Kikuchi, N. Fujita, and T. Miwa, High reliability wire bonding technology by the 120 kHz frequency of ultrasonic, Proc. 1993 International Conference on Multichip Modules, Denver, Colorado, 1993, pp. 366–375.
73. T.H. Spencer, Thermocompression bond kinetics—the four principle variables, International Journal for Hybrid Microelectronics, 5(2), pp. 404–409 (1982).
74. T. Takahashi, E.W. Rutter, Jr., E.S. Moyer, R.F. Harris, D.C. Frye, V.L. St. Joor, and F.L. Oakes, A photodefinable benzocyclobutene resin for thin-film microelectronic applications, Proc. Int. Microelectronics Conference, Yokohama, Japan, 1992, pp. 64–70.
75. K. Takeda, M. Ohmasa, N. Kurosu, and J. Hosaka, Ultrasonic wirebonding using gold plated wire onto flexible printed circuit board, Proc. 1994 International Microelectronics Conference, Oamya, Japan, 1994, pp. 173–177.
76. A.A.O. Tay, K.S. Yeo, and J.H. Wu, The effect of wirebond geometry and die setting on wire sweep, IEEE Trans. on Components, Packaging and Manufacturing Technology, Part B, 18(1), pp. 201–209 (1995).
77. A. Thomas and H.M. Berg, Micro-corrosion of Al-Cu bonding pads, Proc. 23rd IEEE Reliability Physics Symposium, Orlando, Florida, USA, 1985, pp. 153–158.
78. J. Tsujino, T. Mori, and K. Hasegawa, Characteristics of ultrasonic wire bonding using high frequency and complex vibration systems, Proc. 25th Annual Ultrasonic Industry Association Meeting, Columbus, Ohio, 1994, pp. 17–18.
79. D.B. Tuckerman, D.J. Ashkenas, E. Schmidt, and C. Smith, Die attach and interconnection technology for hybrid WSI, 1986 Laser Pantography States Report UCAR-10195, Lawrence Livermore Laboratories, 1986.

80. R.R. Tummula, E.J. Rymazewski, and A.G. Klopfenstein, Microelectronics Packaging Handbook, Vols. I, II, & III, Chapman Hall, NY, USA, 1997.
81. S. Wakabayashi, A. Murata, and N. Wakobauashi, Effects of grain refinement in gold deposits on aluminum wire-bond reliability, Plating and Surface Finishing, V, pp. 63–68 (1981).
82. J.A. Weiner, G.V. Clatterbaugh, H.K. Charles, Jr., and B.M. Romenesko, Gold ball bond shear strengths effects of cleaning, metallization and bonding parameters, Proc. IEEE 33rd Electronic Components Conference, Orlando, Florida, USA, 1983, pp. 208–220.
83. The Welding Handbook, Vol. 2, eighth edition, Ultrasonic Welding, 1991, pp. 784–812.
84. Y.F. Yao, T.Y. Lin, and K.H. Chua, Improving the deflection of wirebonds in stacked chip scale packages CSP, Proc. 53rd Electronic Components and Technology Conference, New Orleans, LA, 2003, pp. 1359–1363.

4

Metallurgical Interconnections for Extreme High and Low Temperature Environments*

George G. Harman

Semiconductor Electronics Division, National Institute of Standards and Technology, Gaithersburg, MD 20899-8120, USA

Abstract The material properties and requirements for wire bond and flip chip interconnections that can be used in packaging chips for extreme high and low temperature environments [from +460°C (*HTE*) down to −200°C (*LTE*)] are described. The most commonly used Au–Al wire bonds should be avoided in the *HTE* range, along with any other metallurgical interfaces that form brittle intermetallics and/or Kirkendall voids. Gold–gold bonds improve with time and temperature. Thus, a clear preference is given for gold (or other noble metals) in the *HTE* environment for both wire and flip–chip bonds. For *LTE* and intermediate temperature ranges, such as on Mars and most earth satellites, conventional interconnections (Au and Al wire bonds) to Al chip metalization (bond pads) are acceptable. Also, normal flip-chip solder bumps are acceptable, but without plastic underfill. Information and techniques for using extreme temperature range materials, such as coefficient of thermal expansion (CTE) matching between chip and substrate, high temperature polymers, etc., are presented. Unusual failure mechanisms, such as possible electromigration of wire interconnections in *HTE*, are described. It is concluded that, with proper selection of materials, interconnections can be reliable in both extreme environments.

4.1. INTRODUCTION

The materials and requirements for wire bond and flip chip interconnections that can be used in packaging chips for extreme high and low temperature environments [from +460°C (*HTE*) down to −200°C (*LTE*)] are described. Devices capable of operating in these environments are needed for future space probes to other solar system planets, well-logging, geothermal measurements, sensors near rocket and jet engines, and, to a lesser

*This work is expanded from an invited presentation at the "Workshop on Extreme Environments Technologies for Space Exploration," Sponsored by the Jet Propulsion Laboratory in May, 2003. Pasadena, CA.

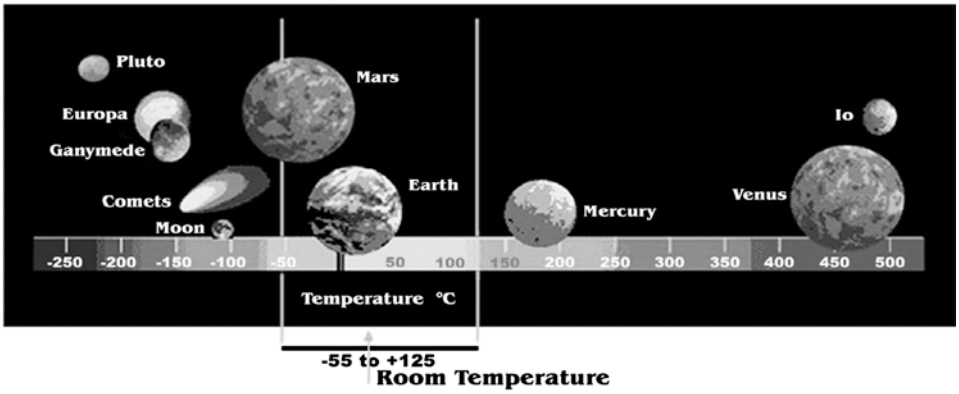


FIGURE 4.1. Pictorial representation of the Solar system planetary temperature ranges for which space probes must be designed in order to survive. [Note: The average nighttime temperature of I_o is about $-170°C$, but it has hot spots over 1000°C.] (Picture, courtesy J. Patel, NASA/JPL, as modified by R. Kirschman.)

extent, on and in internal combustion engines. The normal metallurgy used for interconnections as well as that on chips (wire bonds, flip chips, aluminum metalization) fails in *HTE*, and may have special requirements to be reliable in *LTE*. The required material changes, such as gold–gold interfaces, substrate/chip CTE matching, avoiding underfill, etc., are necessary for systems to survive in these thermal and changing temperature environments. Some of the concepts and materials below have been described in the literature [1]. For an overview of the Solar system planetary temperature ranges that may be encountered by future space probes, see Figure 4.1.

4.2. HIGH TEMPERATURE INTERCONNECTIONS REQUIREMENTS

4.2.1. Wire Bonding

The most commonly used Au–Al wire bonds have been investigated many times for possible high temperature use, with limited success. Such temperature exposure results in the formation of brittle intermetallic compounds and Kirkendall voids, both resulting in interconnection failure. The most successful approach to limiting such failure was to introduce hydrogen into the sealed package environment [2]. The result of this work is illustrated in Figure 4.2.

This approach, however, has not been used because of possible hermetic leakage of the hydrogen and occasional problems of H_2 effecting device performance. The quoted tests were only run for 1 h at 400°C. Therefore, more work should be performed to ascertain any advantages of H_2 in long term *HTE*. Teverovsky [3] studied the effects of high temperature degradation (in a vacuum for space application) of Au–Al bonds in plastic encapsulated devices. A mean lifetime of about 700 h at 225°C was obtained. This was better than in air, but not acceptable for most *HTE* usage. Wire manufacturers are currently developing doped Au bonding wires with $\geq$100 ppm of added (proprietary) stabilizing impurities (vs normal $<$10 ppm). Some have as much as 1%, usually Pd, added. Such wires demonstrate increased bond intermetallic lifetimes [4]. However, as above, preliminary data suggest that this will not be adequate for *HTE*. Thus, for the present, Au–Al interfaces should be avoided in the *HTE* range, along with any other metallurgical interfaces that

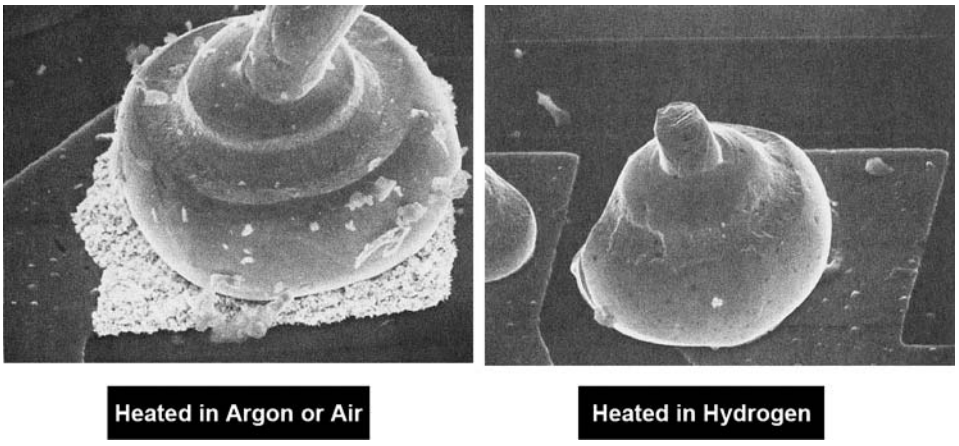


FIGURE 4.2. An example of using H_2 atmosphere to inhibit Au–Al intermetallic compound formation for Au–ball bonds on Al metalization (Heated 400°C for 1 h.) Left photograph bond(s), tested in air, N_2, or Ar (they are equivalent). Note the extensive intermetallic compound surrounding the ball bond. The right figure was tested in an H_2 ambient at 400°C for 1 h also. No intermetallic is evident. Both wires are 25 μm diameter [2].

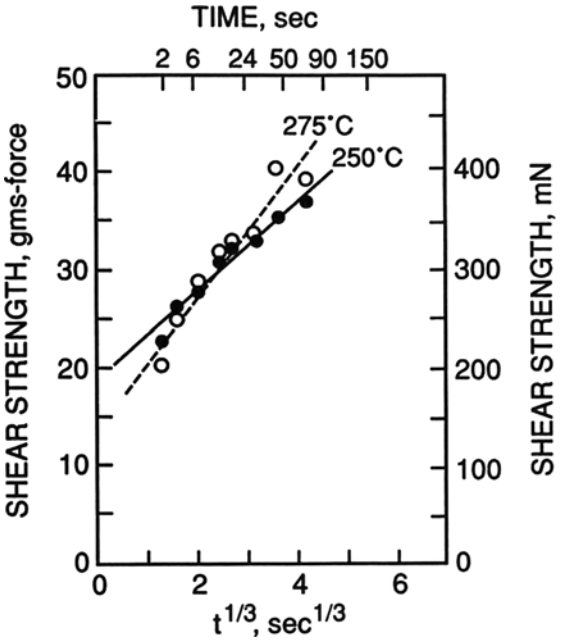


FIGURE 4.3. Improvement in the ball shear strength of organic-contaminated Au–Au wire bonded interfaces with time and temperature. Au–Au bonds are not degraded by high temperatures, as are Au–Al interfaces. After Jellison [5].

form brittle intermetallics and/or Kirkendall voids. In contrast to Au–Al interfaces, Au–Au bonds have long been known to improve with time and temperature [5], Figure 4.3.

Recently, Benoit et al. [6] have exposed Au–Au interfaces to 350°C for up to 300 h, without interface failure. Therefore, a clear preference is given for high melting point

monometallic welds of metals or ones that form solid solutions (e.g., gold to other noble metals, etc.) in the *HTE* environment for both wire and flip-chip bonds. Also consider the Al–Ni interface which has withstood equivalent high temperatures (Section 4.2.1.1) and should be useful, especially for large diameter Al wire bonds on Ni pads, as used on power devices. Tri-metallic diffusion barriers have long been used to prevent interdiffusion on hybrid packages. Examples might consist of a Mo, or Kovar base, which has a top layer of deposited Al and on the other side, a layer of Au. An Al wire would be bonded to the top (Al side), and the bottom would be thermocompression (TC) bonded (or welded) to thick film Au on the substrate. The Mo/Kovar serves as an effective diffusion barrier, preventing Au–Al intermetallic and voiding from forming. Traditionally, these have been called molytabs, but other names have been used, as well.

4.2.1.1. Effect of Wire Annealing in HTE The primary concern in *HTE* is that the bonded/welded interfaces remain strong, even if the wire is annealed and mechanically weaker. *HTE* will result in fully annealing small diameter Al wires, causing their breaking load, and to an extent the bond pull strength, to decrease to less than 30% of initial values. Such annealing does not cause a reliability problem in hermetic (open cavity) packages unless the heel of a wedge bond is overly thinned and subject to temperature cycling (see Section 4.2.1.3). Similar complete annealing occurred in billions of Al wire bonds to Al pads in CERDIP (ceramic-glass) sealed packages (processed at $\geq$400°C for 30 min) without reliability problems. Long periods at high temperatures may possibly change the structure of Al wire, weakening it. However, the weakest, the bamboo structure, has only been reported for wire bonds once in the literature [7] and may not occur in current production wires. (That structure is more associated with electromigration failure, see Section 4.2.1.4.)

Unless bond-lifts are experienced, the pull test is not a good indication of wire strength or reliability in *HTE*, since annealing increases the wire elongation during the test, resulting in some compensation for the breaking load decrease. (The pull test is described by a resolution of forces equation [1].) As the loop height increases due to elongation during the test, the force on the bond heel decreases. This is especially apparent in annealed Al wires which can elongate from 10% to 30%. Aluminum–aluminum or Al–Ni *bond interfaces* [6] will remain strong at high temperatures, (350°C for 300 h). Thus, any annealing of the wire (tensile strength decrease) is not considered a reliability problem (see CERDIP above). Gold wire for ball bonding is stabilized and annealed, and its strength and elongation will change less at high temperature than for Al. However, the strength of small diameter gold wires used for wedge bonding will anneal (soften), but less than that of equivalent Al wire. The many different proprietary dopants and quantities used in Au wire (e.g., Be, Ca, Pd, Pb, rare earths, etc. and in amounts ranging from 10 ppm up to 1000 ppm) make it difficult to predict the exact degree of softening that will occur at high temperatures, and such data will vary among manufacturers.

Large diameter Al wire ($\geq$100 μm) is normally annealed (to varying degrees) when produced, so its breaking load will decrease minimally in *HTE*. The interface reliability is the same as for small diameter wires/pads.

4.2.1.2. Other Noble Metal Wire Bond Characteristics for HTE Platinum [8] and Pd wires are occasionally used in *HTE* and *LTE* as well. They can be used in long term *HTE* (annealing has minimal effect on their strength, compared to Al). Both have high resistivity ($\sim$5$\times$ Au and Al) and low thermal conductivity ($\sim$25% of Au and Al). As such, both wires have been used to thermally isolate chips that may run much colder or hotter than their environment (e.g., superconducting devices—SQUIDS, etc.). However, both Pt and

Pd are harder than Au and Al. Thus, they require more ultrasonic energy (US) and force to bond according to the simplified formula [9], $E \propto (H)^{3/2}$, where E is the ultrasonic energy and H is the metal hardness. The nanohardness of these three metals is, Au = 1.77 GPa, Pt = 3.55 GPa, and Pd = 2.87 GPa. Thus, Pt takes ∼2.8 times and Pd ∼2 times as much ultrasonic power as Au to wedge bond (cold US welding), everything else being constant. Since US energy is a major cause of chip fracturing damage [1], both metal wires will tend to crater semiconductors during bonding, but are acceptable for bonding metalization on ceramic substrates. (SiC chips are as hard and have as high a fracture-toughness as most ceramics, and therefore should be safely bondable with these metals, see Table 4.1, below.) Also, high temperature thermocompression bonding (>300°C) to Si, GaAs, etc. can be achieved, without damage, in most cases. However, the bonding process may weaken the thin-film bond-pad adhesion to the chip, which may then be subject to delamination at high temperatures. This must be qualified for the anticipated temperature/bonding stress.

4.2.1.3. Wire Bond Fatigue in Temperature Cycling Environments Metallurgical fatigue damage to wires can occur during large ΔT temperature/power-cycling in both *HTE* and *LTE* (e.g., system [chip] temperatures on Mars' surface can range from −120°C to 85°C and on Jupiter, possibly from −140°C to >380°C). On Earth, well logging and various sensor applications can be cycled through wide temperature ranges as well. Only minimal and limited work has been done to determine the effects of high temperature cycling on 25 μm diameter Au and Al (1% Si) wire fatigue. Deyhim [10] studied fatigue (cycles-to-failure: C-F) on wires of both materials at temperatures up to 125°C, using various strain rates (0.7%, 5% and 10%). The first is considered more similar to temperature cycling of wires in hermetic packages. In that study, on one type of Au wire (unspecified dopants), the C-F decreased from ∼5000 25°C to ∼600 at 125°C. However, for Au wire made by a different manufacturer, the C-F was ∼30,000 at room temperature, and there was only a minimal decline at the high temperature ranges. No data were obtained at temperatures higher than 125°C in this test, and no test methods are available that are capable of measuring such wire fatigue at *HTE* temperatures. Benoit et al. [6] studied the fatigue of wires (at room temperature and low strain rates) after annealing at 300°C and found that all wires failed more rapidly after annealing for longer times. Thus, if a planetary probe (or other *HTE* application) will experience high temperature cycling, then any gold wires to be used must be qualified for fatigue resistance in the design cycle. This is essential for Au wires, since manufacturers frequently change their proprietary, stabilizing impurities. Fine Al wire has changed minimally in recent years, so older data should suffice.

Small diameter wires (e.g., 25 μm) are flexible and stress is mostly applied at the bond heel, and they typically break there during temperature cycling. A high loop height minimizes the stress on the heel and can prevent or limit such damage, (see Figure 4.4). A variety of factors (wire diameter, shape, loop height, metallurgy, and strain-rate) determine the fatigue susceptibility and life of a wire bond [1]. Also, any cracks or kinks in the heel/neck, or sharp bends in loops, such as can be made in worked-loop formation by modern autobonders, may increase fatigue damage. This has been minimally studied.

Large diameter Al wires are stiff, and stress is concentrated at all bends and at the weld attachments (bond heels); failures can occur at these positions. An example of large diameter Al wire fatigue is shown in Figure 4.5(a). Low bond deformation [see Figure 4.5(b)] and uniform looping gives the best fatigue protection for large diameter wires. Other possible solutions are to use Mg doped wire (0.5–1%) [7] and to make all bends smooth.

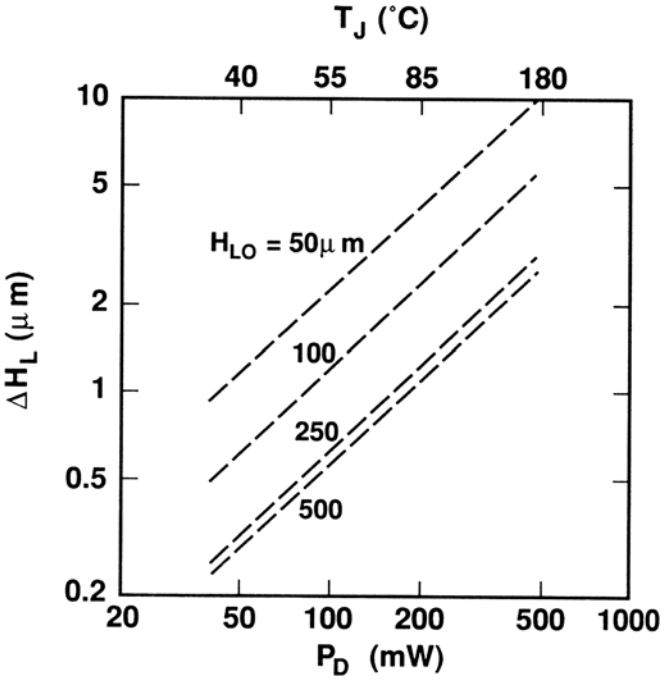
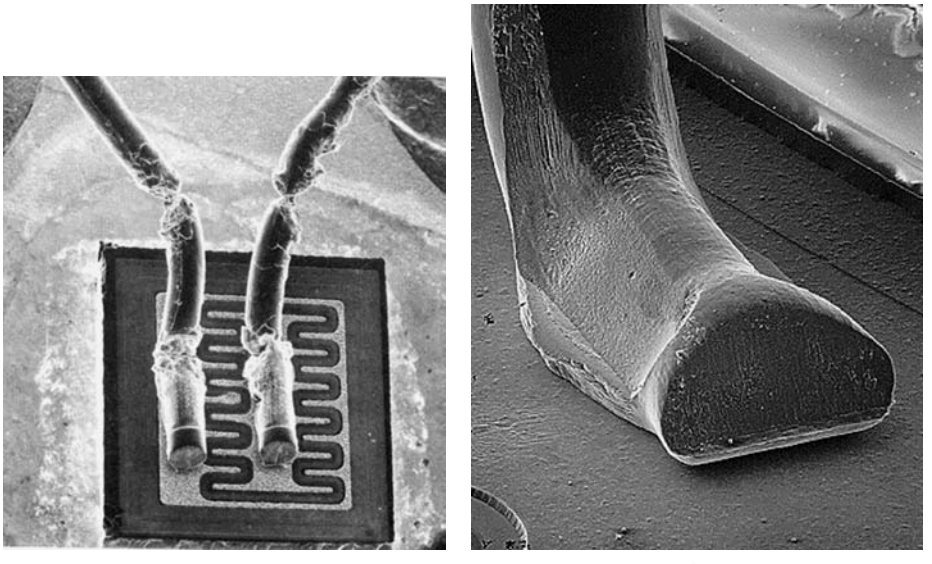


FIGURE 4.4. Calculation of wire bond flexing due to power cycling of a transistor. Top scale is the junction temperature with input power of P_D (mW). The maximum flexing of a semicircular wire bond loop ΔH_L is given on the Y axis. The as-made loop height, H_{LO}, is given by each curve. Minimum flexing occurs with highest loop height. (Analysis was made for a 25 μm diameter Al wire having a 1 mm bond to bond length.)



(a) (b)

FIGURE 4.5. (a) Fatigue of a large diameter (200 μm), 99.99% pure, Al bonding wire resulting from power cycling [1] from 25°C to 180°C. These underwent 18,606 temperature-power cycles from 25°C to 125°C in a telemetry application. (b) The desired bond shape, to minimize heel fatigue, of a 250 μm diameter Al wire (courtesy of Orthodyne Electronics).

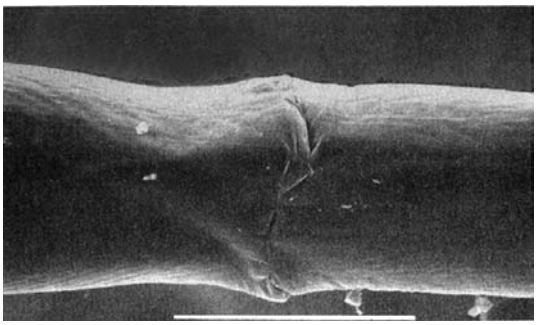


FIGURE 4.6. A SEM photograph of an electromigration-induced bamboo structure in a 25 μm Al wire with a "knuckle" perpendicular to the high current flow [13]. The bar at the bottom is ~25 μm.

4.2.1.4. Current Carrying Capacity of Conductors at High Temperature The burnout current for various wire diameters and compositions has been discussed in many papers, where measurements/calculations were obtained in room temperature ambients [11,12]. For *HTE*, but not for *LTE*, the current carrying capacity of these wires must be appropriately derated to avoid burnout at low current densities. Aluminum wire, with its 660°C melting point, must be derated more than Au (melting point 1064°C). Tse and Lach [13] found electromigration failures in bonded Al wires exposed to high current density and temperature over several years, and this mechanism could certainly contribute to *HTE* failures. As the ambient increases, this failure mode should increase according to Black's equation [14], ($t_{50} \propto J^{-n} e^{E/T}$, where t_{50} is the median lifetime, J is the current density, T is the absolute temperature, n is ~2, and E is the activation energy, which varies from 0.5 eV to 0.7 eV for Al, and ~0.8 eV to 0.9 eV for Au). Aluminum wires will fail more rapidly than Au, Pt, or Pd, by this mechanism. However, values of E for the latter two are not available at this time. Aluminum wires may also develop the weak bamboo structure, as in [13], where knuckle-like joints develop perpendicular to the current flow (see Figure 4.6). These would be very susceptible to ΔT fatigue (Section 4.2.1.3). The Author notes that electromigration is well documented in thin films; however, [13] is the only reported observation of electromigration in bonding wires, and this effect needs to be further studied. It also needs to be verified in contemporary Au bonding wire, since those are the most logical choice for interconnections in extreme *HTE*.

The final wire failure will result from the combination of increasing wire resistance due to the increasing resistivity with ambient (*HTE*) temperature, the wire self-heating due to I^2R heating, as well as any electromigration that may occur. This affects all of the discussed wires (Au, Pt, Pd, and Al), but Al, with its low melting point, will be the most affected. Thus the potential for wire failure will result in greatly reducing the rated current density of wires in both power and small-wire devices when used in *HTE*.

4.2.2. The Use of Flip Chips in HTE

Efforts have been made to evaluate traditional soft-solder bump flip chips for use in higher temperature environments [15]. Obviously, if the melting point of the solder is approached, this technology will not be viable. *(Hard-solder flip-chip bumps will strain and crack the chips and thus cannot be used.)* If underfill is required, due to chip-substrate CTE mismatch, this will also limit such use, since underfill is not usable above its glass

transition temperature, (T_g). The low temperature material properties of underfill are not available and thus may rule out its use in extreme low temperatures, *LTE*. Normal plastic substrates/boards, such as FR-4, BT, etc., cannot be used in *HTE* or *LTE*. New high temperature plastics may be used for boards in the future, but more development is needed (see Section 4.5 below). Gold ball bumps can be used in place of the normal solder-ball flip-chip interconnections. Chips can be ball-bumped or stud-bumped with gold and then thermocompression/thermosonic bonded to the gold metalization on ceramic substrates for reliable *HTE* flip chip interconnections, Figure 4.7.

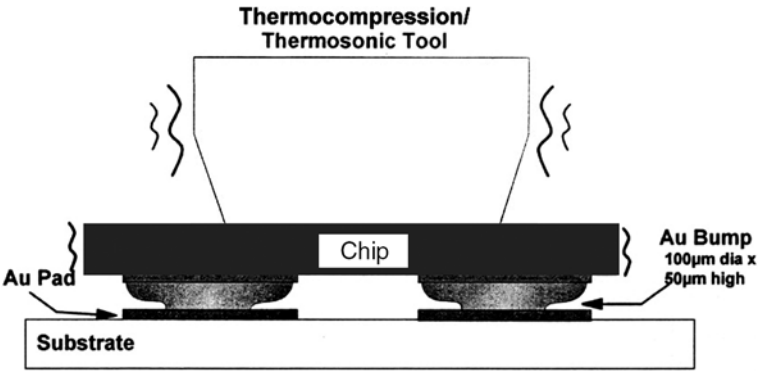


FIGURE 4.7. A gold ball-bumped flip chip that is being TC or TS bonded to gold metalization on a ceramic substrate. Note that these are usually used for low leadcount devices, which is typical of ones used for *HTE*. (Courtesy of Karl Puttlitz.)

TABLE 4.1.
Some material properties of chips and substrates for use in both *HTE* and *LTE*. Components should be chosen to minimize expansion differences in no-underfill flip-chips. Normal letters/numbers are approximate matches for equivalent chip and substrates and large, ***bold-underlined*** for the others. Generally, for flip chips, the substrate should have slightly higher CTE than the chip (which can heat up). But, for face-up die attach, a more exact match is desirable.

Component	Material	CTE (ppm/°C)	Thermal cond. (W/cm °C)	Fracture toughness MPa $m^{1/2}$
	Si	2.6	1.57	0.83–0.95
	SiC (β)	1–3 (T-dep.)	~5 (>T-dep.)	2.8
	GaN	~3	1.3	0.8
	GaAs	***5.7***	***0.48***	***~0.5***
Substrate	SiC	1–3	0.8–2	2.5–4.5
	AlN	4.6	1.75	2.8
	Si_3N_4	2.8–3.6	0.3–0.4	5.0–6.1
	Al_2O_3	***~6***	***0.35***	***3.1–3.3***
	BeO	***~6–7***	***~3***	***3.7***

[Note: Thermal-conductivity values can/will change at low temperatures, and exact data may not be available in the literature. Values can change for ceramic/polycrystalline substrates depending on temperature range, preparation, impurities, etc. and for single crystals, the orientation. Such values decrease in high temperatures.] Data from SRC-CINDAS and NIST Database. Values are rounded and should be used as a guide only. Obtain exact data from manufacturer or other specific measurements.

If extreme temperature cycling is expected, then a double or triple height ball-bumped structure can be applied at each bond pad to achieve higher chip standoff and minimize stress in the chip and substrate. For both extreme environments, substrate-and-chip CTE matching is essential for flip chips, if temperature cycling is expected (and it usually does occur). See Table 4.1 for some material property choices in designing packages for HTE flip chips. Components should be chosen to minimize the *expansion differences* in no-underfill flip-chips. (Ideally the substrate should have slightly higher CTE than the chip and substrate, since the chip is electrically heated.) However, for face-up chip die attach in *HTE*, one must also consider the thermal conductivity of both chip and substrate. High CTE substrates tend to equalize the temperature difference between the chip; thus a closer match in CTEs' is needed than for flip chip devices to be reliable in large temperature change environments as in Table 4.1.

4.2.3. General Overview of Metallurgical Interfaces for Both HTE and LTE

An overview of metallurgical interface (bonding) reliability for both *HTE* and *LTE* is shown in Figure 4.8. Dark lines indicate the interfaces appropriate for *HTE*. All interfaces, except for Al–Ag, are considered acceptable for *LTE* applications.

4.3. LOW TEMPERATURE ENVIRONMENT INTERCONNECTION REQUIREMENTS

Although a larger portion of this chapter is devoted to problems of interconnections in the *HTE* environment, most of the solar system is cold. This is readily seen in Figure 4.1. The problems of maintaining good chip and package interconnections is relatively simple when compared to those of *HTE.* (Other problems, such as battery performance, operation of some semiconductor devices, resistors, plastic delamination, etc., are of major impor-

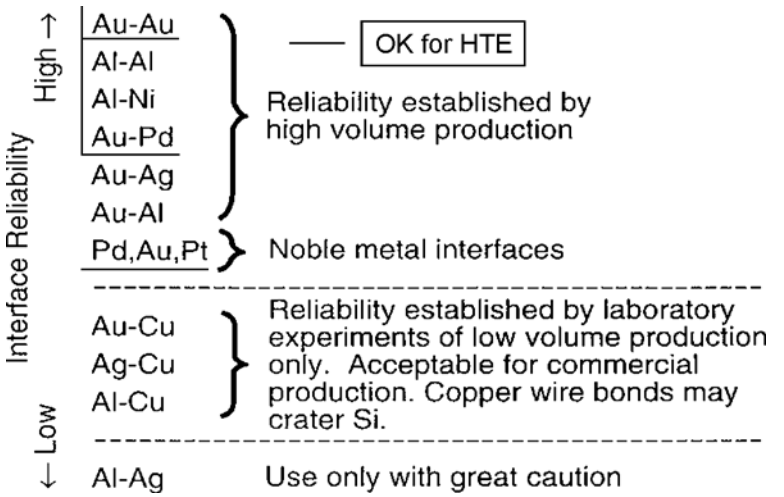


FIGURE 4.8. Metallurgical interfaces that can be used reliably in both extreme environments. All are acceptable for *LTE*, including soft solder based flip chips. However, only the dark underlined, boxed ones are acceptable for *HTE*.

tance but beyond the scope of this work.) Nevertheless, extreme low temperature interconnections do present challenges. Some of these are further discussed along with appropriate HTE problems or in combined sections such as in Section 4.4.

For use in *LTE* on the outer planets as well as for intermediate temperature ranges, such as on Mars, conventional Si or SiGe based devices can be used with normal Al metalization. Interconnections can be made with Au and Al wire bonds. Also, normal flip-chip solder bumps are acceptable on ceramic substrates, but without plastic underfill. The plastic-to-chip CTE mismatch at extreme low temperatures, could fracture the chip. (Future underfill development may improve this.) One can use normal epoxy glass-laminated substrates (e.g., BGAs), but ceramic is preferred. For both extreme environments, but especially for *HTE*, ceramic substrate-and-chip CTE matching is essential for flip chips, if temperature cycling is expected (and it usually does occur). See Table 4.1 for some material property choices for various chips with different ceramic substrates and packages.

4.4. CORROSION AND OTHER PROBLEMS IN BOTH *HTE*, AND *LTE*

There is no *liquid* H_2O at the temperatures of either extreme environment (Mars is intermediate, water condenses and materials may corrode); therefore, electro/chemical corrosion of interconnections and metalization is less likely. However, all such devices are built, qualified, and stored in normal Earth environments where water (moisture) and ionic impurities are plentiful and corrosive processes could initiate. Chips must be packaged using the best high-reliability, hermetic procedures. *HTE* chips would normally have gold/noble metalization, and, while electric-field-driven metal migration and high temperature gaseous attack is possible for these metals, it rarely occurs without *liquid* water.

Devices intended for *LTE* operation typically use normal aluminum-metalization on the chips and are subject to normal corrosion. These devices may be cycled through the liquid water (corrosion) range at various times during the device/system life. For Mars, the effective temperature range of semiconductor devices can vary from $-120°C$ to $85°C$, which includes self heating. Thus, appropriate low-moisture hermetic precautions are required. [*For comparison, Earth's south-polar-region temperatures range from approximately* $-80°C$ *to* $-20°C$ *and, in-situ temperature cycling would result in neither water based corrosion nor significant fatigue*.]

The packages for *HTE* will usually be made of metal/glass/ceramic (classical thick film hermetic hybrid packages). Reliability problems from hysteresis, creep, and/or cracking of normal glass-metal seals (Kovar-glass) in hermetic packages can cause failure during temperature cycling in both *HTE* and especially in *LTE* (see Kohl [16]). The glass-metal seals undergo expansion/contraction hysteresis resulting in cracking or delamination in the range starting about $-100°C$. Metal leads extending through *HTE*-softened glass seals will yield under stress in the *HTE* environment. Thus ceramic-metal seals should be used for both extreme environments.

Diffusion processes generally follow the Arrhenius equation ($K = Ae^{-E/RT}$, where K is the rate constant, E the activation energy, T is the absolute temperature, and R is the molar gas constant). Thus, diffusion will be greatly accelerated in *HTE*. Possible failure of metal adhesion layers between gold metalization and ceramic, as well as any diffusion barriers under/over the chip metalization must be considered as potential reliability hazards as well. However, the reverse is true in *LTE*, and various detrimental diffusion reactions such as Au–Al intermetallic compounds will not form or measurably increase in thickness.

Chips used for *HTE* will not be conventional Si based. Most likely they will be SiC or possibly GaP, GaAs, etc. Their metalization will most likely consist of complex layers with noble metals and diffusion barriers between the chip and the conductors. Interdiffusion of these layers can lead to adhesion failure [8] and degradation of the device as with the substrate metalization problems, above. The possibility of electromigration of the chip intraconnecting metalization as well as the wire bonds must be considered, as in Section 4.2.1.4. The die attach will have to be metallurgical rather than polymer/epoxy, as currently used on normal earth-bound devices. In some intermediate temperature ranges, silver-glass (max $T \sim 350°C$) [6] my be a satisfactory die attach material. For SiC chips, other possibilities could be Ni/Ti/$TaSi_2$, Ti/$TaSi_2$/Pt, Ni/Ti/Pt/Au, or Au. Currently, neither these nor any other die attach materials have been qualified for long term *HTE* at 460°C. Further investigations must be carried out before such missions take place.

4.5. THE POTENTIAL USE OF HIGH TEMPERATURE POLYMERS IN HTE

Some new high temperature polymers, such as Lo-k materials developed for advanced chips with copper metalization, may be considered for insulators and circuit boards in future extreme environments in cases where non-ceramic boards for interconnections are needed. (See Table 4.2, below.) Some of these materials can be used at temperatures above 400°C. Many of the problems of circuits on polymers, such as metal adhesion at high temperatures, have been solved by the semiconductor industry, but further development is required for specific *HTE* uses. Recently, the high temperature organic materials listed in Table 4.2 [17] have not been acceptable for semiconductor low-k incorporation because of processing or manufacturing compatibility reasons (e.g., SiLK, HOSP, FLARE). If these are available, however, they could still be useful in *HTE*. Recent developments continue in the field of Lo-k, and high temperature materials. For example, high modulus carbon substituted borazine polymers [18] may be appropriate for *HTE* circuit boards. Their temperature characteristics would be similar to the polymers in Table 4.2. Research in other

TABLE 4.2.
Low dielectric constant materials and the maximum operating temperature of ones possibly useful for *HTE* circuit boards or other *HTE* polymer/insulator uses. Weight losses at the indicated temperatures are available from manufacturers, who also supplied the quoted data (right hand column) (Ref. [17]).

High temperature, low dielectric constant, insulator materials						
Material	Max temp (°C)	Modulus (GPa) 25°C	Hardness (GPa)	Fracture-tough (MPa $m^{1/2}$)[c]	CTE[b] (10^{-6}/°C)	References/ sources
DVS-BCB	375	2.9	0.37	0.37	52	Dow Chem.
SiLK-H	450	2.45	0.31	0.6+ to 0.42	62	Dow Chem.
FLARE	>350	2.5	0.35	—	≈60	Honeywell (Allied Signal)
Parylene AF-4	450	2.28	—	—	30–80	Union Carbide

[a] Trade names are used to describe a material when no other identifier is available. This does not imply any endorsement.

[b] CTEs of organic LoK materials generally increase with temperature. Reported values are average and in the range of 25 EC to 100 EC.

[c] Fracture toughness of "Material" interface with SiO_2, SiN, Ta, or TaN.

—Development is dynamic and any product above may be improved or discontinued.

high temperature polymers (such as phthalonitriles, cyanate esters, and inorganic-organic hybrid polymers) for operation between 300°C and 510°C in oxidizing atmospheres, also show promise for *HTE* applications [19]. In addition, polyimidebenzoxazole [19] has been studied for such high temperature applications. However, much more development must be accomplished before these can be used.

4.6. CONCLUSIONS

This chapter presented possible materials as well as design considerations for chip interconnections/systems for extreme high temperature and/or low temperature environments. They are needed for future space-craft solar system exploration, well-logging, geothermal measurements, sensors near rocket and jet engines, etc. These cover the range from about −200°C to 460°C. Unusual problems, seldom encountered in normal environments can occur. Examples might be electromigration of interconnection wires and extreme temperature cycling induced fatigue. By using noble metal interconnections and ceramic circuit boards/substrates or possibly new high temperature polymers (still under development), these interconnection needs should be met. Other areas still needing specific development, such as die attach materials for *HTE*, are discussed. Except for these, most requirements are or will be achievable in the near future. For additional overviews of high and low temperature electronics materials, devices, and interconnections, see Kirschman [20] and McCluskey [21].

ACKNOWLEDGMENTS

The author acknowledges valuable discussions/information from R. Kirschman, Extreme-Temperature Electronics, J. Patel, NASA/JPL, and P. McCluskey, UMD. The paper could not have been written without support from NIST Office of Microelectronics Programs and the Semiconductor Electronics Division.

REFERENCES

1. G.G. Harman, Wire Bonding in Microelectronics, Materials, Processes, Reliability, and Yield, Second Edition, McGraw Hill (1997).
2. D.Y. Shih and P.J. Ficalora, The reduction of Au–Al intermetallic formation and electromigration in hydrogen environments, 16th Annual Proc. IEEE Reliability Physics Symposium, San Diego, California, 1978, pp. 268–272 (Figure 4.2, © IEEE, 1978).
3. A. Teverovsky, Effect of vacuum on high-temperature degradation of gold/aluminum wire bonds in pems, 42nd Annual Proc. IEEE International Reliability Physics Symposium, Phoenix, 2004, pp. 547–556.
4. C. Breach, F. Wulff, K. Dittmer, D.R. Calpito, M. Garnier, V. Boillot, and T.C. Wei, Reliability and failure analysis of gold ball bonds in fine and ultra-fine pitch applications, Proc. 2004 Semicon, Singapore, May 4–6, 2004, pp. 1–10.
5. J.L. Jellison, Kinetics of thermocompression bonding to organic contaminated gold surfaces, IEEE Trans. Parts, Hybrids, and Packaging, PHP-13, pp. 132–137 (1977) (Figure 4.3, © IEEE, 1977).
6. J. Benoit, S. Chen, R. Grzybowski, S. Lin, R. Jain, and P. McCluskey, Wire bond metallurgy for high temperature electronics, Proc. 4th Int'l High Temperature Electronics Conference, Albuquerque, NM, June 14–18, 1998, pp. 109–113.
7. K.V. Ravi and E. Philofsky, The structure and mechanical properties of fine diameter Al, 1-pct Si wire, Met. Trans., V2 (March), pp. 711–717 (1972). Also see same authors, Reliability improvement of wire bonds

subjected to fatigue stresses, 10th Annual Proceedings IEEE Reliability Physics Symposium, Las Vegas, Nevada, April 5–7, 1972, pp. 143–149.
8. J.S. Salmon, R.W. Johnson, and M. Palmer, Thick film hybrid packaging techniques for 500°C operation, Proceedings of the 4th International High Temperature Electronics Conference, Albuquerque, NM, June 16–19, 1998, pp. 103–108.
9. Simplified version of an empirical equation derived from US welding, Welding Handbook 8th Edition, Vol. 2, Am. Welding Soc., 1991. The constants of the full equation were developed for the US welding of thick materials, but the hardness relationship is indicative of the values observed in microelectronics.
10. A. Deyhim, B. Yost, M. Lii, and C.-Y. Li, Characterization of the fatigue properties of bonding wires, Proc. 1996 ECTC, Orlando FL, May 28–31, 1996, pp. 836–841.
11. E. Loh, Heat transfer of fine wire fuse, IEEE Trans. CHMT, V-7(Sept.), pp. 264–267 (1984).
12. A. Mertol, Estimation of aluminum and gold bond wire fusing current and fusing time, IEEE Trans. CPMT, Part B, V-18(Feb.), pp. 210–214 (1995).
13. P.K. Tse and T.M. Lach, Aluminum electromigration of 1-mil bond wire in octal inverter integrated circuits, Proc. 45th IEEE ECTC, Las Vegas, NV, 1995, pp. 900–905 (Figure 4.6, © IEEE, 1977).
14. J.R. Black, Electromigration—a brief survey and some recent results, 6th Annual Proc. IEEE Int. Reliability Physics Symposium, Dec. 1968, pp. 338–347.
15. T. Braun, K.F. Becker, M. Koch, V. Bader, R. Aschenbrenner, and H. Reichi, High temperature potential of flip chip assemblies, Intl. High Temperature Electronics Conference, Santa Fe, NM, May 17–20, 2004, pp. TP1-3. See also, Flip chip technology for high temperature automotive applications, 36th International Symposium on Microelectronics, Boston, MA, 2003, pp. 853–858.
16. W.H. Kohl, Materials Technology for Electron Tubes, Reinhold, N.Y., 1951.
17. G.G. Harman and C.E. Johnson, Wire bonding to advanced copper-low-K integrated circuits, the metal/dielectric stacks, and materials considerations, IEEE Trans. CPT, V-25(4), pp. 677–683 (2002).
18. I. Masami, S. Sekiyama, K. Nakamura, S. Shishiguchi, A. Matsuura, T. Takuya, Fukuda, H. Yanazawa, and Y. Uchimaru, Borazine-siloxane organic/inorganic hybrid polymer, Review of Advanced Material Science, V-5, pp. 392–397 (2003).
19. T.M. Keller, D.D. Dominguez, and M. Laskowski, High temperature polymers for geothermal and electronic packaging applications, Intl. High Temperature Electronics Conference, Santa Fe, NM, May 17–20, 2004, pp. TP2. Also see D.A. Dalman and F.F. Hoover, Eds., PBIO film dielectric for advanced microelectronics packaging, ibid, TA2.
20. R. Kirschman, Ed., High Temperature Electronics, IEEE Press, New York, NY, 1999. Also see ibid, Low-Temperature Electronics, 1986, These both present excellent overviews, as well as problems and solutions for extreme environment packaging.
21. F.P. McCluskey, R. Grzybowsky, and T. Podlesak, High Temperature Electronics, CRC Press, New York, 1997. A good recent overview and data on high temperature electronics.

5

Design, Process, and Reliability of Wafer Level Packaging

Zhuqing Zhang[b] and C.P. Wong[a]

[a]*School of Materials Science and Engineering, Georgia Institute of Technology, 771 Ferst Drive, Atlanta, GA 30332-0245, USA*
[b]*Hewlett-Packard Co., 1000 NE Circle Blvd., Corvallis, OR 97330, USA*

Abstract Wafer level packaging (WLP) has been growing continuously in electronics packaging due to its low cost in batch manufacturing and the potential of enabling wafer test and burn-in. A variety of wafer level packages have been devised, among which four important categories are identified including thin film redistribution and bumping, encapsulated package, compliant interconnect, and wafer level underfill. This chapter reviews the different WLP technologies with an emphasis on challenges and processes of the wafer level underfill. The wafer level packaging integrated with wafer burn-in, test and module assembly shows great attraction due to the dramatic cost reduction. Cost effective ways of building wafer level test and burn-in are under investigation.

5.1. INTRODUCTION

As a result of rapid advances in integrated circuit (IC) fabrication and the growing market for faster, lighter, smaller, yet less expensive electronic products, high performance low cost packaging is needed by the electronics industry. The conventional discrete IC packaging is inefficient, as such, a paradigm shift to wafer level packaging is apparent. Wafer level packaging (WLP) is a packaging technology where most or all of the IC packaging process steps are carried out at the wafer level. In the conventional discrete IC packaging process, the wafers are diced into individual IC chips first and then the chips are redistributed and packaged individually. In the WLP process, redistribution and packaging are performed at the wafer level. After wafer dicing, the individual components are ready to ship and assemble onto the substrates or printed wiring boards (PWBs) by the standard surface mount technology (SMT) process. A comparison of the conventional discrete packaging and the wafer level packaging is illustrated in Figure 5.1. The WLP makes possible 100% silicon efficiency (defined as the ratio of IC area over the entire IC package area) and low packaging cost due to the wafer level batch processing.

There are two major market drivers for wafer level package. In the cost-driven market, the wafer level Chip Scale Packaging (CSP) has the advantage that the cost per device

goes down as the wafer size increases and/or the IC size decreases [1]. The wafer level CSPs (WLCSPs) are mainly designed for small dies and low input/output (I/O) devices in consumer product market. Many of the technologies developed for these applications are based on simple peripheral pad redistribution followed by the solder ball attachment. These technologies are finding applications in low I/O counts functions, integrated passives and Rambus™ DRAMs. In this market, the WLP is a technology targeting at lower cost for packaging, and therefore the packaging cost per wafer and the number of chips per wafer (CPW) are the critical measures for success. The transfer to 300 mm wafer fabrication favors the WLP by increasing the number of CPW significantly.

The other market driver of WLP lies in large dies and high I/Os, high performance devices, for which flip-chip is currently the dominant first-level interconnect method. However, the major concern of the flip-chip is the solder joint reliability that is shortened by the thermo-mechanical shear stress due to the coefficient of thermal expansion (CTE) mismatch between the silicon chip and the organic substrate. The use of underfill increases the reliability of the packaging by stress redistribution, but it also increases the cost of flip-chip assembly due to the tedious dispensing and curing process. Wafer level packaging, through its design of stress buffering and/or compliance, promises to improve the reliability of the flip-chip without the additional underfilling steps in the assembly process. In both applications, wafer level packaging may enable wafer test and burn-in, resolving the known good die (KGD) issue, which will further reduce the cost of electronics manufacturing.

A variety of wafer level packages have been devised, among which four major categories are identified as follows:

- Thin film redistribution and bumping.
- Encapsulated package.
- Compliant interconnect.
- Wafer level underfill.

There often exists confusion between the concepts of flip-chip and wafer level packaging. WLP, by definition, requires no more packaging or encapsulation at the board level assembly. This means flip-chip on board (FCOB) can also be considered as a WLP. However, in most cases of FCOB, underfill is needed as mentioned in the above context. Some argued that flip-chip with underfill is not a WLP. Nevertheless, if underfill is moved onto the wafer level and no additional underfilling step is needed at board level assembly, flip-chip

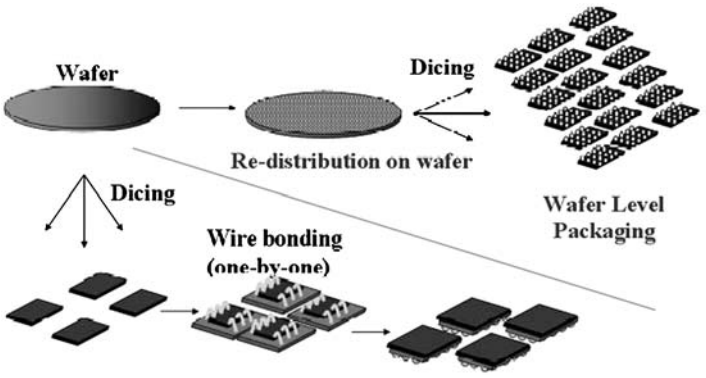


FIGURE 5.1. A comparison between the conventional packaging and the wafer level packaging.

with wafer level underfill is often considered as a type of WLP. In this chapter, different WLP technologies are reviewed and compared with an emphasis on the wafer level underfill. The first three categories of WLP would be discussed in Section 5.2 and wafer level underfill in Section 5.3. The recent development in wafer level test and burn-in is also reviewed.

5.2. WLCSP

5.2.1. Thin Film Redistribution

The thin film redistribution packages provide cost effective wafer level process and standard SMT compatible assembly, and are the major techniques used in the commercial WLCSPs. One example of this type of package is the Ultra CSP™ by K&S Flip Chip Division [2]. The manufacturing process of the Ultra CSP™ is illustrated in Figure 5.2. It utilizes two layers of benzocyclobutene (BCB) dielectric and one redistribution layer of Al/NiV/Cu. After the fabrication of the thin film layers, the solder balls are attached by flux, ball placement and reflow. One advantage of the Ultra CSP™ concept is that it uses standard IC processing technology for the package manufacturing. This makes the Ultra

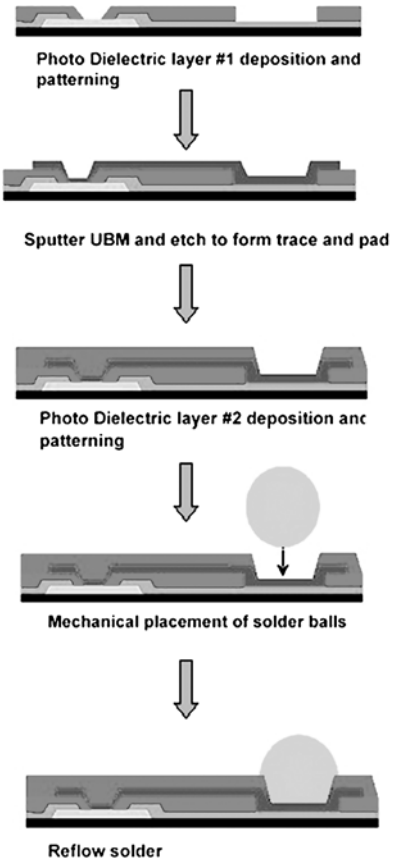


FIGURE 5.2. Process flow of Ultra CSP™.

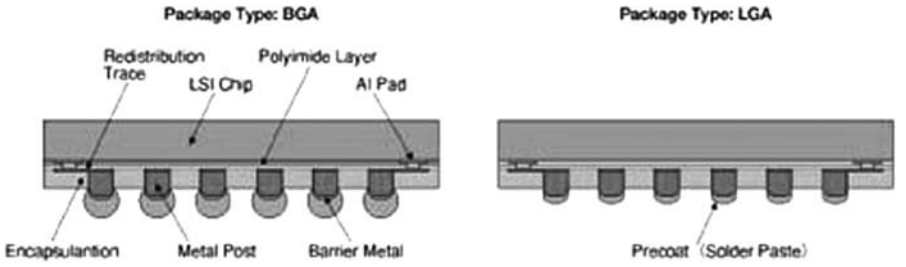


FIGURE 5.3. Structure of the Super CSP™.



FIGURE 5.4. Structure of FIP double bump wafer level package.

CSP™ ideal for both insertion at the end of the wafer fab as well as facilitation of wafer level test and burn-in options.

Typical products using the Ultra CSP™ are small packages with low number of I/Os. In order to increase the solder joint reliability of larger packages, a polymer reinforcement was designed and a technology called Polymer Collar WLP™ was developed by K&S Flip Chip Division [3]. In a standard WLP bumping process, a flux layer is usually applied before the solder ball placement to facilitate the solder wetting on the bond pads during wafer reflow. In the Polymer Collar WLP, a polymeric material is used instead of the flux and remains after the reflow to build reinforcement around the solder joint neck so as to block the shear deformation of the solder. As such, the reliability can be increased. In the Ultra CSP™ package with solders of maximum distant to the neutral point (DNP) being 3.18 mm, a 64% increase in cycle fatigue life-time was observed with the "polymer collar."

Another example of the thin film redistribution WLP is the Super CSP™ developed by Fujitsu Ltd. [4]. Figure 5.3 shows the structure of the Super CSP™ BGA (Ball Grid Array) and LGA (Land Grid Array) type packages. The manufacturing process of the Super CSP™ involves the formation of the redistribution layer by a polyimide film and electrolytic-plated metal trace. After redistribution, the resist is patterned and the copper posts are formed by electrolytic plating. Then the whole wafer is encapsulated with an epoxy molding compound (EMC), and solder balls or solder pastes are applied on top of the copper posts. The board level reliability of the Super CSP™ is good mainly due to high stand-offs of the copper post as well as the low CTE of the EMC encapsulation material which effectively reduces the stress occurring in the solder joint interconnect.

Similar to the Super CSP™, the Fab Integrated Packaging (FIP) invented by Fraunhofer IZM, Berlin uses a stress compensation layer (SCL) which embeds the solder balls before the second solder balls are attached on the top of embedded balls as shown in Figure 5.4. According to the thermal mechanical simulation, the SCL reduces the accumulated equivalent creep strain of the solder balls and also serves as mechanical support for the second solder ball to achieve taller solder heights compared to the standard redistribution

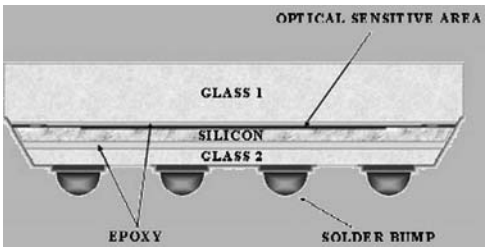


FIGURE 5.5. Structure of ShellOP CSP, an example of encapsulated WLP.

technology [5]. The double bump structure was evaluated by Motorola with different SCL materials [6].

5.2.2. Encapsulated Package

The ShellOP CSP by ShellCase is an example of the encapsulated wafer level packaging [7]. It sandwiches the silicon chip between two glass plates that prevent the silicon from being exposed and ensures mechanical protection as shown in Figure 5.5. The compliant polymer layer under the solder bumps provides board level reliability. This type of package is ideal for optical display.

5.2.3. Compliant Interconnect

For large die applications, many compliant interconnect technologies have been developed to improve the interconnect reliability. Several examples of the compliant interconnect are Microspring Contact on Silicon Technology (MOST) by FormFactor [8], Wide Area Vertical Expansion (WAVE) by Tessera, Sea of Leads (SoL) by Georgia Tech [9], G-Helix Interconnect by Georgia Tech [10], Elastic-Bump on Silicon Technology (ELASTec®) by Infineon [11], On-Wafer Floating Pad Technology by GE Global Research [12], Compliant Bump WLCSP by TI and Fujikura, etc. A common feature of the compliant interconnect is that the interconnect structure is designed to provide movement into x and y directions to accommodate the CTE mismatch during the thermal cycling. In most cases, z direction compliance is also provided to address the substrate coplanarity and wafer testing issues.

The MicroSpring™ technology was first invented for wafer probe cards and LGA production sockets. This technology was recently extended to a wafer level package called MOST™, in which the microspring contacts are fabricated directly on silicon at the wafer level. Figure 5.6 shows a picture of the MOST™ package. The microspring contacts are fabricated by gold wire bonding process and are plated with a Ni alloy, called "spring alloy." These contacts can be attached to the substrate through soldering. They decouple the CTE mismatch between the silicon die and the board. Hence, the reliability far exceeds any solder ball based technology. The microsprings require around one gram of compression force for every 25 micron of displacement and exhibit low contact resistance. They can be used as a fine pitch contact down to 225 microns, a similar pitch size compared with the current flip-chip production. The MOST™ technology integrated with wafer-level test and burn-in has been developed into a "Wafer on Wafer" (WOW) process as discussed later in the text.



FIGURE 5.6. A picture of the MOST™ package.

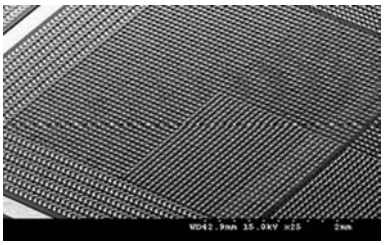


FIGURE 5.7. A picture of the SoL package.

As the semiconductor industry moves toward the development of giga-scale integration, the demand for high packaging density is increasing. Sea of Leads (SoL) by Georgia Tech is an example of the high-density wafer level packaging employing a compliant interconnect technology. SoL extends front end batch processing of the on-chip interconnect on the wafer to include x–y–z compliant chip I/Os through the fabrication of "slippery" leads and embedded air-gaps in the polymer film. A picture of the package is shown in Figure 5.7. There are several methods of allowing the lead movement during the thermal cycling. One method of fabricating the "slippery leads" uses a seed layer plated onto the leads that is selectively etched when the leads are ready to be released from the surface. The embedded air gaps are created through the decomposition of a patterned sacrificial polymer layer on the wafer. The density of the SoL package reaches $12 \times 10^3/\text{cm}^2$. The package supports high frequency signals up to 45 GHz. Similar interconnect structure can be found in the ELASTec by Infineon Technology. In the ELASTec package, the redistribution traces routed from the I/O pads are plated and form a spiral pattern on resilient silicon bumps. where an S-shaped metal layer was plated on a resilient bump made of silicone.

Another example of the high density compliant WLP is the G-Helix also designed by Georgia Tech. The G-Helix is a free-standing compliant interconnect fabricated by photolithography process. Figure 5.8 shows the pictures of the G-Helix Cu interconnect on a 200 μm pitch wafer. The advantage of the compliant interconnect lies in the design flexibility that can be optimized to offer the best mechanical and electrical properties. The drawback may be due to the high cost of the three mask-sets required for the fabrication.

The On-Wafer Floating Pad Technology by GE Global Research and the Compliant Bump WLCSP by TI/Fujikura shared the same characteristic of building the solder bumps on an array of polymer islands. Figure 5.9 shows the structure of Compliant Bump WLCSP. In this case, a resin post (polyimide core) is formed on the wafer to provide a compliant stand-off for the bumps. An encapsulant is molded over the structure to provide a protective

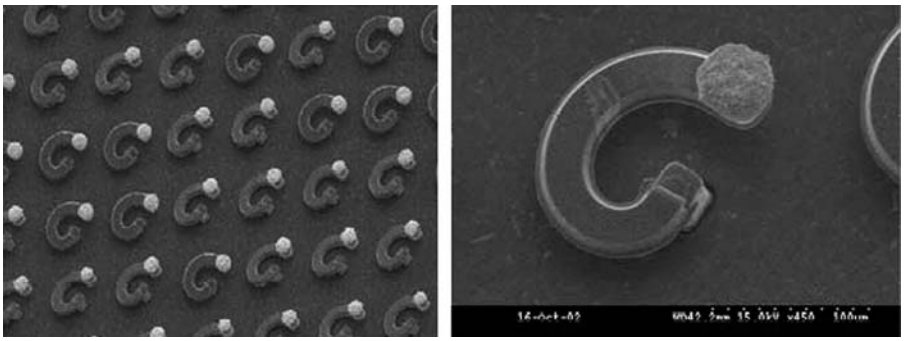


FIGURE 5.8. SEM pictures of G-Helix compliant interconnect WLP.

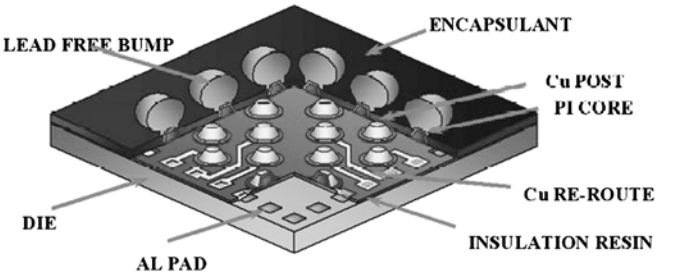


FIGURE 5.9. Structure of compliant bump WLCSP.

layer. The polymer core absorbs the strain between the mounted chip and PWB and can also provide some lateral movement when compressed.

The compliant interconnects usually build onto the thin film redistribution WLCSP, but have showed much superior thermal mechanical reliability, which intrigues enormous research effort from different companies. Nevertheless, the major drawback of the compliant interconnects is their high fabrication cost and the lack of infrastructure. The added inductance of some compliant interconnects also limits their application in high frequency devices. For these reasons, they are still at R&D level and the market is yet to mature.

5.3. WAFER LEVEL UNDERFILL

The wafer level underfill was initially proposed as a SMT compatible flip-chip process to achieve low cost and high reliability [13–16]. It can be used on WLSCPs as well to enhance their board level reliability (MicroFill by National Semiconductor, for instance). The schematic process steps are illustrated in Figure 5.10. In this process, the underfill is applied either onto a bumped wafer or a wafer without solder bumps, using a proper method, such as printing or coating. Then the underfill is B-staged and wafer is diced into single chips. In the case of unbumped wafer, the wafer is bumped before dicing when the underfill can be used as a mask. The individual chips are then placed onto the substrate by standard SMT assembly equipment.

It is noted that in some types of WLCSP, a polymeric layer is also used on the wafer scale to redistribute the I/O and/or to enhance the reliability. However, this polymeric layer

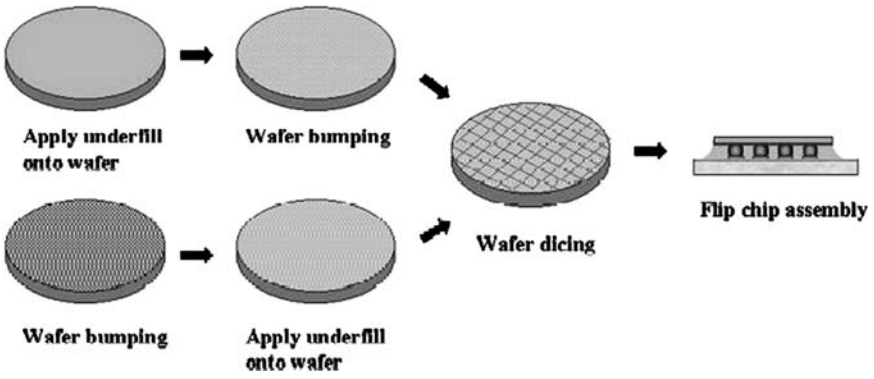


FIGURE 5.10. Process steps of wafer level underfill.

usually does not adhere to the substrate and cannot be considered as underfill. The wafer level underfill discussed here is an adhesive to glue chip and substrate together and functions as a stress-redistribution layer rather than a stress-buffering layer. The attraction of the wafer level underfill lies in the potential low cost (since it does not require a significant change in the wafer back-end process) and high reliability of the assembly enhanced with the underfill. Since this process suggests a convergence of front-end and back-end in package manufacturing, close cooperation between chip manufacturers, packaging companies, and material suppliers are required. Several cooperative research programs have been carried out in this area, including the team between Motorola, Auburn University, and Loctite Electronic Materials sponsored by National Institute of Standards and Technology Advanced Technology Program (NIST-ATP) [17], the team between National Semiconductor, IBM, National Starch and Chemical Company, and Georgia Institute of Technology sponsored also by NIST-ATP [18], and the team between 3M Company and Delpi-Delco Electronic Systems [19].

5.3.1. Challenges of Wafer Level Underfill

The material and process challenges for wafer level underfill have been identified and can be summarized as follows. First, a robust underfill deposition process is required; the resulted underfill layer must be of sufficient uniformity and consistency to enable a high yield in the assembly process, good solder joint formation and acceptable underfill fillet. Different deposition processes have been explored including spin coating, vacuum lamination, screen printing and stencil printing, etc. The underfill needs to be partially cured, or B-staged, if the original form is a liquid to facilitate the later handling including dicing and storage. One method is to use solvent in the deposition process and then drive off the solvent to B-stage the underfill. However, the use of unreactive solvent might leave residue which is likely to cause voiding during the later assembly [20]. B-stage cure can be used with careful control of the curing degree not to interfere with solder joint formation in the solder reflow. Wafer dicing presents another challenge for the underfill since the uncured material would be exposed to water that is used for cooling. If the wafer is to be diced with the underfill, the material also needs good mechanical property to prevent cracking. Unlike liquid underfill that is usually freeze-stored, wafer level underfill requires long shelf-life for packing, shipping and storage of the dies. Fortunately, B-staged material

usually has the glass transition temperature above room temperature, at which the mobility of the molecules is low to prevent large-scale reactions.

The issues related to the wafer level underfill in the assembly process start with the vision recognition at the placement machine. Normally, either fiducials or solder bumps on the die are located using the vision system in a pick-and-place equipment for flip-chip bonding alignment. Being covered by the underfill that is often heavily filled with silica fillers and hence translucent, these registration marks are difficult to be recognized by the vision system. Fortunately, many placement machines can adjust illumination angle, light intensity and image acceptance transforms, etc. to optimize imaging [21]. The coating color can also be adjusted to enhance the recognition. Some work has shown that black color provides the best contrast to the coated bumps [22]. If no additional flux is to be dispensed on the board, the wafer level underfill has to provide some tackiness to hold the chip in place. Several methods have been proposed including heating the board, heating the chip in a separate station, and heating the underfill through the pick-up nozzle. Similar to no-flow underfill, self-fluxing capability is required to eliminate the flux dispensing process. However, flux is known to degrade the stability of epoxy-based systems and shorten the shelf-life of the wafer level underfill. Hence, wafer-level underfill usually contains separating materials with different functions to achieve the desired result [23]. The solder wetting process with a wafer level underfill presents challenges to high interconnect yield, because the wetting is constrained by the presence of the partially cured underfill. Numerical simulation has been performed to predict the solder joint formation under constrained boundaries [24]. The solder joint interconnection is highly dependent on the fluxing capability and the viscosity of the underfill. However, it was found that the wetting process could be complicated by underfill outgassing and chip motion driven by forces other than surface tension of the solder [25]. The thickness of the underfill coating was critical for an optimal solder joint formation; deficiency in underfill could result in a gap between the bumps and excess underfill would hinder the solder joint formation. Other issues such as the desire for no post cure and reworkability are being addressed as well in the wafer level underfill process.

5.3.2. Examples of Wafer Level Underfill Process

In order to address the previous challenges, different wafer level underfill processes and the corresponding materials have been developed by various research teams, each providing unique solutions to the issues mentioned above. Illustrated in Figure 5.11 is the wafer scale applied reworkable fluxing underfill process developed by Motorola, Loctite and Auburn University [17]. Since uncured underfill materials are likely to absorb moisture that leads to potential voiding in the assembly, in this process, wafer is diced prior to underfill coating. Two dissimilar materials are applied; the flux layer coating by screen or stencil printing and the bulk underfill coating by a modified screen printing to keep the saw street clean. The separation of the flux from the bulk underfill material preserves the shelf life of the bulk underfill as well as prevents the deposition of fillers on top of the solder bump so as to ensure the solder joint interconnection in the flip-chip assembly. In this process, no additional flux dispensing on board is needed and hence the underfill needs to be tacky in the flip-chip bonding process to ensure the attachment of the chip to the board, as discussed previously.

Underfill deposition on wafer using liquid material via coating or printing requires subsequent B-staging, which is often tricky and problematic. The process developed by 3M and Delphi-Delco circumvents the B-stage step using film lamination [26]. The process

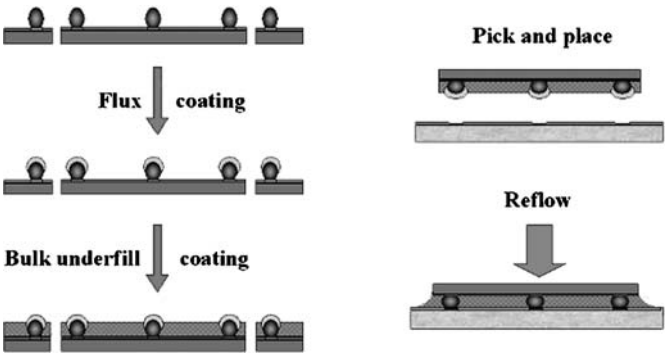


FIGURE 5.11. A wafer scale applied reworkable fluxing underfill process.

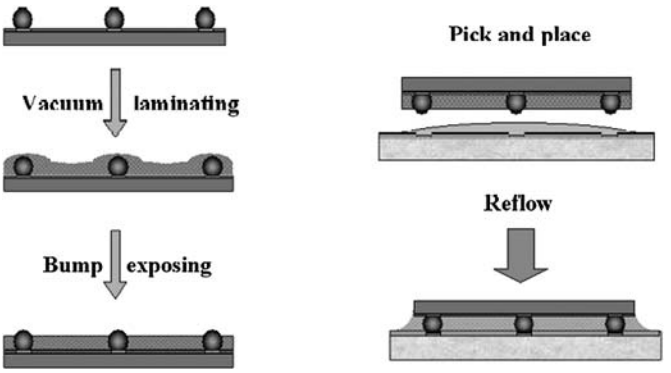


FIGURE 5.12. A wafer-applied underfill film laminating process.

steps are shown in Figure 5.12, in which the solid film comprised of thermoset/ thermoplastic composite is laminated onto the bumped wafer in vacuum. Heat is applied under vacuum to ensure the complete wetting of the film over the whole wafer and to exclude any voids. Then a proprietary process is carried out to expose the solder bump without altering the original solder shape. The subsequent flip-chip assembly is carried out with a curable polymeric flux adhesive pre-applied on the board.

Wafer level underfill can also be applied before the bumping process. Figure 5.13 shows a multi-layer wafer-scale underfill process developed by Aguila Technologies, Inc. [27]. The highly filled wafer level underfill is screen printed onto an unbumped wafer and then cured. Then this material is laser-ablated to form microvias that expose the bond pads. The vias are filled with solder paste and reflowed. Bumps are formed on top of the filled vias. The flip-chip assembly is similar to the previous approach with a polymer flux.

The wafer level underfill process has been successfully implemented in a commercial WLCSP MicroSMD by National Semiconductor [28]. Full area array at 200 micron pitch flip-chip assembly with wafer level underfill was also demonstrated by Georgia Tech [29].

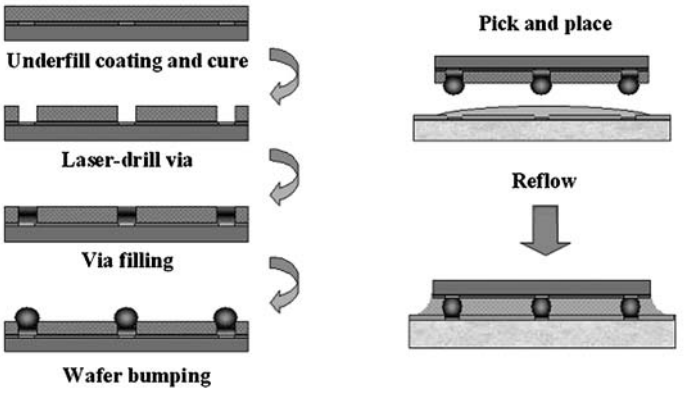


FIGURE 5.13. A multi-layer wafer-scale underfill process.

5.4. COMPARISON OF FLIP-CHIP AND WLCSP

Both flip-chip and WLCSP have the advantage of the actual package size being the same as the chip size. Miniaturization and low profile are the drivers of these two packages. As the interconnect/terminal is concerned, these two packages are similar in that most WLCSPs employ solder technology in an area array. In the assembly process, depending on the reliability requirement, WLCSPs are usually assembled in a standard SMT process without the need of underfill, while for flip-chip on organic substrate, underfill is usually required to ensure the reliability. The use of underfill substantially increases the assembly cost for flip-chip packages. In addition, the repair on the board level becomes difficult due to the fact that most underfill materials are thermosetting resins that are non-reworkable.

A major drawback of the flip-chip technology is the Known Good Die (KGD) issue. Test and burn-in for flip-chip packages are usually conducted at the component level. Wafer level package, on the other hand, has the potential of enabling wafer level test and burn-in, which would substantially reduce the cost and solve the KGD issue. However, due to the complexity of the IC wafer, a low cost and robust wafer level test and burn-in process is still under development.

The current market for WLCSP is low cost, low I/O, and small devices. The high performance devices mainly rely on flip-chip due to its high I/O capability and good electrical performance. However, flip-chip underfill technology is facing challenges as the I/O density increases and the pitch distance decreases. Convergence of the IC fab front-end and back-end of packaging manufacturing and potential cost reduction through wafer level test and burn-in drive the development of high density wafer level package. Compliant interconnects are attempts to bring the high density packaging to the wafer level. On the other hand, wafer level underfill is converging flip-chip technology with wafer level packaging, providing a low cost solution for flip-chip manufacturing.

5.5. WAFER LEVEL TEST AND BURN-IN

One of the attractions of wafer level packaging is the possibility of wafer level test and burn-in. Test (at speed) and burn-in is an expensive step in the semiconductor business. The transfer to full parallel test on wafer level could dramatically reduce the overall cost.

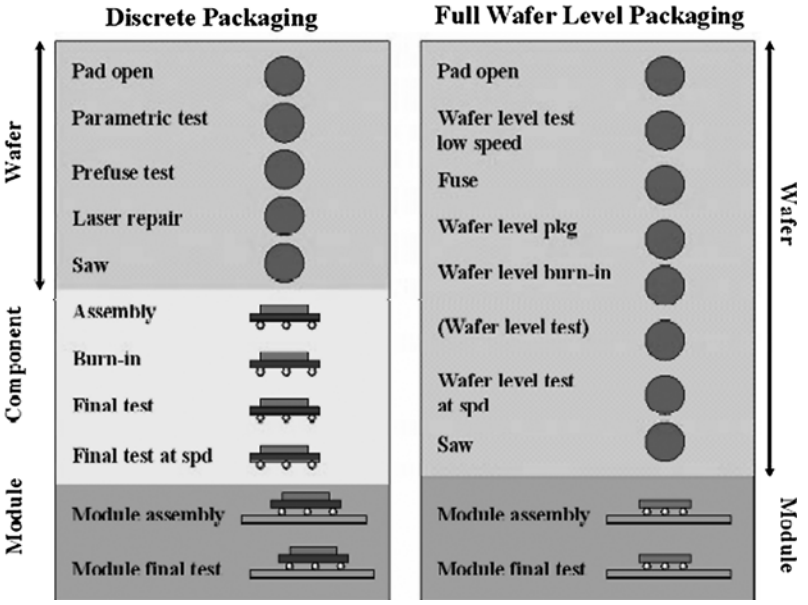


FIGURE 5.14. Assembly and test in discrete packaging and full wafer level packaging.

Estimation indicates that savings up to 50% for the transfer from component to wafer test. Figure 5.14 shows the assembly and test process flow for conventional discrete packaging and full wafer level packaging. The conventional process flow is shown in the left side in which the assembly and tests are done on the component level. On the right side, the integrated assembly and test process on wafer level is illustrated with test and burn-in performed on the wafer level.

However, several critical issues need to be addressed before the implementation of wafer level test and burn-in. The challenges for the probe card include high-density interconnects onto the wafer, CTE matching of the contactor to silicon, coplanar probe tips, high forces to make electrical connection with low resistance, uniform load to all the bumps, etc. In addition, precise alignment of probe to wafer is needed. Thermal management in wafer level burn-in is also critical. All the dies on the wafer should be subjected to a uniform stress; therefore the voltage, temperature, and the ramping rate need to be carefully controlled. Above all, the main barrier to the success of wafer level test and burn-in is the cost/performance ratio.

Several wafer level test and burn-in examples have been demonstrated in industry. Motorola announced the wafer level burn-in technology in 1998 with the partnership of Motorola Semiconductor Products Sector, Tokyo Electron Limited (TEL), and W.L. Gore and Associates, Inc. (GORE) [30]. The developed approach uses TEL wafer-probe technology in a controlled environment and allows each chip on a silicon wafer to be electrically stressed across a range of temperatures from 125°C to 150°C. Using this new technology, a silicon wafer of completed circuits is placed on a thermal chuck with an extremely flat surface. An electrical contact head, with thousands of contacts, is aligned to the wafer and contact is made through a sheet of contact material as shown in Figure 5.15. Critical to the process is the unique full-wafer contact material, called GoreMate™ wafer contactor, placed between the contact head and the test wafer. GORE also developed a thermally

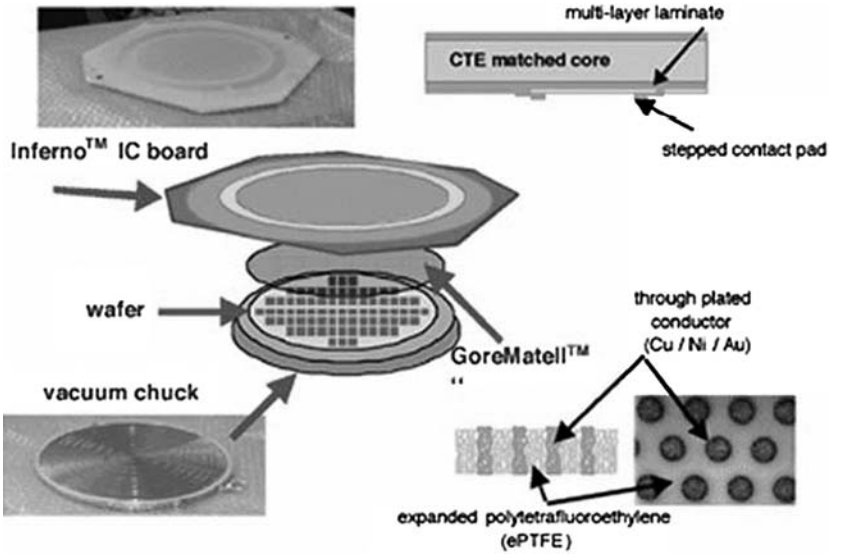


FIGURE 5.15. Motorola wafer level burn-in strategy.

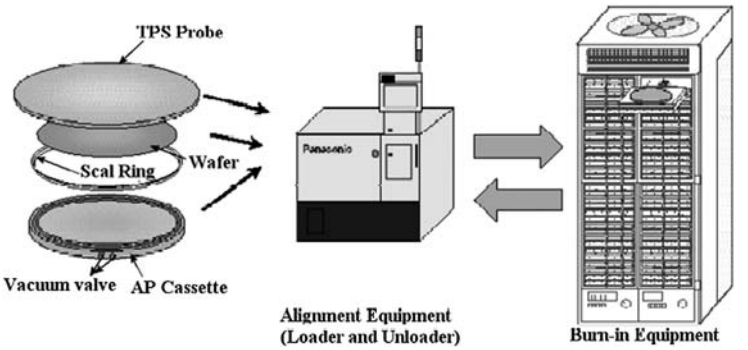


FIGURE 5.16. Matsushita wafer level burn-in overview.

matched (Inferno™) interconnect board, designed to have the same coefficient of expansion as silicon. As estimated from Motorola, through simplification and consolidation of product testing operations, manufacturing cost savings are expected to be as high as 15 percent and improvements in manufacturing cycle time will range up to 25 percent.

Matsushita Electric Industrial Co. Ltd. has also developed a wafer level burn-in strategy as shown in Figure 5.16. A three-part-structure (TPS) probe is used which consists of a glass substrate multilayer wiring board, a compliant z-axis conductor using conductive rubber, and a polyimide membrane with bumps for contacting [31]. The structure of the TPS probe is shown in Figure 5.17. A uniform contact force is provided by the atmosphere when vacuum is applied between the wafer and the TPS probe through the vacuum valve on the AP Cassette as shown in Figure 5.16. The conductive rubber acts to provide the absorption of the bump height differences. Firm contacts have been achieved on 2756 bumps which have remained stable up to 125°C.

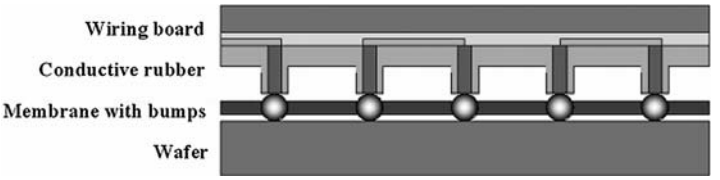


FIGURE 5.17. TPS probe structure of Matsushita wafer level burn-in.

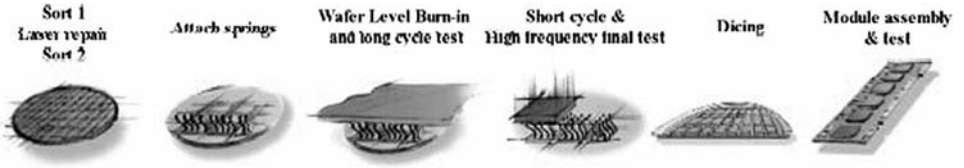


FIGURE 5.18. Process flow of WOW™.

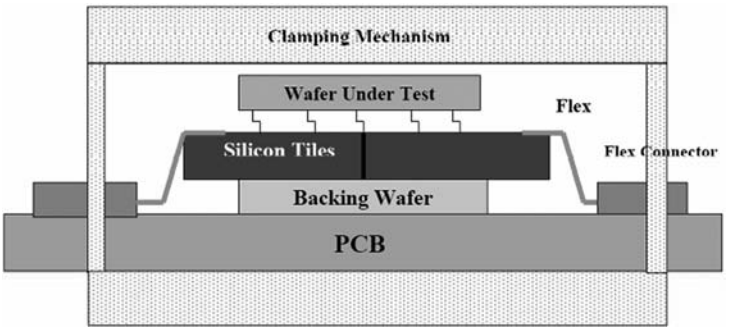


FIGURE 5.19. WOW™ wafer level burn-in structure.

Wafer level packaging, through the design of interconnect on wafer, is enabling full wafer test and burn-in to construct an integrated wafer level packaging, test and burn-in and assembly process to achieve ultimate low cost of electronic packaging. Many compliant interconnect techniques aim at providing flexible bumps that can be pressed down by a low force onto a flat contactor board. ELASTec WLP by Infineon has illustrated the benefit of the resilient bumps. Approximately 2 grams per bump are enough to form a reliable contact, taking into account the height tolerances of bumps and board pads [11]. The compliance of the interconnects also serves to solve the CTE mismatching problem of the wafer and the test board. A good example of WLP enabling wafer level test and burn-in is illustrated by the WOW™ (wafer on wafer) technology by FormFactor. WOW™ is IC industry's first back-end process that provides fully integrated wafer level package, burn-in, test and module assembly.

The microsprings of the MOST™ technology can provide the permanent interconnect onto the final product, as well as temporary connection under pressure during test and burn-in. These microsprings can be located anywhere on the die surface including directly on the bond pads. Figure 5.18 shows the process flow of integrated wafer level package, burn-in, test and assembly in WOW™. The wafer level burn-in structure can be seen in Figure 5.19. Silicon wafer is used for building the contactors due to the matched CTE with the

wafer under test and also the well-understood interconnect materials and process. However, it is challenging to build perfect yield wafer larger than 200 mm cost effectively. Therefore silicon tiles with smaller area are placed on and connected to a backing wafer. The test wafer is clamped against the contactors and tested from 25°C to 150°C. The test can also be carried out on a single die level and a multi die (module) level in addition to wafer level for different testing scenario. FormFactor's WOW™ process opens the door for vast business opportunities. However, cost effective wafer alignment and clamping systems, wafer temperature forcing systems, and wafer level test and burn-in electronics are to be sought.

5.6. SUMMARY

Wafer level packaging has been growing continuously in electronics packaging because its low cost in batch manufacturing and the potential of enabling wafer level test and burn-in. A variety of wafer level packages have been devised, among which four important categories are identified including thin film redistribution and bumping, encapsulated package, compliant interconnect, and wafer level underfill. The current WLCSPs mainly use thin film redistribution technology due to its low cost and are found in the consumer electronics market. Many compliant interconnect structures have been developed for large die applications. These compliant interconnects are designed to provide compliance in x, y, and z direction to accommodate the CTE mismatch between the chip and the substrate, and to address the substrate coplanarity as well. Wafer level underfill process suggests the convergence of flip-chip underfill and wafer level CSP, and may provide a low cost solution for high density WLP. However, the unique process of the wafer level underfill presents great challenges for the materials. Several wafer level underfill processes are reviewed and discussed. The wafer level packaging integrated with wafer burn-in, test and module assembly shows great attraction due to the dramatic cost reduction. Cost effective ways of building wafer level test and burn-in are under investigation.

REFERENCES

1. P. Garrou, Wafer level chip scale packaging: an overview, IEEE Transactions on Advanced Packaging, 23(2), pp. 198–205 (2000).
2. P. Elenius, S. Barrett, and T. Goodman, Ultra CSP™—a wafer level package, IEEE Transactions on Advanced Packaging, 23(2), p. 220 (2000).
3. D.H. Kim, P. Elenius, M. Johnson, S. Barrett, and M. Tanaka, Solder joint reliability of a polymer reinforced wafer level package, Proceedings of the 52nd Electronic Components and Technology Conference, 2002, p. 1347.
4. T. Kawahara, Super CSP™, IEEE Transactions on Advanced Packaging, 23(2), p. 215 (2000).
5. M. Topper, J. Auersperg, V. Glaw, K. Kaskoun, E. Prack, B. Beser, P. Coskina, D. Jager, D. Petter, O. Ehrmann, K. Samulewiez, C. Meinherz, S. Fehlberg, C. Karduck, and H. Reichl, Fab integrated packaging (FIP): a new concept for high reliability wafer-level chip size packaging, Proceedings of the 50th Electronic Components and Technology Conference, 2000, pp. 74–80.
6. B. Keser, E.R. Prack, and T. Fang, Evaluation of commercially available, thick, photosensitive films as a stress compensation layer for wafer level packaging, Proceedings of the 51st Electronic Components and Technology Conference, 2001, pp. 304–309.
7. A. Badihi, Ultrathin wafer level chip size package, IEEE Transactions on Advanced Packaging, 23(2), p. 212 (2000).
8. J. Novitsky and D. Pedersen, FormFactor introduces an integrated process for wafer-level packaging, burn-in test, and module level assembly, Proceedings of 1999 International Symposium on Advanced Packaging Materials, 1999, p. 226.

9. M.S. Bakir, H.A. Reed, P.A. Kohl, K.P. Martin, and J.D. Meindl, Sea of leads ultra high-density compliant wafer-level packaging technology, Proceedings of the 52nd Electronic Components and Technology Conference, 2002, p. 1087.
10. Q. Zhu, L. Ma, and S.K. Sitaraman, Design and optimization of a novel compliant off-chip interconnect—one-turn Helix, Proceedings of the 52nd Electronic Components and Technology Conference, 2002, p. 910.
11. H. Hedler, T. Meyer, W. Leiberg, and R. Irsigler, Bump wafer level packaging: a new packaging platform (not only) for memory products, Proceedings of 2003 International Symposium on Microelectronics, 2003, pp. 681–686.
12. R. Fillion, L. Meyer, K. Durocher, S. Rubinsztajin, D. Shaddock, and J. Wrigth, New wafer level structure for stress free area array solder attach, Proceedings of 2003 International Symposium on Microelectronics, 2003, pp. 678–692.
13. S.H. Shi, T. Yamashita, and C.P. Wong, Development of the wafer-level compressive-flow underfill process and its required materials, Proceedings of the 49th Electronic Components and Technology Conference, 1999, p. 961.
14. S.H. Shi, T. Yamashita, and C.P. Wong, Development of the wafer-level compressive-flow underfill encapsulant, IEEE Trans. on Components, Packaging, Manuf. Technol., Part C, 22(4), p. 274 (1999).
15. K. Gilleo and D. Blumel, Transforming flip chip into CSP with reworkable wafer-level underfill, Proceedings of the Pan Pacific Microelectronics Symposium, 1999, p. 159.
16. K. Gilleo, Flip chip with integrated flux, mask and underfill, W.O. Patent 99/56312, Nov. 4, 1999.
17. J. Qi, P. Kulkarni, N. Yala, J. Danvir, M. Chason, R.W. Johnson, R. Zhao, L. Crane, M. Konarski, E. Yaeger, A. Torres, R. Tishkoff, and P. Krug, Assembly of flip chips utilizing wafer applied underfill, Presented at IPC SMEMA Council APEX 2002, Proceedings of APEX, San Diego, CA, 2002, pp. S18-3-1–S18-3-7.
18. Q. Tong, B. Ma, E. Zhang, A. Savoca, L. Nguyen, C. Quentin, S. Lou, H. Li, L. Fan, and C.P. Wong, Recent advances on a wafer-level flip chip packaging process, Proceedings of the 50th Electronic Components and Technology Conference, 2000, pp. 101–106.
19. S. Charles, M. Kropp, R. Kinney, S. Hackett, R. Zenner, F.B. Li, R. Mader, P. Hogerton, A. Chaudhuri, F. Stepniak, and M. Walsh, Pre-applied underfill adhesives for flip chip attachment, IMAPS Proceedings, International Symposium on Microelectronics, Baltimore, MD, 2001, pp. 178–183.
20. B. Ma, Q.K. Tong, E. Zhang, S.H. Hong, and A. Savoca, Materials challenges for wafer-level flip chip packaging, Proceedings of the 50th Electronic Components and Technology Conference, 2000, pp. 171–174.
21. C.D. Johnson and D.F. Baldwin, Wafer scale packaging based on underfill applied at the wafer level for low-cost flip chip processing, Proceedings of the 49th Electronic Components and Technology Conference, 1999, p. 951.
22. M. Chason, J. Danvir, N. Yala, J. Qi, P. Neathway, K. Tojima, W. Johnson, P. Kulkarni, L. Crane, M. Konarski, and E. Yaeger, Development of wafer scale applied reworkable fluxing underfill for direct chip attach, Assembly Process Exhibition and Conference 2001, San Diego, CA, 2001.
23. L. Crane, M. Konarski, E. Yaeger, A. Torres, R. Tishkoff, P. Krug, S. Bauman, W. Johnson, P. Kulkanari, R. Zhao, M. Chason, J. Danvir, N. Yala, and J. Qi, Development of wafer scale applied reworkable fluxing underfill for direct chip attach, Part II, Presented at IPC SMEMA Council APEX 2002, Proceedings of APEX, San Diego, CA, 2002, pp. S36-2-1–S36-2-6.
24. L. Nguyen and H. Hguyen, Solder joint shape formation under constrained boundaries in wafer level underfill, Proceedings of the 50th Electronic Components and Technology Conference, 2000, pp. 1320–1325.
25. J. Lu, S.C. Busch, and D.F. Baldwin, Solder wetting in a wafer-level flip chip assembly, IEEE Transactions on Electronics Packaging Manufacturing, 24(3), pp. 154–159 (2001).
26. R.L.D. Zenner and B.S. Carpenter, Wafer-applied underfill film laminating, Proceedings of the 8th International Symposium on Advanced Packaging Materials, 2002, pp. 317–325.
27. R.V. Burress, M.A. Capote, Y.-J. Lee, H.A. Lenos, and J.F. Zamora, A practical, flip-chip multi-layer pre-encapsulation technology for wafer-scale underfill, Proceedings of the 51st Electronic Components and Technology Conference, 2001, pp. 777–781.
28. L. Nguyen, H. Nguyen, A. Negasi, Q. Tong, and S.H. Hong, Wafer level underfill—processing and reliability, Proc. 27th Int. Electron. Manuf. Tech. Symp., July 17–18, San Jose, CA, 2002.
29. Z. Zhang, Y. Sun, L. Fan, R. Doraiswami, and C.P. Wong, Development of wafer level underfill material and process, Proceedings of 5th Electronic Packaging Technology Conference, Singapore, Dec. 2003, pp. 194–198.
30. G.W. Flynn, Wafer level burn-in (WLBI) at Motorola—outline, Fleck Research Chip Scale International, 1999.
31. Y. Nakata and M. Kawai, A wafer-level burn-in technology, ULSI Process Technology Development Center, Matsushita Electronics Corporation.

6

Passive Alignment of Optical Fibers in V-grooves with Low Viscosity Epoxy Flow

S.W. Ricky Lee and C.C. Lo

Electronic Packaging Laboratory, Center for advanced Microsystems Packaging, Hong Kong University of Science & Technology, Clear Water Bay, Kowloon, Hong Kong

Abstract

Optical fibers are one of the most commonly used light transmitting media in optoelectronic systems for telecommunication applications. Because the core diameter of optical fibers is very small, active alignment methods are usually employed for the coupling between optical fibers and other optoelectronic devices. In general, the equipment cost of active alignment is very high and the processing time is relatively long, especially for fiber array alignment. Therefore, the conventional fiber alignment process becomes rather expensive and the throughput is quite low. In recent years, passive alignment using low cost epoxy adhesives and precisely etched V-grooves on silicon optical benches is attracting more attention due to its reduced production cost and short processing time. During the passive alignment process, the optical fiber may be lifted up by the buoyancy of epoxy flow and, hence, an extra cover plate is required to press the fiber against the walls of the V-groove. An effort is made to develop a modified passive alignment method without using the cover plate. Several parameters may affect the yield and need to be optimized. It is found that the amount of epoxy dispensed to the V-groove is critical in the process. Also the viscosity of the epoxy determines the characteristics of the flow in the V-groove and, hence, affects the results of passive alignment. In this chapter, the design and configuration of the modified passive alignment method will be introduced. The effect of the volume and viscosity of epoxy will be presented. The application to multiple fiber alignment will be demonstrated. The newly developed passive alignment method is capable of aligning an array of 8 fibers up to 1 micron accuracy.

6.1. INTRODUCTION

Alignment of optical fibers is very critical for optoelectronic packaging. A slight offset in any direction will affect the performance of the photonic devices. The tolerance of alignment is very tight, especially for single-mode optical fibers of which the core diameter is only 9 μm [1–7]. Active alignment method is commonly used in the industry since

the coupling efficiency is optimized. However, the processing time of active alignment is relatively long and the equipment cost is rather high [8–10].

Recently, passive alignment of optical fibers is attracting more attention due to its lower manufacturing cost and shorter processing time, compared with active alignment. Passive alignment is usually implemented on a silicon optical bench (SiOB) with V-grooves [11–17]. The position of optical fibers in passive alignment is defined by the geometry of the V-groove. The conventional method is to dispense the mounting epoxy a glob-top manner. However, the optical fiber may be lifted up due to the buoyancy. In order to avoid this problem, a cover plate is usually required to press the fiber against the wall of the V-groove [18–23]. Although the fiber is well aligned by pressing the cover plate, the applied stress may deform the optical fiber and affect the optical performance. If this pressing process is not well controlled, the optical fiber may be damaged and the reliability of the package will be decreased. Besides, in some applications, there may be not enough space for the mounting of the cover plate [24,25].

In this chapter, a new method of dispensing the epoxy with passive alignment capability will be introduced [26]. It is observed that if a suitable amount of epoxy is flowing to the gap between the optical fiber and the bottom of the V-groove, the optical fiber will not be lifted up by the buoyancy of the epoxy. On the other hand, the optical fiber will be pulled downward and sit against the walls of the V-groove. Based on this self-alignment property, a new design of V-grooves on the SiOB is developed. A "reservoir" is placed right next to the V-groove. The purpose of this reservoir is to let the epoxy flow into the gap between the optical fiber and the bottom of the V-groove. In order to obtain a steady flow with a gentle motion, the epoxy should have relatively low viscosity. The reservoir is patterned and etched together with the V-grooves so that there is no additional cost and time for fabricating the reservoir.

The proposed new passive alignment design has been characterized experimentally. The testing results show that, with the present approach, an optical fiber with an initial offset of 60 μm from its intended position will be aligned to the centre of the V-groove with a deviation less than 0.5 μm. Once the fiber is aligned, more epoxy is dispensed in a glob-top manner to enhance the fiber mounting strength and reliability. In this chapter, the design and fabrication of SiOB, the specifications of the epoxy, and the testing procedures and results will be presented in details.

6.2. DESIGN AND FABRICATION OF SILICON OPTICAL BENCH WITH V-GROOVES

In this section, the design and fabrication of SiOB with V-grooves for the passive alignment will be introduced. Figure 6.1 shows the mask design of SiOB. Two sizes of V-grooves are fabricated on the SiOB. They are used to hold the jacket and the cladding of the optical fibers. In addition to V-grooves, two newly designed features, "reservoir" and "canal," are also fabricated at the same time with the V-grooves. The reservoir and canal are only used in the modified passive alignment method illustrated in the later section. The design shown in Figure 6.1 is suitable for all the other experiments discussed throughout this chapter.

In this experiment, both low and high profile SiOB are fabricated. When the optical fiber is placed on the former SiOB, it is completely underneath the surface as the V-groove is deep enough to hold the cladding. On the other hand, part of the fiber is above the top

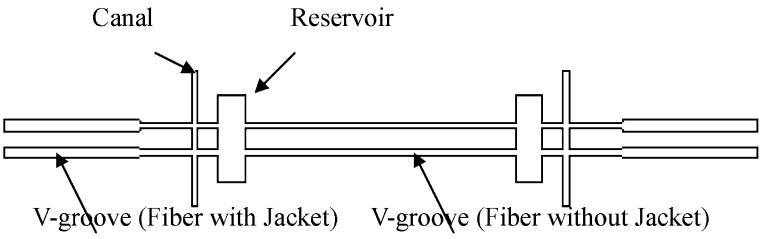


FIGURE 6.1. Mask layout of SiOB for passive alignment of optical fiber.

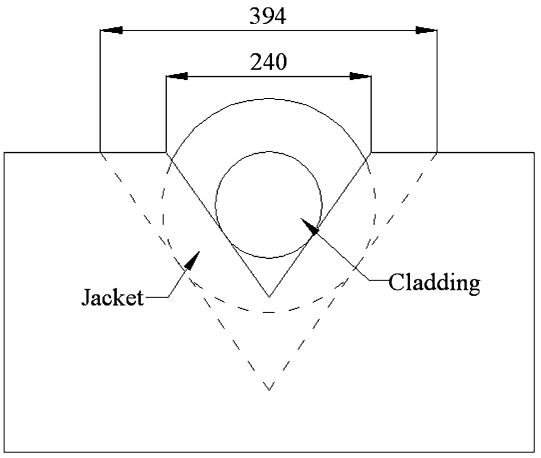


FIGURE 6.2. Low profile SiOB (all units are in micron).

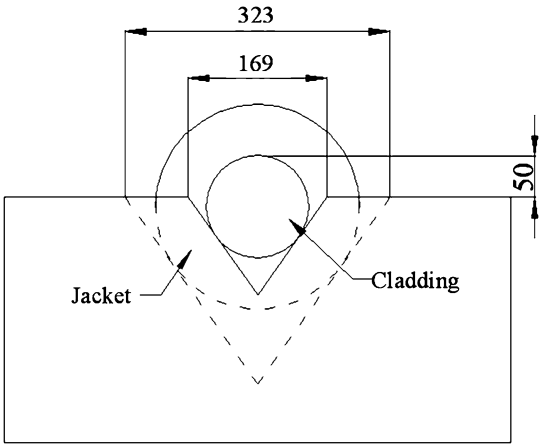


FIGURE 6.3. High profile SiOB (all units are in micron).

surface when it is placed on the latter kind of SiOB. The corresponding dimensions of the low and high profile SiOB are shown in Figures 6.2 and 6.3 respectively.

In the experiment, a 4-inch {1 0 0} silicon wafer with about 500 μm thickness is used. The wafer is cleaned before fabrication. After cleaning, a thin layer of low stress

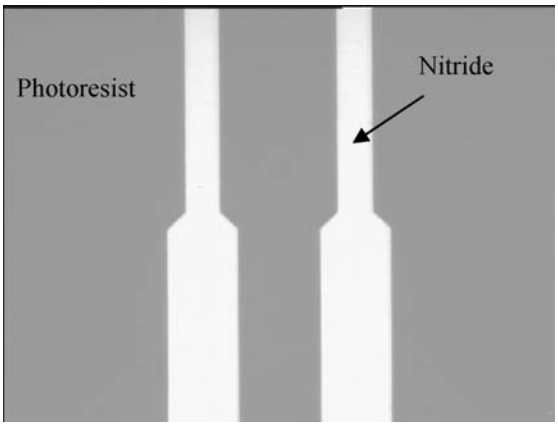


FIGURE 6.4. Patterning of low stress nitride by photoresist.

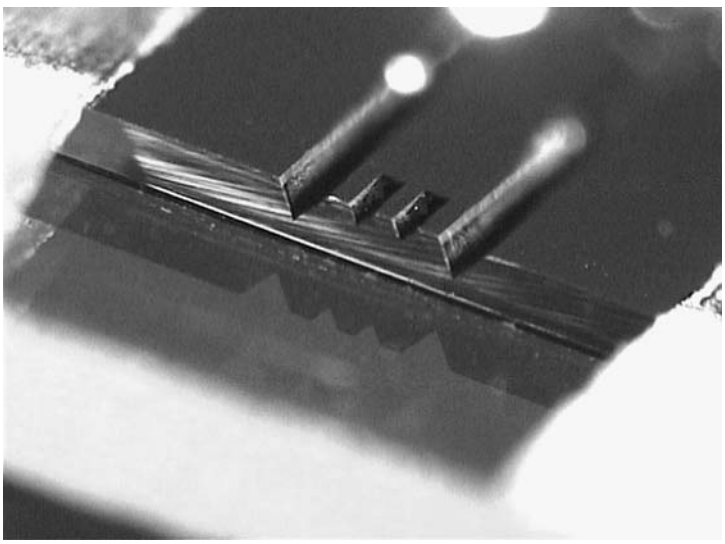
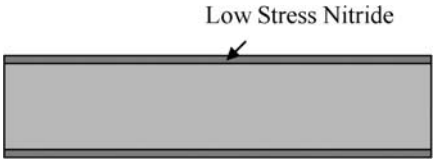
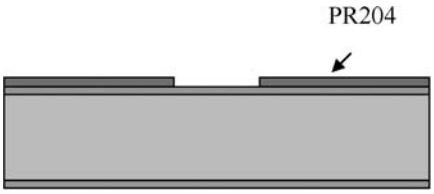
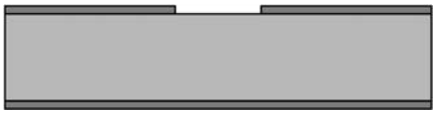
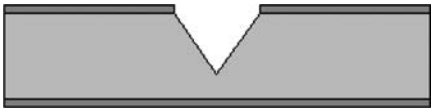


FIGURE 6.5. 3D view of V-grooves on silicon substrate.

nitride with 7000 Å thickness is then deposited on both sides of the wafer. This setup is critical because the wafer is then etched with strong alkaline at high temperature. A thin layer of low stress nitride is needed to act as the passivation layer. The nitride is patterned by photolithography process as shown in Figure 6.4. After hard baking the photoresist, the nitride which is not covered by the photoresist is then dry-etched away. The photoresist is stripped off after the dry etching process and the wafer is now ready for wet etching. The whole process is summarized in Table 6.1.

V-grooves formed by the {1 1 1} planes are obtained after the etching process. During wet etching, {1 1 1} planes are gradually formed [29–36]. When two {1 1 1} planes touch together and form the V-groove, the etching process stops. It is because the etch rate of {1 1 1} planes are the slowest among all crystal planes, and only the {1 1 1} planes are exposed to the solution. The depth of the V-groove depends on dimension of the window

TABLE 6.1.
Process flow of V-groove fabrication.

Process flow	Cross-section view
(1) Wafer cleaning – Dip in concentrated sulfuric acid for 10 minutes at 120°C – Dip in HF:H20 (1:50) solution for 1 minute at room temperature	
(2) Deposit 7000 Å low stress nitride on both side	
(3) Patterning – Spin coat photoresist PR204 on the wafer at 4000 rpm for 30 seconds – Soft bake at 110°C for 1 minute – Expose to UV light for 5 seconds – Develop by FDH-5 for 60 seconds – Hard bake at 120°C for 30 minutes	
(4) Passivation opening – Dry etch the nitride layer – Remove the photoresist by dipping the MS2001 solution at 70°C for 5 minutes	
(5) Silicon wet etching – Dip in 30% KOH at 85°C for 4 hours	

opening. Because of this, it is possible to have a design which has V-grooves with different depths on the same substrate by wet etching.

Figure 6.5 shows a 3D view of the V-groove obtained. Figure 6.6 shows V-grooves formed by {1 1 1} crystal planes. The cross-sectional view of V-grooves with different sizes and depths is clearly observed.

Figure 6.7 shows the cascaded V-grooves for holding the jacket and the cladding of the optical fiber fabricated on the same SiOB. The newly designed features, reservoir and canal are shown in Figure 6.8. It is observed that the reservoir and the canal are not rectangular in shape on designed the mask due to undercutting at the corner. However, this geometry is useful in the modified passive alignment method and will be discussed in following section.

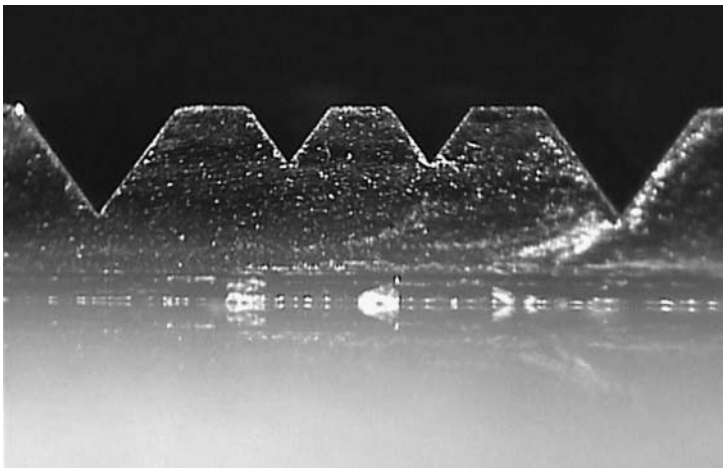


FIGURE 6.6. Cross-section view of V-grooves in different sizes.

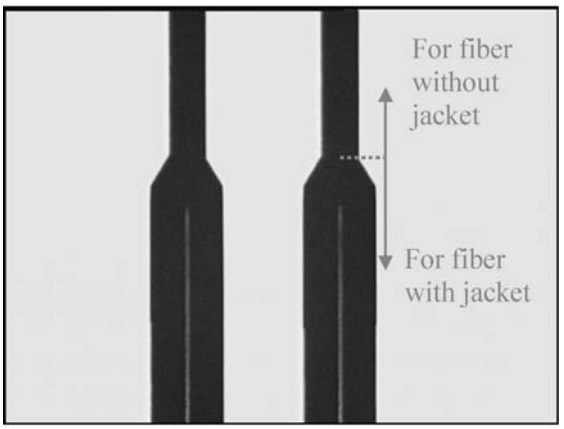


FIGURE 6.7. Cascaded V-grooves.

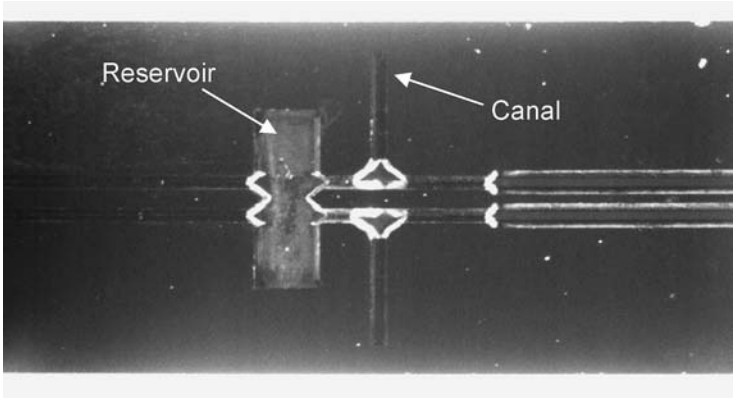


FIGURE 6.8. Reservoir and canal.

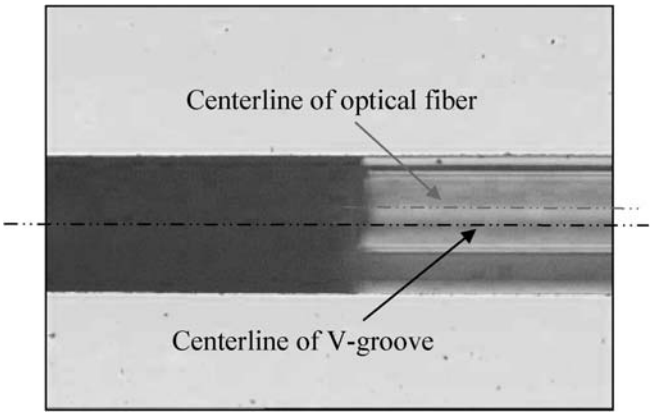


FIGURE 6.9. Initial misalignment.

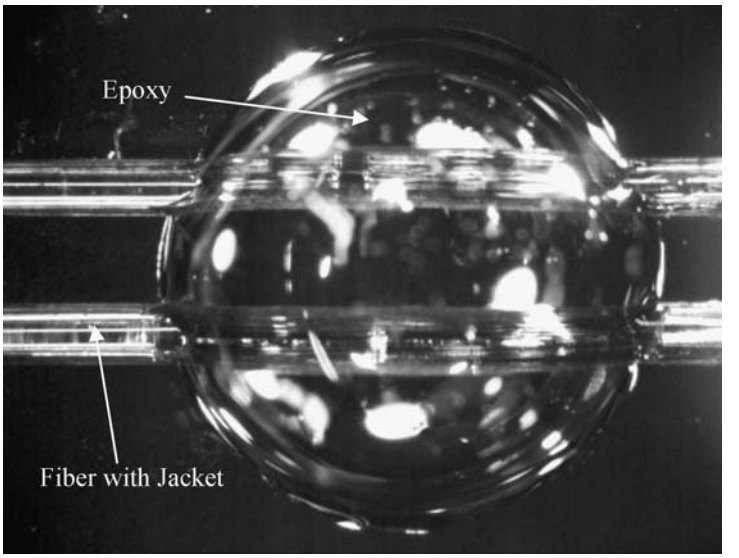


FIGURE 6.10. Fixing jacket by epoxy.

In the present study, the jacket of the optical fiber is stripped off at the beginning. The stripped length is dependent on the experiment performed. The cladding portion is dipped into the IPA solution for cleaning. Then, the fiber is placed onto the SiOB with V-grooves fabricated as mentioned earlier. In order to test the alignment properties of different passive alignment method, the fiber is intentionally misaligned in the V-groove as shown in Figure 6.9. Placement of the sample fiber is performed under an optical microscope.

Once the fiber is placed in the desired position, epoxy is dispensed onto the jacket in a glob-top manner. After curing the epoxy, the optical fiber can only move in transverse direction but not in axial direction as shown in Figure 6.10.

Two types of epoxies, Epoxy A and Epoxy B, are used in the experiment as listed in Table 6.2. Both are UV curable. Epoxy A is used to hold the jacket, as it has a higher vis-

TABLE 6.2.
Epoxies specifications.

	Epoxy A	Epoxy B
Viscosity	~14000 mPa s	~5000 mPa s
UV Cure	Yes, 365 nm UV light	Yes, 365 nm UV light
Heat Cure	Yes, 30 mins at 121°C	No

cosity. Therefore, it will not flow along with the V-groove rapidly and affect the experiment process at the room temperature.

6.3. ISSUES OF CONVENTIONAL PASSIVE ALIGNMENT METHODS

In this section, some conventional passive alignment methods and their drawbacks will be discussed.

6.3.1. V-grooves with Cover Plate

In order to prevent the buoyancy of the epoxy which may lift up the fiber and cause misalignment, a new procedure is added to the experiment by pressing the fiber with a cover plate. In this experiment, epoxy is first dispensed on top of the fiber. An additional cover plate, which is made by silicon wafer, is then placed on top. A small dead weight, around 2 g, is placed on the cover plate to provide a static force to press the fiber. Figure 6.11 shows the experimental setup. Epoxy A instead of Epoxy B is used, as the epoxy is in-between the SiOB and the cover plate. UV light cannot be used to cure the epoxy. The sample is placed in an oven at 121°C for 30 minutes for the curing process.

Figure 6.12 shows the cross-section inspection obtained by pressing the fiber with a cover plate on a low profile SiOB. The experiment shows that the fibers are not well aligned with the V-grooves even if a cover plate is placed on top. The whole optical fiber is underneath the top surface. This proves that, the cover plate cannot press the fiber effectively. Voids are also found in the cross-section inspection, which add an additional unfavouring result to the experiment.

In order to check whether a low profile SiOB is the factor leading to these unfavouring results, the experiment is repeated with a high profile SiOB. The experimental results are shown in Figure 6.13. From the figure, it is observed that both fibers are well aligned with the V-grooves when a high profile SiOB is used. The cover plate can effectively press the fibers down. However, voids are still found between the cover plate and the SiOB.

This experiment is repeated again by changing one experimental element. This time, only one fiber is placed on the SiOB. The cross-section inspection of this experiment is shown in Figure 6.14.

Figure 6.14 proves that the application of a cover plate is not suitable when there is only one fiber on the SiOB. The cover plate may be tilted when it is placed on the SiOB, because there is only one fiber to support the cover plate.

A die crack is found on the cover plate. It may be resulted from the tilted plate during the curing of epoxy. When cured, the volume of epoxy decreases due to shrinkage. The tilting of the plate leads to asymmetric distribution of epoxy between the gap. During the curing process, a larger epoxy volume change is experienced in the region with more epoxy.

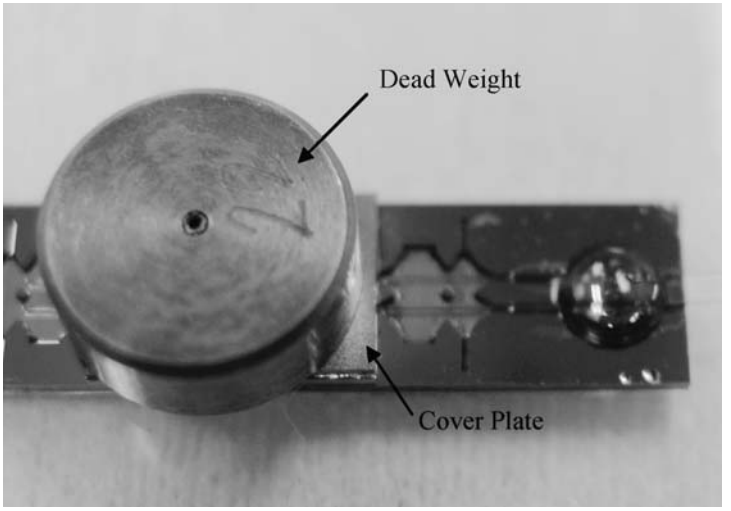


FIGURE 6.11. Application of cover plate.

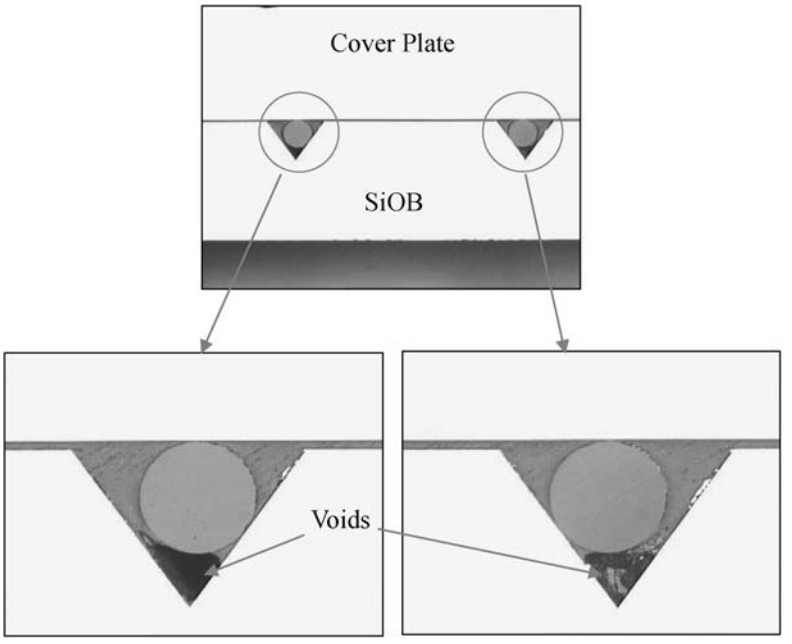


FIGURE 6.12. Cross-section inspection (with cover plate and low profile SiOB).

Therefore, a larger downward force is created on that region. The non-uniformly disturbed force will bend the plate, cracking the die when the force is large enough.

The epoxy shrinks not only during the curing process, but also due to a temperature drop. As the curing temperature of epoxy is 121°C, there is a hundred degree difference when compared with the room temperature. When the sample is removed from the oven,

epoxy starts to shrink immediately. The degree of shrinkage depends on the coefficient of thermal expansions of the epoxy. Based on the location of the die crack, it shown that the highest stress is developed at the tip of the optical fiber.

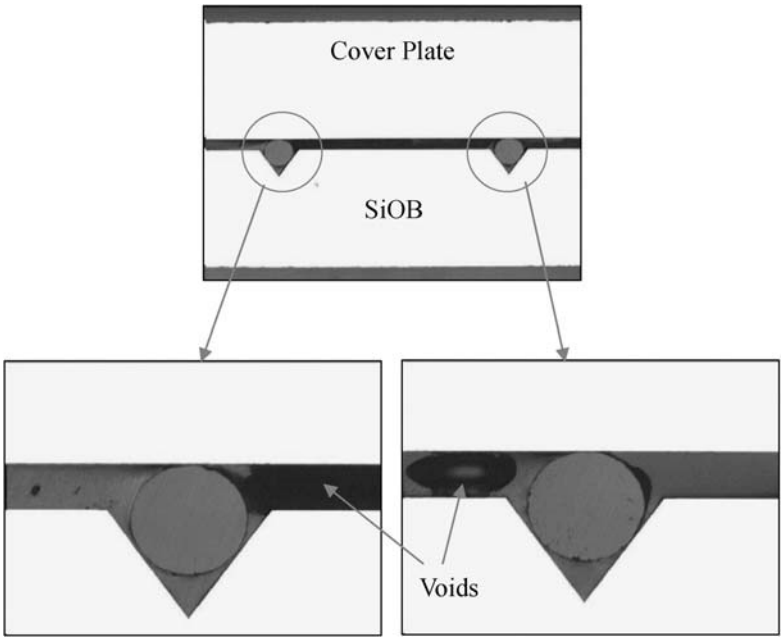


FIGURE 6.13. Cross-section inspection (with cover plate and high profile SiOB).

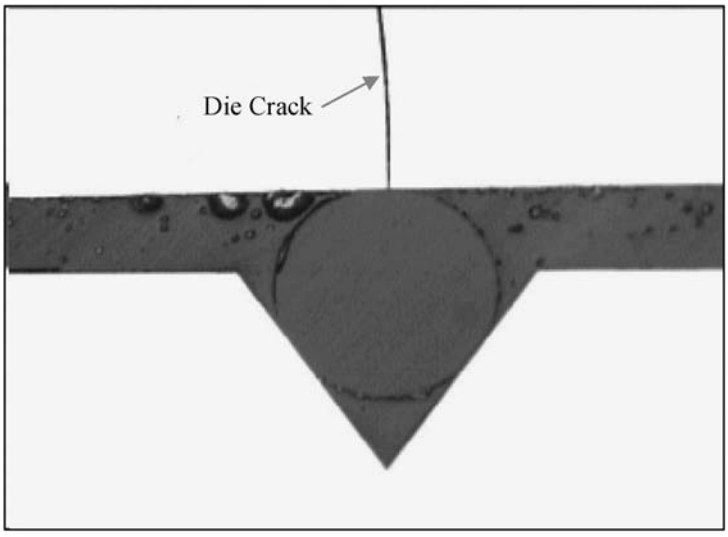


FIGURE 6.14. Cross-section inspection (with cover plate, high profile SiOB, single fiber).

6.3.2. Edge Dispensing of Epoxy

Besides glob top dispensing of epoxy and pressing the fiber with a cover plate, another passive alignment method, edge dispensing of epoxy is studied. A cover plate is used in this method. However, unlike the methods evaluated earlier, this time the cover plate is put on top of the fiber first. Epoxy is then dispensed at the edge of the cover plate. Again, Epoxy A is used because thermal cure is needed. Epoxy will flow into the gap as its viscosity decreases at high temperature.

Figure 6.15 shows the experimental setup and the epoxy dispensing direction. The way epoxy is dispensed is similar to the one used in dispensing underfill in flip chip technology. Figure 6.16 shows the results when the epoxy is dispensed at the edge.

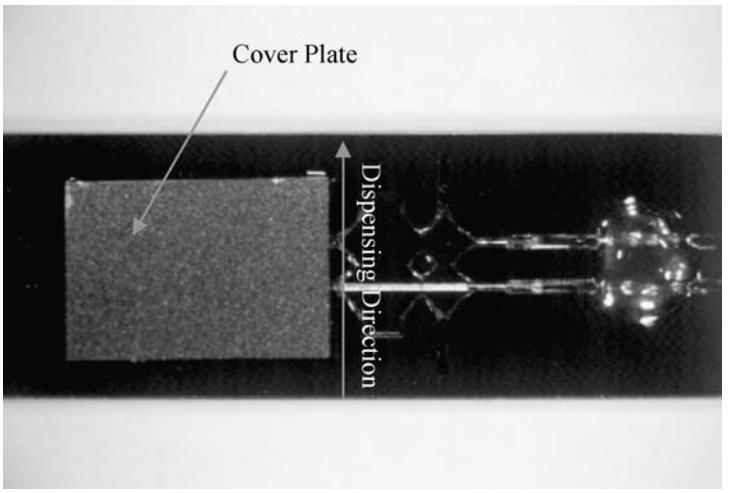


FIGURE 6.15. Edge dispensing method.

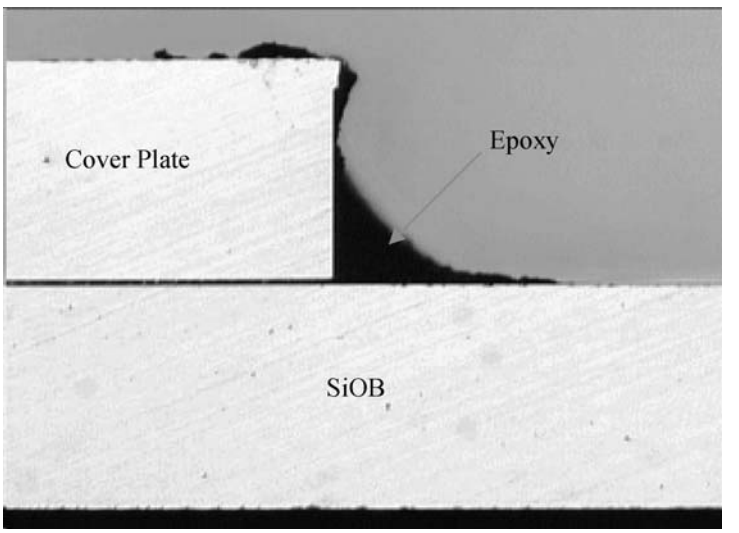


FIGURE 6.16. Epoxy dispensed at the edge.

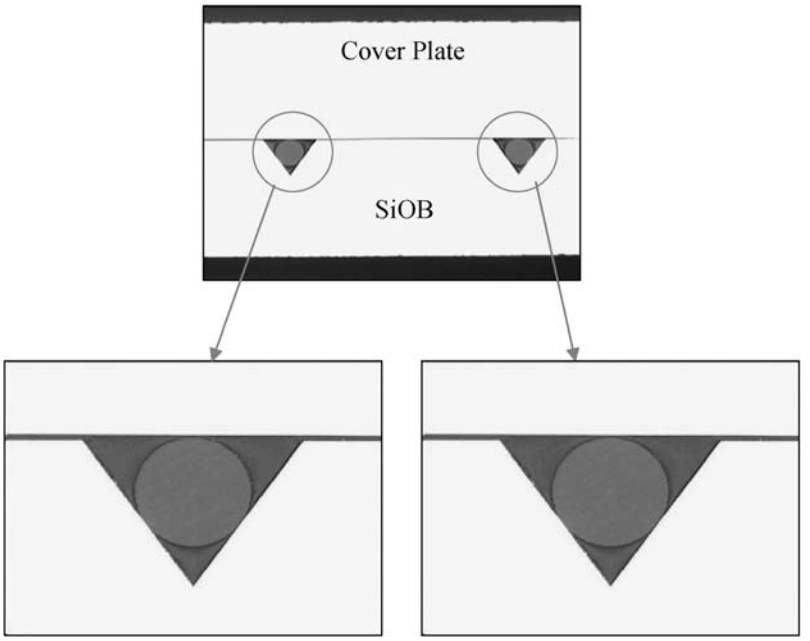


FIGURE 6.17. Cross-section inspection (edge dispensing method, low profile SiOB).

Figure 6.17 shows the cross-section inspection obtained by using edge dispensing method on a low profile SiOB. The figure shows that the optical fibers are well aligned and no void and air bubble is found in the epoxy because the gap is filled up by the epoxy flow. The buoyancy of epoxy is minimized so that both the optical fibers and the cover plate are not lifted up.

The results of applying edge dispensing of epoxy is encouraging. The fibers are well aligned without air bubble. However, the cover plate is still tilted when there is only one fiber on the SiOB. Figure 6.18 shows the tilted cover plate. Although no crack is found from the cross-section inspection, the stress of the cover plate and the fiber may still be very high due to the reasons mentioned before.

6.4. MODIFIED PASSIVE ALIGNMENT METHOD

In earlier sections, some drawbacks of the conventional passive alignment methods are presented. In this section, new modified passive alignment method will be introduced. This method focuses on aligning a single fiber on a SiOB without the help of a cover plate.

6.4.1. Working Principle

The problems found in the conventional passive alignment method are mainly caused by the buoyancy of the epoxy and the application of the cover plate. If the dispensing method is not well controlled, voids and air bubbles may exist in the epoxy. This can be solved by applying the edge dispensing method. However, if only one fiber is placed in the SiOB, the cover plate will be tilted. This leads to non-uniform stress distribution and die cracking.

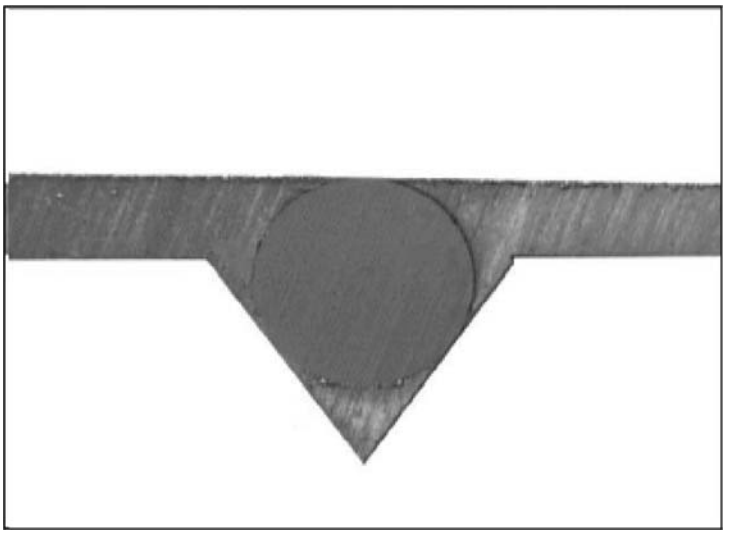


FIGURE 6.18. Cross-section inspection (edge dispensing method, high profile SiOB, single fiber).

From the lessons learned, the edge dispensing method is developed which has a good alignment result with the help of epoxy flow. The steady flow of epoxy from the edge to the gap will prevent any void or air bubble from being trapped. However, if the cover plate is removed, the flow of epoxy will not be well controlled.

To improve the conventional passive alignment method, two new features, reservoir and canal are introduced. Their function is to induce a steady epoxy flow into the gap between the optical fiber and the bottom of the V-groove. This can completely replace the use of the cover plate.

Since the epoxy flow into the gap between the fiber and the bottom of the V-groove, it only wets the bottom part of the fiber. Unbalanced surface tension is acted on the fiber during the epoxy flow. This surface tension will pull down the fiber against the wall of the V-groove. Consequently, the modified passive alignment method has a self-aligning property like the reflow of solder joints in surface mount technology (SMT).

6.4.2. Alignment Mechanism

In order to have epoxy flowing into the gap between the optical fiber and the bottom of the V-groove, epoxy is dispensed and accumulated in the reservoir as shown in Figure 6.19. The epoxy then gathers around the center of the reservoir. It runs along the axial direction of the fiber and finally flows into the V-groove by capillary effect. As presented earlier, undercutting is created during the fabrication process and the reservoir will no longer be rectangular in shape. The feature is similar to a funnel which guides the epoxy to flow steadily into the gap instead of inducing a sharp turn. The gap between the fiber and the bottom of the V-groove is then filled up completely. However, as epoxy flows both toward the fiber tip and the back, if too much epoxy is accumulated at the back, the fiber will be lifted up. At this stage, the flow direction cannot be controlled. Therefore a canal is added to act as a stopper to prevent the epoxy from flowing backward and drive the excess epoxy away. It is redundant to have a canal if the reservoir is placed in the middle of the package. However, this will increase the package size.

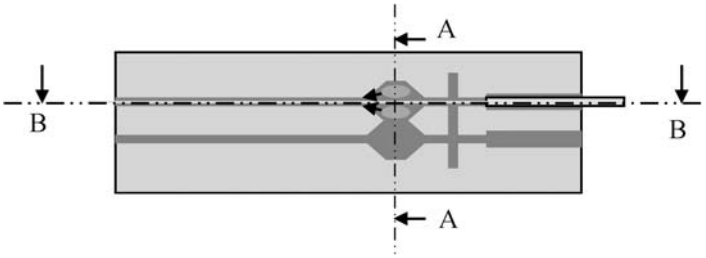


FIGURE 6.19. Alignment mechanism (top view).

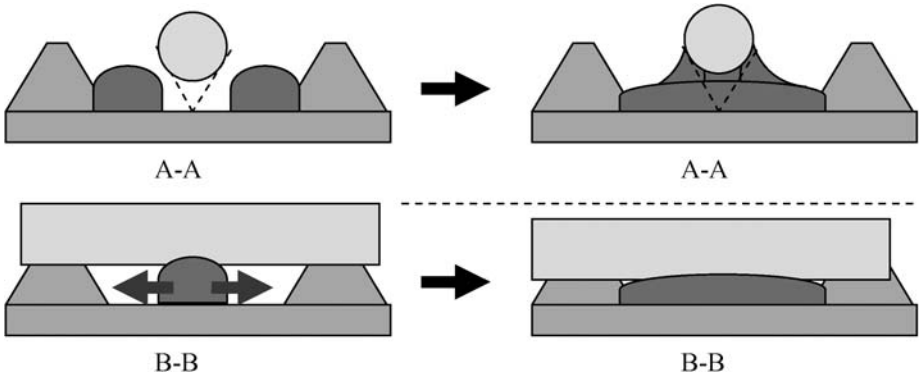


FIGURE 6.20. Alignment mechanism (cross-sectional view).

The cross-section shown in Figure 6.20 explains the alignment mechanism. When the epoxy which gathered in the reservoir touches the fiber, it wets the bottom part of the optical fiber only. This generates an unbalanced surface tension. With the flow of the epoxy along the axial direction by capillary effect, the fiber will be pulled down by surface tension. Thus, the fiber is then aligned against the wall of the V-groove without using an additional cover plate to press it.

6.4.3. Design of Experiment

From the preliminary exercise, it is identified that the amount of epoxy dispensed may play an important role in the quality of passive fiber alignment. Besides, since the overhang length of the stripped fiber (the part with cladding) may affect the gap spacing between the fiber and the walls of the V-groove, this parameter is also considered as a potential factor that may have a certain effect on quality of the passive alignment. Hence, a series of parametric studies with various combinations of epoxy volume and overhang fiber length are performed as listed in Table 6.3. Unfortunately, due to the availability of epoxy materials, it is unable to investigate the effect of various epoxy viscosity on the quality of the passive alignment.

6.4.4. Experimental Procedures

When the fiber is placed in the desired position, Epoxy B is dispensed into the reservoir. Epoxy B is used because its viscosity is lower. Excess epoxy is directed to the canal

TABLE 6.3.
Test matrix for parametric study.

Epoxy weight	Fiber length (w/cladding)		
	8 mm	11 mm	14 mm
0.3 mg	Sample #1	Sample #2	Sample #3
0.4 mg	Sample #4	Sample #5	Sample #6
0.6 mg	Sample #7	Sample #8	Sample #9

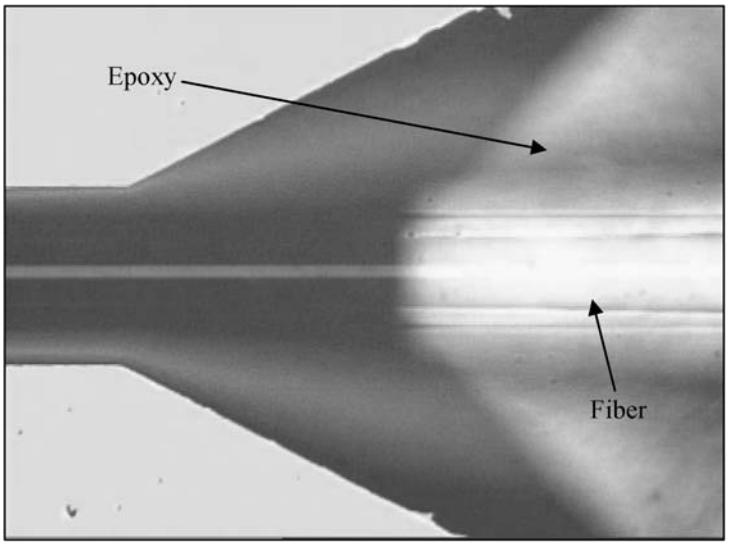


FIGURE 6.21. Epoxy flow into the V-groove from the reservoir.

next to the reservoir. As a result, epoxy flows from the reservoir into the V-groove as shown in Figure 6.21.

During the running of epoxy flow, the fiber is pulled by the surface tension. Once the lateral surface of the fiber touches the walls of the V-groove, the fiber is aligned accordingly. For the current configuration and dimensions, the epoxy flow normally takes about 5 minutes to stop. Then, the epoxy is cured under UV light to fix the fiber position. Here, UV light is used as the fiber is transparent.

6.4.5. Experimental Results

In most optoelectronic applications, the accuracy requirement for optical fiber alignment may reach 0.5 μm. Therefore, in the present study, 0.5 μm offset from the perfectly aligned position is used as a benchmark to evaluate the quality of fiber alignment. The results of the parametric study mentioned earlier are listed in Table 6.4. It is found that some differences exist in the initial misalignment, it is because the fiber is placed by human hand. The fiber aligns with the wall of V-groove even if the initial misalignment is greater than 60 μm. Although there are some differences in the wetted length (the length of the fiber with cladding that is wetted by the epoxy), there is only one case fails to meet the 0.5 μm criterion among all 9 cases.

TABLE 6.4.
Results of parametric study.

	Initial misalignment (μm)	Wetted length (mm)	Within 0.5 μm
Sample #1	48.0	3.3700	Yes
Sample #2	68.0	3.1035	Yes
Sample #3	47.5	3.5670	No
Sample #4	44.8	3.6465	Yes
Sample #5	51.5	3.1865	Yes
Sample #6	39.0	2.7200	Yes
Sample #7	56.0	2.2175	Yes
Sample #8	49.0	3.5300	Yes
Sample #9	51.3	3.1175	Yes

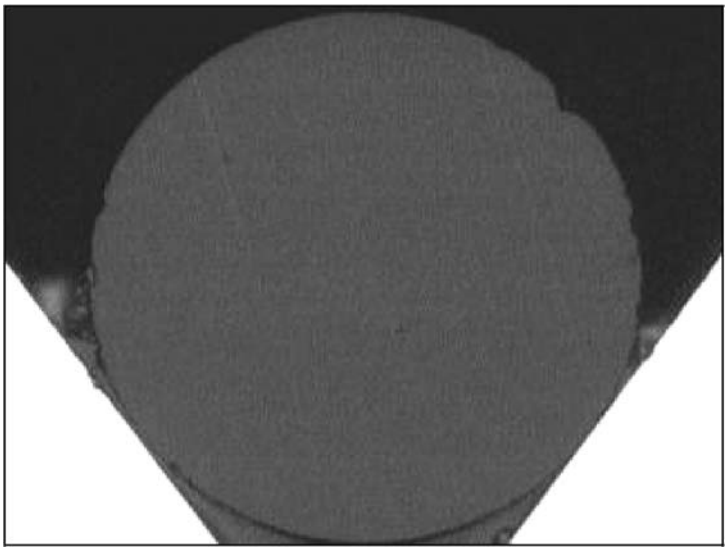


FIGURE 6.22. Well aligned fiber with V-groove.

Figure 6.22 shows the cross-section of a well-aligned fiber using the modified passive alignment method. It is observed that the surface of the fiber is well aligned with the walls of the V-groove. It should be noted that there is little epoxy in the neighborhood of contact points. In fact, most epoxy gathers underneath the fiber, which generates the force to pull down the fiber during the running of epoxy flow. Figure 6.23 shows the cross-section of a fiber with poor alignment. This defect is resulted from epoxy overflow. Several parameters, such as the initial misalignment and the amount of the epoxy added are related to this overflow.

The capability of the newly designed passive alignment method is also tested by monitoring the movement of the optical fiber during the process in real time. The experiment is conducted by coupling optical fibers on one SiOB. By detecting the power received, the movement of the fiber during the passive alignment process can be monitored.

In the experiment, one fiber is initially misaligned and the other one is aligned to the V-groove by the method mentioned above as shown in Figure 6.24. The cladding lengths of

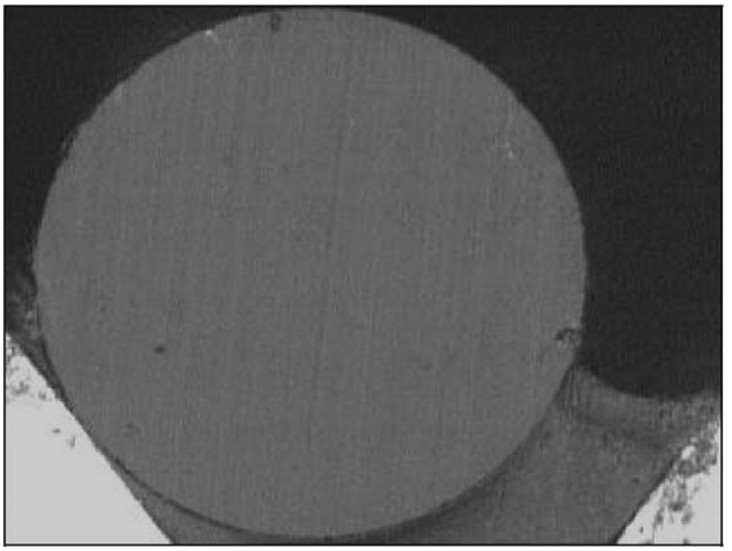


FIGURE 6.23. Misaligned fiber with V-groove.

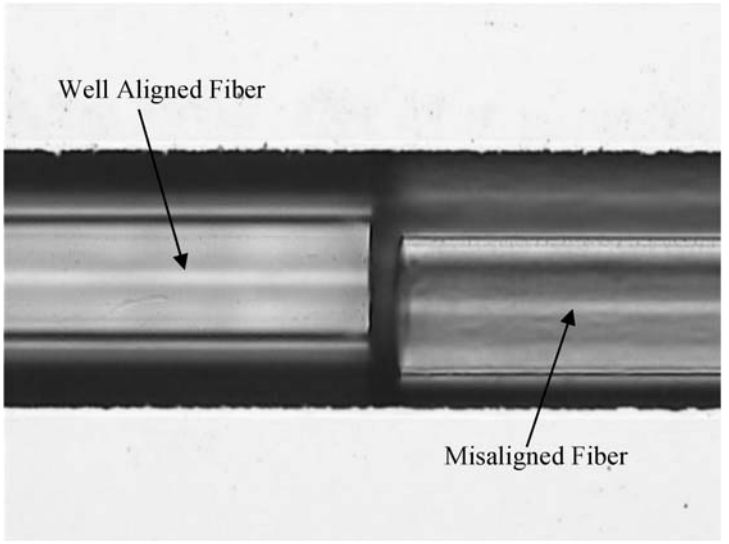


FIGURE 6.24. Initial misalignment.

the fibers are 8 mm. Epoxy B is dispensed in the reservoir next to the misaligned fiber. Light is coupled from the misaligned fiber to the aligned fiber during the alignment process. By simply monitoring the power received from the aligned fiber, the movement of the optical fiber can be evaluated in real time.

Both single-mode and multi-mode optical fiber are used. Figure 6.25 shows the results obtained by coupling a single-mode fiber to another single-mode fiber, and Figure 6.26 shows the results obtained by coupling a multi-mode fiber to another multi-mode fiber. The input power of the single-mode fiber is 2.3 mW where that of the multi-mode fiber

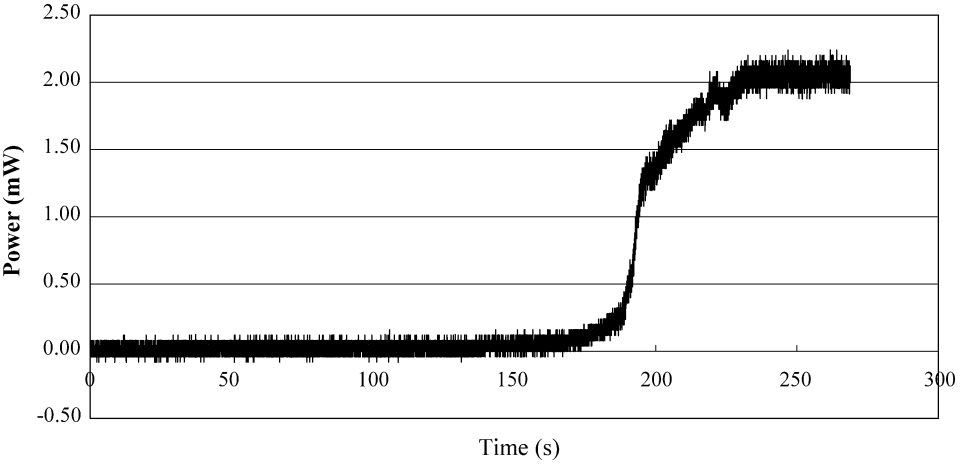


FIGURE 6.25. Real time power monitoring (single-mode to single-mode).

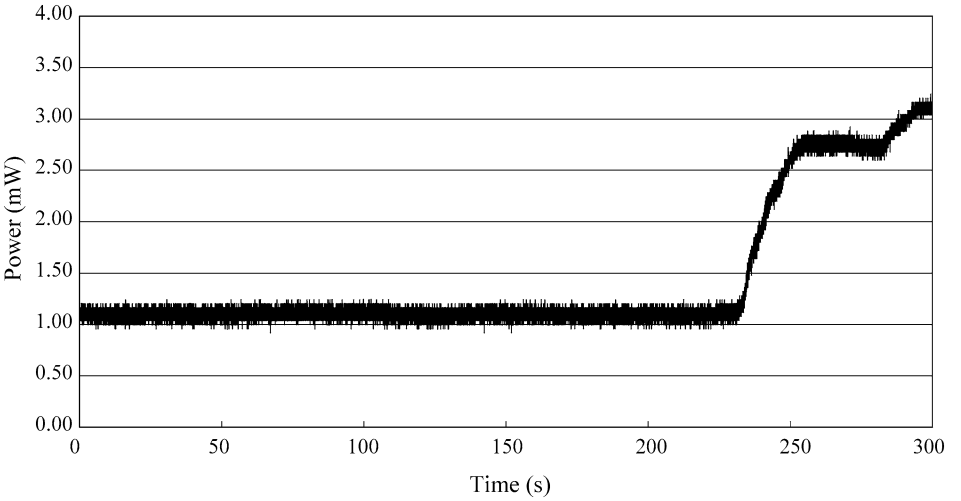


FIGURE 6.26. Real time power monitoring (multi-mode to multi-mode).

is 3.0 mW. The results show that the method achieves higher than 90% coupling efficiency. The whole alignment process takes less than one minute but the flow time is about 5 minutes.

6.5. EFFECTS OF EPOXY VISCOSITY AND DISPENSING VOLUME

From the experimental results, it is observed that the epoxy viscosity is very critical in the alignment process. The viscosity of epoxy affects the flow length, process time and the performance. Besides, the dispensing volume plays a very important role in the yield. If too much epoxy is dispensed, epoxy may overflow and cause misalignment. However, if too little epoxy is dispensed into the reservoir, the epoxy may not flow into the gap between

optical fiber and the wall of V-groove. Also, it is difficult to control if the dispensing volume is too small.

In this section, four types of epoxy with different viscosity are used. For each type of epoxy, different epoxy volume is dispensed. In this parametric study, only one fiber is placed on the SiOB. Cross-section inspection is preformed to verify the alignment of the optical fiber. It is claimed to be aligned if the offset of the core centre is less then 1 μm. The theoretical position of the well aligned fiber is shown in Figure 6.27. The test matrix and the results are shown in Table 6.5. The columns of the matrix are arranged by increasing the viscosity and the row of the matrix are arranged by increasing the dispensing volume of epoxy. By completing the matrix, the effect of the epoxy viscosity and dispensing volume are analyzed.

For each combination of epoxy viscosity and dispensing volume, ten samples are tested and analyzed. From the experimental result, it shows both epoxy viscosity and dispensing volume are important factors. The yield of alignment process decrease when the dispensing volume increases. It is because the epoxy may overflow if too much epoxy is dispensed. The canal cannot accommodate that excess epoxy and hence the optical fiber is lifted up can cause the misalignment. Experiment results also show that high viscosity epoxy causes misalignment. If the viscosity of the epoxy is too high, the epoxy cannot completely flow into the gap between the optical fiber and the wall of V-groove. Therefore, the optical fiber is not pulled down by the surface tension of the epoxy. The self alignment capability of the process is further reduced by the epoxy accumulated in the reservoir which may lift up the fiber. In general, dispensing low viscosity epoxy with a right amount of volume provides the best results.

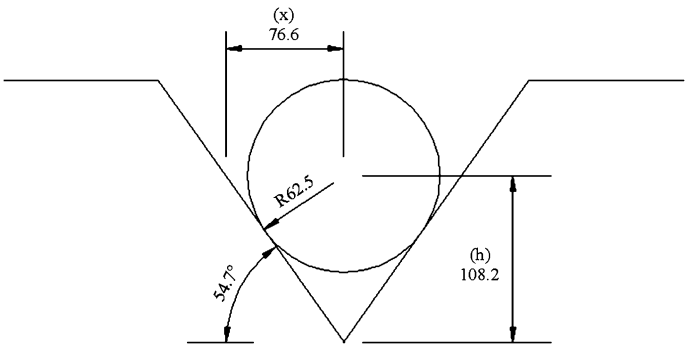


FIGURE 6.27. Theoretical position of well aligned fiber.

TABLE 6.5.
Effect of dispensing volume and viscosity.

	Dispensing volume (mm^3)			
Viscosity (Pa s)	0.28	0.56	0.84	1.12
7.5	7/10 aligned	3/9 aligned	2/10 aligned	1/10 aligned
12.2	4/10 aligned	4/8 aligned	1/10 aligned	0/10 aligned
19.5	2/10 aligned	4/10 aligned	4/10 aligned	1/10 aligned
24.8	2/10 aligned	0/10 aligned	0/10 aligned	1/10 aligned

6.6. APPLICATION TO FIBER ARRAY PASSIVE ALIGNMENT

The newly invented method is also applied to aligning fiber arrays. In the present study, fiber arrays with 2, 4 and 8 channels are also tested. There are two pitch sizes, 250 μm and 500 μm, for each array configuration. Figures 6.28 and 6.29 show the mask layout of 4 fiber array with 250 μm pitch and 500 μm pitch, respectively. The fabrication process is the same as mentioned above. Figure 6.30 shows the fiber array SiOB with reservoir and canal.

The experimental procedures are same as aligning single fiber. Fibers are first fixed on the SiOB with high viscosity epoxy. All the fibers on the SiOB are placed with intended initial misalignment. Low viscosity epoxy is dispensed on the reservoir. After the epoxy flow, the epoxy is cured by UV light and cross-section inspection is performed to analyze the alignment performance.

The experimental results are tabulated in Table 6.6. Cross-section inspection is performed to measure the alignment. Figures 6.31 and 6.32 show the cross-section view of 8 fibers array with 250 μm pitch and 500 μm pitch respectively. The samples are clamped to be aligned only all the fibers on the fiber array are aligned within 2 μm.

From the experimental results, it is found that the yield decrease when the number of fibers on the array and the pitch size increase. In this study, the epoxy is only dispensed once on the reservoir for aligning the fiber array. The epoxy is not evenly flow into each gap. Some fibers may be lifted up by the excess epoxy and the outermost fiber may not be aligned due to lack of epoxy. These are the possible reasons to the relatively low yield is obtained when the method is applied to fiber array.

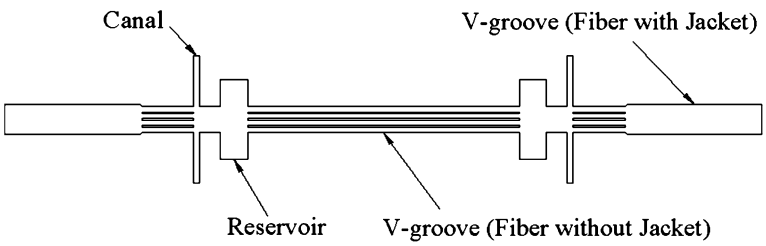


FIGURE 6.28. Mask layout of 4 fibers array (250 μm pitch).

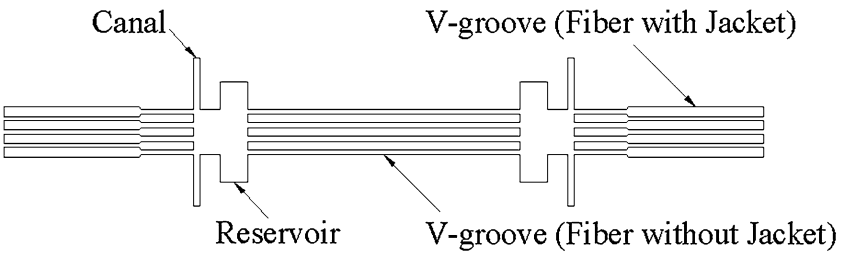


FIGURE 6.29. Mask layout of 4 fibers array (500 μm pitch).

TABLE 6.6.
Experimental results of fiber array.

	Pitch (μm)	Results
2 Fibers array	250	6/10 aligned
	500	8/10 aligned
4 Fibers array	250	6/10 aligned
	500	6/10 aligned
8 Fibers array	250	4/10 aligned
	500	2/10 aligned

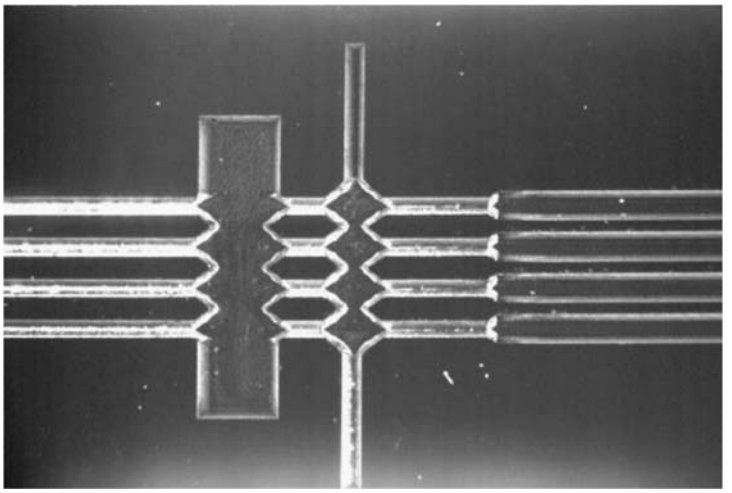


FIGURE 6.30. 4 fibers array (500 μm pitch).

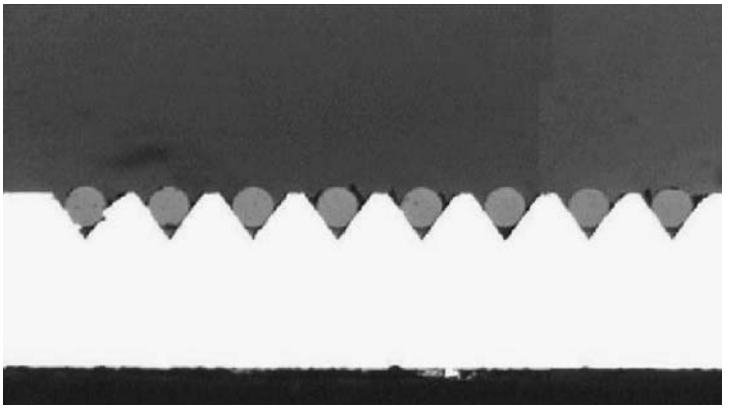


FIGURE 6.31. Cross-section view of 4 fibers array (250 μm pitch).



FIGURE 6.32. Cross-section view of 8 fibers array (500 μm pitch).

6.7. CONCLUSIONS AND DISCUSSION

In this chapter, several conventional passive alignment methods have been discussed. These methods have certain drawbacks which are mainly caused by the buoyancy of epoxy. Though the buoyancy can be overcome by using a cover plate, a high profile SiOB instead of a low profile one must be used. However, voids and air bubbles are easily found in the epoxy. This creates problem in the long-term reliability and position stability.

The situation can be improved by dispensing epoxy at the edge of the cover plate. Experiments prove that in this case, a low profile SiOB can be used and no void and air bubble are trapped in the epoxy. However, the cover plate is tilted when there is only one fiber on the SiOB. The epoxy underneath the tilted cover plate generates non-uniform distributed force. Both the optical fiber and the cover plate suffer from high compressive stresses.

Based on the advantages of the epoxy flow observed in the edge dispensing process, a modified passive alignment method is introduced. The epoxy is dispensed in the reservoir and flow into the gap between the optical fiber and the bottom of the V-groove. Parametric studies and real time monitoring experiment show this method has a self-align property. With this method, the optical fiber will align with the V-grooves without the use of cover plate.

The modified passive alignment method only eliminates the use of cover plate but does not improve the alignment accuracy. The overall alignment depends heavily on the precision of the V-groove fabrication process and the geometry of the optical fiber. This is the major disadvantage of the passive alignment when compared with active alignment.

REFERENCES

1. M.F. Dautartas, J. Fisher, H. Luo, P. Datta, and A. Jeantilus, Hybrid optical packaging, challenges and opportunities, Proc. 52nd ECTC, San Diego, CA, May 2002, pp. 787–793.
2. M.W. Beranek, et al., Passive alignment optical sub-assemblies for military/aerospace fiber-optic transmitter/receiver modules, IEEE Transactions on Advanced Packaging, 23(Aug.), pp. 461–469 (2000).
3. G. Keiser, Optical Fiber Communications, McGraw-Hill, New York, 2000.
4. D.K. Mynbaev and L.L. Scheiner, Fiber-Optic Communications Technology, Prentice Hall, New Jersey, 2001.
5. R.R. Tummala, Fundamentals of Microsystems Packaging, McGraw-Hill, New York, 2001.
6. F.G. Smith and T.A. King, Optics and Photonics, John Wiley & Sons, Chichester, 2000.
7. J.A. Buck, Fundamentals of Optical Fibers, John Wiley & Sons, Chichester, 1995.
8. P. Karioja, et al., Comparison of active and passive fiber alignment techniques for multimode laser pigtailing, Proc. 50th ECTC, Las Vegas, ND, May 2000, pp. 244–248.
9. S.H. Law, T.N. Phan, and L. Poladian, Fibre geometry and pigtailing, Proc. 51st ECTC, Orlando, FL, May 2001, pp. 1447–1450.
10. K. Ishikawa, An integrated micro-optical system for laser-to-fiber active alignment, Proc. IEEE 15th MEMS, Jan. 2002, pp. 491–494.
11. J. Goodrich, A silicon optical bench approach to low cost high speed transceivers, Proc. 51st ECTC, Orlando, FL, May 2001, pp. 238–241.
12. R. Hauffe, U. Siebel, K. Petermann, R. Moosburger, J.-R. Kroop, and F. Arndt, Methods for passive fiber chip coupling of integrated optical devices, IEEE Transactions on Advanced Packaging, 24(Nov.), pp. 450–455 (2001).
13. R.A. Boudreau, Passive alignment in optoelectronic packaging, Optical Fiber Communication, OFC 97, Feb. 1997, pp. 109–110.
14. S.J. Park, et al., A novel method for fabrication of a PLC platform for hybrid integration of an optical module by passive alignment, IEEE Phton. Technol. Lett., 14(Apr.), pp. 486–488 (2002).
15. H. Mori, et al., LD and PD array modules assembled in a new plastic package with auto-alignment projections for silicon optical bench, Pro. OFC/IOOC, 3(Feb.), pp. 198–200 (1999).

16. G. Grand and C. Artigue, Hybridization of optoelectronic components on silicon substrate, Proc. ECOC'94, 1994, pp. 193–200.
17. R. Moosburger, B. Schüppert, U. Fischer, and K. Petermann, Passive alignment technique for all-silicon integrated optics, Proc. Integr. Photon. Res., Boston, MA, IWH3, Apr. 1996, pp. 565–568.
18. R. Moosburger, R. Hauffe, U. Siebel, D. Arndt, J. Kropp, and K. Petermann, Passive alignment of single mode fibers to integrated polymer waveguide structures utilizing a single mask process, IEEE Photon. Technol. Lett., 11, pp. 848–850 (1999).
19. M.W. Beranek, et al., Passive alignment optical sub-assemblies for military/aerospace fiber-optic transmitter/receiver modules, IEEE Trans. Advanced Packaging, 23(Aug.), pp. 461–469 (2000).
20. M.F. Grant, et al., Self-aligned multiple fibre coupling for silica-on-silicon integrated optics, Proc. 9th Annual European Fibre Optic Conference, London, UK, Jun. 1991, pp. 269–272.
21. J.W. Osenbach, et al., Low cost/high volume laser modules using silicon optical bench technology, Proc. IEEE 48th ECTC, May 1998, pp. 581–587.
22. K. Kurata, et al., A surface mount single-mode laser module using passive alignment, IEEE Transactions on Components, Packaging, and Manufacturing Technology, 19(3), pp. 524–531 (1996).
23. K. Yamauchi, et al., Automated mass production line for optical module using passive alignment technique, Proc. 50th ECTC, Las Vegas, ND, May 2000, pp. 15–20.
24. C.B. Probst, A. Bjarklev, and S.B. Andreasen, Experimental verification of microbending theory using mode coupling to discrete cladding modes, 7(Jan.), pp. 55–61 (1989).
25. C. Unger and W. Stocklein, Investigation of the microbending sensitivity of fibers, Journal of Lightwave Theory, 12(Apr.), pp. 591–596 (1994).
26. J.C.C. Lo and S.W.R. Lee, Experimental assessment of passive alignment of optical fibers with V-groove on silicon optical bench, Proc. 6th EPTC, Singapore, December 2004, pp. 375–380.
27. J. Lo, R. Lee, S. Lee, J.S. Wu, and M. Yuen, Modified passive alignment of optical fibers with low viscosity epoxy flow running in V-grooves, Proc. IEEE 54th ECTC, Jun. 2004, pp. 830–834.
28. J. Lo, C.S. Yung, R. Lee, S. Lee, J.S. Wu, and M. Yuen, Passive alignment of optical fiber in V-groove with low viscosity epoxy flow, Proc. ASME IMECE, Nov. 2003, paper IMECE 2003/43902.
29. K.E. Bean, Anisotropic etching of silicon, IEEE Trans Electron Devices, ED-25, pp. 1185–1193 (1978).
30. C.W. Chang and W.F. Hsieh, Micromachined double-side 45° silicon reflectors for dual-wavelength DVD optical pickup heads, Proc. IEEE 54th ECTC, Jun. 2004, pp. 1390–1395.
31. C. Strandman, et al., Fabrication of 45° mirrors together with well-defined v-grooves using wet anisotropic etching of silicon, Journal of Microelectromechanical System, 4(Dec.), pp. 213–219 (1995).
32. S.A. Campbell and H.J. Lewerenz, Semiconductor Micromaching Volume 1 Fundamental Electrochemistry and Physics, John Wiley & Sons, Chichester, 1998.
33. S.A. Campbell and H.J. Lewerenz, Semiconductor Micromaching, Volume 2, Techniques and Industrial Applications, John Wiley & Sons, Chichester, 1998.
34. E. Bassous, Fabrication of novel three-dimensional microstructures by the anisotropic etching of (100) and (110) silicon, IEEE Trans Electron Devices, ED-25, pp. 1178–1185 (1978).
35. M. Sekimura, Anisotropic etching of surfactant-added TMAH solution, Proc. IEEE 12th MEMS, Jan. 1999, pp. 650–655.
36. W. Sonphao and S. Chaisirikul, Silicon anisotropic etching of TMAH solution, Proc. IEEE ISIE, Jun. 2001, pp. 2049–2052.

RELIABILITY AND PACKAGING

7

Fundamentals of Reliability and Stress Testing

H. Anthony Chan

Department of Electrical Engineering, University of Cape Town, Rondebosch, 7701, South Africa

This chapter discusses the concepts which in the author's opinion are fundamental to understand the reliability of electronics and packaging. It also summarized some conventional reliability backgrounds.

Reliability in electronics and packaging is often interpreted differently in different contexts. In some reliability programs, reliability is an interdisciplinary science aimed at predicting, analyzing, preventing and mitigating failures over time. To a manufacturer, reliability may simply be the probability of "not failing" for a "specified" period of time and under "specified" conditions when used in the manner and for the purpose intended. In a seller market where the demand from the customers exceeds the availability of the merchandise, it is tempting for the manufacturers to make the specifications only according to the technical capability of their products. Yet the customers who encounter failures may often be upset. The customers may disagree with the manufacturers on various issues, especially when the specifications are made from the perspectives of the manufacturers alone. Especially in a buyer market, reliability is then the avoidance of failures as experienced by the customers and defined by the customers.

It follows that a good approach in reliability is to understand the causes of failures and then to avoid these causes.

Section 7.1 gives a non-technical discussion of the challenges and trends in reliability, which is a non-technical introduction for management staff, application engineers, and anyone interested in reliability without going into technical depth. Two questions are often addressed. One wants to know how often failures occur. The fundamentals on failure rate and failure distributions are given in Section 7.2. One also wants to understand why failures occur by conducting failure analysis and root cause analysis. Section 7.3 list the failure mechanisms in electronics and packaging. Reliability programs to improve reliability are explained in Section 7.4. The fundaments of product weaknesses and stress testing are given in Section 7.5. Finally, the formulation of stress testing is explained in Section 7.6.

7.1. MORE PERFORMANCE AT LOWER COST IN SHORTER TIME-TO-MARKET

The challenges faced in the electronics industry is not just how to make more reliable products but to do it in the ever more competitive market. Reliability is becoming essential in electronics as the electronics are being used in more products that have become part of human life and culture. Product performance is increasing. Yet product cycles are short, and product cost continues to be cut.

7.1.1. Rapid Technological Developments

The electronics and packaging technology has been rapidly progressing toward higher complexity, system integration, and product miniaturization. With these fast technological improvements, one is always working on the reliability of a new product. Failure modes that did not show up in a previous technology may now be important under factors like increased power, density and speed of operation. New failure modes may also arise with new materials and manufacturing processes under a new technology.

In addition, the three major products of communications, computing and consumer electronics have merged. Before the late 1990's, these three types of products had different reliability requirements. A telecommunication system may not tolerate any failure in over 30 years. Very high reliability is an important consideration for a customer to buy such a system although it may cost more to achieve that reliability level. On the other hand, the sale of a consumer product may be affected primarily by its price and features. A comparatively lower reliability level may be enough. For example, a failure fraction of 2–5% or higher over a product life of 5–10 years is often reported for consumer products. Software products with high failure rates have been penetrating the market of personal computers. Yet, when a low cost product is merged with other systems that need to be highly reliable, this product is contributing to and interacting with the reliability of the overall system. The reliability requirements for these low cost products need be higher now.

7.1.2. Integration of More Products into Human Life

Products from calculators to microwaves and cellular phones are changing human life. As people are becoming more dependent on these products, product failures are affecting people more than before. In the last millennium, vendors often advertised their products in terms of functionality alone. The warranty periods were also short for these products (e.g., 1 year or 90 days). Today, more customers are concerned about possible failures and tend to check with their friends and relatives before they buy a product. Customers also often read reports on the repair history of a commodity under different brand names. As for manufacturers, they now tend to adopt a longer warranty policy to compete in the market. Warranty periods from three to seven years are becoming competitive edges.

7.1.3. Diverse Environmental Stresses

The field stresses seen by customers of electronic devices are diverse and dynamic. Portable and hand held electronics are also vulnerable to outdoor environments. The extreme temperatures (e.g., -20 to 70°C) and temperature cycles differ in different parts of the world. Humidity and the extent of corrosive contamination in the air also differ. Handling by hands and tools often imposes hazardous stresses. These include mechanical

shocks when accidentally dropped and electrostatic discharges of several thousand volts when a metallic lead is touched by another conducting body. In addition, users are usually not trained operators and often do not read manuals. Random on and off switching, unexpected modes of operation, attempts to plug into an improper power source and attempts to connect to an incompatible interface are common mistakes. If the products are not robust against the stresses under these user environments, excessive failures may result.

7.1.4. Competitive Market

Today, the world has moved to an Age of Cost Reduction, which is affecting industry, government, commerce and practically everything else. The need to cut cost in highly competitive markets tends to leave reliability efforts to a minimum. Yet, short-sighted cost cutting at the expense of product reliability is expensive in the long run. Early product failures result in warranty repairs, which usually cost much more than would reliability programs to avoid these failures. Early product failures also affect the buying decisions of customers on both current and future products under the same brand name.

7.1.5. Short Product Cycles

Short product cycles make traditional reliability programs difficult to implement owing to the short failure history available and a lack of data for failure analysis for new products. The urgency to bring a product to market in a short product cycle can no longer accommodate reliability programs that are passively only in response to field returns. A pro-active approach is needed to consider reliability, starting before the product design stage.

7.1.6. The Bottom Line

Despite the reliability concerns, revenue growth is the bottom line for investors and corporate owners whose investments govern manufacturers and their research and development programs. Cost-cutting programs have been abundant for many companies since the mid 1980s. The short-term benefits of cost- cutting usually dominate over the long-term health considerations of the business. Few product and process owners realize the long-term importance of reliability to the business. Even for the minority who are willing to put resources into reliability programs, they still need to see the monetary returns of their investment in reliability.

A major selling point of reliability is the avoidance of repair costs, which grow exponentially as a product goes through various stages from early design to maturity. A full stream cost consideration, which includes the repair cost, is needed. In addition, the effects of failures on customers may be serious. Yet, the effects on the vendor may again be judged by how much business will be lost when the investment in proper reliability efforts is not in place.

Some products may be critical to the customer's revenue. For example, a service provider that buy telecommunication systems ought to be cautious that a failure may result in substantial losses to the provider. The cost of providing redundancy to avoid such losses can also be high. For such situations, it is critical for the manufacturer to ensure the highest reliability of the product being sold to the service provider. In other markets, such as consumer products, a customer may buy a product based principally on first cost or features. Yet, the product reliability should still be sufficient for the customer. Excessive failures will

damage the brand name of the manufacturer, resulting in loss of sales even for other unrelated products under the same brand name. For either high reliability systems or consumer products, it is important to have a clearly defined reliability objective and understand the economic factors influencing the setting of that objective.

7.2. MEASURE OF RELIABILITY

Two questions are often addressed. One wants to know how often failures occur by measuring the failure rate. One also wants to understand why failures occur by conducting failure analysis and root cause analysis. These data are helpful to the design and manufacturing processes to prevent failures. We summarize the measure of reliability first.

The reliability of different systems may be characterized in different ways.

Failure Rate (Hardware components and systems): The metric for the reliability of hardware components and systems is often expressed in terms of the measure of "unreliability" or the failure rate, which is also known as hazard rate.

Failure Intensity (Software): The metric for the reliability of software is the same as that for hardware systems but is called failure intensity.

Availability (Service): The metric for reliability of service is often called availability. The availability of a system is the probability that the system will be available to perform the intended actions.

Downtime (Computer, Telecommunication): The reliability for computer and telecommunication is often measured as the downtime.

Risk (network): The reliability for a network is also called the risk.

7.2.1. Failure Rate

Cumulative distribution function, or Cumulative fraction failed, $F(t)$ is the probability that a system first fails at or before time t. Denoting the service life of a product by ts Cumulative fraction failed over the service life is $F(ts)$.

Reliability function $R(t)$, or Survivor function $S(t)$ is the probability that a system survives to time t without failure:

$$R(t) = S(t) \equiv 1 - F(t). \tag{7.1}$$

Probability density function $f(t)$ is the probability of failure per unit time (per unit product born at time $t = 0$) occurring at time t. $f(t)$ is related to $F(t)$ by

$$f(t) \equiv \frac{d}{dt} F(t), \tag{7.2}$$

but $f(t)$ is NOT the failure rate.

The (instantaneous) failure rate $\lambda(t)$, or Hazard rate $h(t)$, is defined as the probability of failure per unit product that was working at time $<t$. This failure rate is equal to $f(t)$ over the fraction that survives to time t.

$$h(t) = \lambda(t) \equiv \frac{f(t)}{1 - F(t)}. \tag{7.3}$$

The probability density function $f(t)$ is therefore not identical to the failure rate because some product units born at $t = 0$ would have already failed before time <0, leaving only a the fraction $1 - F(t)$ to survive to time t.

Failure in time FIT is a unit of $\lambda(t)$ often used for electronic components. It is equal to the number of failures per 10^9 hours.

When $\lambda(t)$ is equal to n FITs, it means

$$\lambda(t) = \frac{n}{10^9 \text{ hours}}. \tag{7.4}$$

$$10 \text{ FITs} \cong 1 \text{ in} 10^4 \text{ per year}. \tag{7.5}$$

Failure rate is time dependent.

Instead of counting the number of failures per working product unit per unit time, an alternate measure is the time from time = 0 to the time a working unit has just failed. This is called the time to failure for any given particular product unit. The arithmetic average over a large number of product unit of such time to failure is called the mean-time-to-failure, MTTF, of the product. Many electronic components, modules, and sub-systems are non-repairable, so that the only meaningful MTTF for them is the mean-time to the first failure.

Mean Time to Failure MTTF is the average time to first failure:

$$\text{MTTF} \equiv \int_0^\infty dt\ t f(t). \tag{7.6}$$

For a repairable product unit, it may be repaired after the first failure and is then working again. After some time, it may fail again. This time is the time between the first and second failures. In the particular case that the average time (over many units of the same product) between the first and second failures is equal to that between the second and third failures and so on, we can define the mean-time-between-failures, MTBF, for that product. Systems comprising separate components or modules are often repaired by replacing the defective sub-systems with new sub-systems, so that one may talk about the MTBF if we assume that the rest of the system were as good as when they were new.

For repairable systems and assuming the repairs are ideal, the Mean time between failures MTBF is the average cycle time including operation and down times for the system:

$$\text{MTBF} \equiv \text{MOT} + \text{MDT}, \tag{7.7}$$

where $\text{MOT}(t)$ is mean operating time, and $\text{MDT}(t)$ is mean downtime.

The availability is related to MOT and MDT by

$$\text{Availability} \equiv \frac{\text{MOT}}{\text{MOT} + \text{MDT}}. \tag{7.8}$$

7.2.2. Systems with Multiple Independent Failure Modes

A system usually has multiple failure modes. In the simplest case, let us assume or take the approximation that these failure modes are independent of each other. Such a system may be modeled as a series system or parallel system.

A series system fails when any (one) failure mode occurs. An example of a series system is an IC that may fail caused by electromigration, overvoltage, or corrosion, etc. Another example is a system consisting of different subsystems.

Consider a series system with N independent failure modes. If the reliability functions for each of the individual failure modes are $R_1(t)$, $R_2(t), \ldots, R_N(t)$, the reliability function of the system $R_{12\ldots N}(t)$ is then given by

$$R_{12\ldots N}(t) = R_1(t)R_2(t)\ldots R_N(t). \tag{7.9}$$

When $F_1(t), F_2(t), \ldots, F_N(t)$ are small, i.e., $R_1(t) = 1 - F_1(t)$ is close to unity, one can take the approximation to relate the cumulative functions:

$$F_{12\ldots N}(t) \cong F_1(t) + F_2(t) + \cdots + F_N(t). \tag{7.10}$$

When $F_{12\ldots N}(t)$ is also small, one may take further approximation to relate the failure rates by:

$$\lambda_{12\ldots N}(t) \cong \lambda_1(t) + \lambda_2(t) + \cdots + \lambda_N(t). \tag{7.11}$$

A parallel system with N failure modes fails when all N failure mode occurs. Consider a communication system linking A and B with N channels. If some of these channels are down, communication between A and B will be downgraded. Yet in the case that a low data rate is needed, the communication will be out when all the N channels are out.

If the cumulative distribution functions for each of the individual failure modes are $F_1(t), F_2(t), \ldots, F_N(t)$, the cumulative distribution function of the system $F_{12\ldots N}(t)$ is then given by

$$F_{12\ldots N}(t) = F_1(t)F_2(t)\ldots F_N(t). \tag{7.12}$$

When $F_1(t), F_2(t), \ldots, F_N(t)$ are small, one can take approximation to relate the failure rates by:

$$\lambda_{12\ldots N}(t) \cong \lambda_1(t)\lambda_2(t)\ldots\lambda_N(t). \tag{7.13}$$

The parallel system design provides a way to achieve much smaller failure rate of the overall system than that of the individual subsystems by providing redundant subsystems. Yet there is cost in providing redundancy.

7.2.3. Failure Rate Distribution

The dependency of failure rate on time is typically pictured with a bath-tub curve. At large times, the failure rate rises up as the product wears out. At much shorter time typically within the first year for electronics, the failure rate may also be higher. These early failures are also known as infant mortality and their failure mechanisms are generally different from those of the long-term reliability.

7.2.3.1. Weibull Failure Rate Distribution One way to model the failure rate distribution is to use the Weibull distribution:

$$\lambda(t) \equiv \lambda_1 t^{-\alpha}, \text{ where } \alpha < 1. \tag{7.14}$$

Weibull distribution may be derived for the failure rate of a system with a large number of independent and competing failure mechanisms, and is therefore applicable to many different systems. This failure rate is quite general in the sense that it remains constant (exponential distribution) for $\alpha = 0$, decreases with time for $0 < \alpha < 1$, and increases with time for $\alpha < 0$. In order to model the bathtub behavior of the failure rate, one may model the different regions separately.

The reliability function for Weibull distribution is given by

$$1 - F(t) = R(t) = \exp\left(-\frac{\lambda_1 t^{1-\alpha}}{1-\alpha}\right) \equiv \exp\left[-\left(\frac{t}{t_e}\right)^{1-\alpha}\right]. \tag{7.15}$$

The term $1 - \alpha$ is called the Weibull shape, and the time when $R(t)$ has dropped to $1/e$ is t_e.

It is customary to plot $\ln F(t)$ against $\ln t$, called the Weibull plot.

Taking natural log twice for the reliability function, one obtains

$$\ln\left[\ln \frac{1}{1 - F(t)}\right] = (1-\alpha)\ln t + \ln \frac{\lambda_1}{1-\alpha}. \tag{7.16}$$

In the approximation with small $F(t)$, the left hand side is approximately equal to $\ln F(t)$, so that the relation between $\ln F(t)$ and $\ln t$ is linear.

$$\ln F(t) \cong (1-\alpha)\ln t + \ln \frac{\lambda_1}{1-\alpha}. \tag{7.17}$$

The slope in the Weibull plot is the Weibull shape. When there are 2 different failure modes that dominates in different regions of time, the Weibull plot will change from a straight line segment in one range of time with one value of slope to another straight line segment in another range of time with a different value of slope. This is called the S-shape Weibull plot.

The above Weibull distribution is also called 2-parameter Weibull distribution. One physical difficulty with the 2-parameter distribution is that there is a non-zero probability of occurrence at all values of $t < 0$. A generalization is the 3-parameter Weibull distribution with the reliability function given by:

$$1 - F(t) = R(t) \equiv \exp\left[-\left(\frac{t - t_0}{t_e - t_0}\right)^{1-\alpha}\right], \quad \text{for } t > t_0, \tag{7.18}$$

and

$$1 - F(t) = R(t) \equiv 1, \quad \text{for } t < t_0. \tag{7.19}$$

This distribution includes a threshold time to failure t_0 so that failure may not occur before the time t_0. This is the case when certain processes must proceed to certain extent

before an event may occur. An analogy is an egg hatching process. A series of certain metabolism and growth processes must precede taking a minimum number of days. The egg therefore cannot hatch before that minimum amount of time.

7.2.3.2. Lognormal Failure Rate Distribution Another way to model failure rate distribution is to use lognormal distribution for which the probability density function $f(t)$ versus the natural logarithm of time $\ln(t)$ exhibits normal distribution with a mean of $\ln(t_{50})$ and a standard deviation of σ.

$$f(t) \equiv \frac{1}{t\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\ln t - \ln t_{50}}{\sigma}\right)^2\right]. \tag{7.20}$$

The cumulative distribution function is given by

$$F(t) \equiv \frac{1}{\sigma\sqrt{2\pi}} \int_0^t \frac{dt'}{t'} \exp\left[-\frac{1}{2}\left(\frac{\ln t' - \ln t_{50}}{\sigma}\right)^2\right], \tag{7.21}$$

which can be expressed as a function of $\ln(t)$ as follows:

$$F(t) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\frac{\ln t}{\sqrt{2}\sigma}} d\left(\frac{\ln t'}{\sqrt{2}\sigma}\right) \exp\left[-\left(\frac{\ln t'}{\sqrt{2}\sigma} - \frac{\ln t_{50}}{\sqrt{2}\sigma}\right)^2\right]. \tag{7.22}$$

The lognormal failure rate distribution is usually applicable when there is a process such as diffusion, electromigration, corrosion, etc. that takes time to proceed to failure.

7.3. FAILURE MECHANISMS IN ELECTRONICS AND PACKAGING

An often asked question is why failures occur. It is customary to conduct failure analysis on failed product units to find out the failure modes. The reliability program will then try to avoid these failures. A list of failure mechanisms in electronics and packaging are given here.

7.3.1. Failure Mechanisms at Chip Level Include

Bulk semiconductor: second breakdown, latch-up, single-event upset, bulk radiation effects, chip fracture.

Semiconductor–dielectric interface: alkali ion migration, slow trapping instability, hot-carrier effects, surface charge spreading, polarization, ionizing radiation effects.

Dielectric: time-dependent dielectric breakdown, dielectric wear-out, fracture, passivation layers, EOS/ESD breakdown.

Conductor and metalization: electromigration, microcracks, corrosion, metal migration, contact spiking, hillock formation.

7.3.2. Failure Mechanisms at Bonding Include

Die bond: voids in bonding alloys or adhesives, disbonding, excessive thermal resistance, contamination from die bonding materials, die cracking by die-bond-induced stress.

Wire bonds: underbonding, overbonding, purple plague (Au-Al intermetallic), contaminant effects, wire break above bond.

7.3.3. Failure Mechanisms in Device Packages Include

Hermetic packages: leak–moisture–corrosion/electrical isolation, movement of bond wires and loose particles, contaminants–halide–Al metal film corrosion, electrolytic etching of Au wire bonds.

Plastic encapsulation: moisture penetration along plastic–lead interface or through plastic—delamination and popcorning during IR solder reflow—corrosion of chip metalization; package stresses from temperature cycling with TCE mismatch (large package)—shear bond wire, dielectric stress cracking, open, short; Ionic impurities, fire retardant—corrosion; radiation from plastic materials.

Electrochemical: solderability, bondability, delamination/adhesion, corrosion, ionic contamination, electromigration.

Thermomechanical: thermal conductivity deficiency, thermal stress due to CTE mismatch, delamination/popcorning, lead bend fatigue, lead break, stress/die attach pad shift.

7.3.4. Failure Mechanisms in Epoxy Compounds Include

Package assembly: moldability (flow and release), wire sweep, void level, mark permanency, wire bond pull strength, line movement (chip metal traces), board adhesion.

Failure mechanisms in soldering include:

Package defect: delamination, popcorn, microcrack.

Joint: intermetallic, non-wetting (oxidation, contamination), no solder—tombstone problem (for small chip discretes); No solder—lead coplanarity problem; no solder—large heat sinking, no flux; Icycle—insufficient flux for wave soldering; re-reflow problem—especially for high lead-count fine-pitch packages; low solder—weak against temperature cycling; excessive voids.

7.3.5. Failure Mechanisms at Shelf Level Include

Mechanical: poor tolerance—misalignment, structural failure, loose interconnection, connector fall off.

Thermal—improper cooling: nonuniform airflow, blocking airflow—some circuits too hot; excessive back pressure; airflow leakage.

7.3.6. Failure Mechanisms in Material Handling Include

Shelf life, electrolytic capacitors (oxide dielectric dissolves); lead solderability (oxidation); moisture absorption—popcorn; connector gold finger contamination; gouged thick film; vacuum pickup, kitting, board flexure; bent lead; mechanical shock—crystals/oscillators; ESD/EOS.

7.3.7. Failure Mechanisms in Fiber Optics Include

Component: output power degradation; facet erosion; dark-line defect—nonradiative recombination due to dislocations; misalignment at fiber-component connection–assembly problem: mounting, soldering, welding, fiber jacket shrinkage—connector misalignment.

Fiber: microcrack growth; fiber connection misalignment; fatigue-stress from fiber bend, stress corrosion; coating problem, moisture.

Fiber cut: digging ground, shark bite, animal.

Static stress fatigue from surface damage: surface cleanliness—fiber manufacture process, OH environment, minimum bending radius.

Passive components (splitters, couplers): failure modes of active components are absent but need high yield, high reliability interconnections.

7.3.8. Failure Mechanisms in Flat Panel Displays Include

Polycrystalline Si defects; rubbing process problem (line mura); cell gap problem (uniformity, chromaticity, flicker, contrast variations); spacer balls/backlight nonuniformity; high concentration of spacer balls (spot mura); backlight variations (color defects); polarizer misalignment; Na ion; open/short in array (pixel, line defect); electrical bias (flicker); close grouping of pixel defects (cluster); poor bonding, defective driver (driver related pixel and line defects); fillport contamination (fillport mura); fillport seal degradation (moisture leak to liquid crystal); uncured epoxy reacting with liquid crystal (edge mura).

7.4. RELIABILITY PROGRAMS AND STRATEGIES

Reliability needs attention in every stage of product development and manufacture. In the design process, one may build in enough design margins to avoid failures or make the design more robust against failures. In the manufacturing process, one may tighten process variations or use more robust processes to avoid failures. In the testing process, one may test whether the products have achieved adequate level of reliability and gather useful information needed to avoid future failures.

Many different reliability methodologies exist. As failure mechanisms are interdisciplinary, caution is needed to extrapolate beyond the bounds of any successful program. In the earlier days of electronics (before 1980's), much emphasis was to achieve long-term reliability. The studies for long-term reliability were primarily using elevated temperature to accelerate a large class of failure mechanisms in electronics that obey Arrhenius equation. As those failure mechanisms were under control during the IC fabrication process, the additional failure mechanisms introduced in the electronic packaging processes are different and do not need to follow Arrhenius equation. In addition, elevated temperature is not the only important acceleration factor. Many other stresses need to be taken into account.

While the field of long-term reliability is generally mature, infant mortality is less well understood. Early failures are now presenting more problems in some products. In addition, many products have a low cost requirement. The additional cost of built-in redundancy may be perceived as incompatible to cost reduction efforts. Some vendors also tend to focus on short-term cost reduction and overlook the effects of reliability on the long-term business.

Although various reliability programs are available for implementation at different stages of development for any given product, resources for reliability programs may be limited. It is therefore a challenge to allocate these resources to implement the most cost effective reliability programs. Reliability engineers generally agree that pro-active reliability measures are more cost effective and deserve higher emphasis. Yet in the commercial

world, the measures to fix problems in response to field returns and test data may be more common, even though such measures are more expensive than the pro-active measures. The reason is that the measures needed to improve product robustness may be cost effective only for the full-stream cost in the overall product cycle. The need to invest in pro-active reliability programs at an early stage of product development is therefore often not obvious. Pro-active reliability programs are consequently candidates for cancellation in this Age of Cost Cutting. Even when pro-active reliability programs have been approved, the available resources are limited and one still needs to optimize the effectiveness of the investment in reliability.

Both short term and long-term reliability strategies are needed for every product family. Yet, the current state-of-the-art for the reliability programs of many products is lagging behind the technologies that produce these products.

The field of reliability needs much exploratory work to advance. As the world turns to focusing more on short-term investments, support from companies and the government for more forward looking work are desired. Publications, sharing of information and discussions among different companies are fruitful. While the research efforts of individual companies may be limited, sharing of information among these companies provides a pool of resources to advance this field. If this field advances faster, a better understanding of the basic science and more effective reliability programs will become available to all, creating a win-win situation for these companies.

7.5. PRODUCT WEAKNESSES AND STRESS TESTING

7.5.1. Why do Products Fail?

The causes of failures are product dependent and are numerous for any given product or process. A partial list of failure modes are given in the previous section. Yet, the causes of hardware failures may be categorized as problems in design, materials, manufacture, test and field use, as shown in Figure 7.1. The causes of software failures may be categorized in a similar way.

The following is a systematical understanding of failures and stress testing without being complicated by the specific behavior of different products and their numerous failure behaviors.

Any unit of a product at any stage during development and manufacture may be grouped according to the explicitness of their defects. Thus, these units may belong to the good (solid line), weak (dashed line) or bad (dotted line) groupings, as shown in Figure 7.2.

A product unit can be thought of as going through a process, such as design, manufacture or a test process at every stage of development and production. Each process can be modeled as a process operator. The product unit may either pass or be dropped out in passing through an operator. For each product unit that passes the operator, it may undergo a transition from one group (e.g., good) to another group (e.g., bad).

Figure 7.3 shows three types of operators: process operator, ideal stress operator, and ideal test operator.

Good product units meet all design objectives under all nominal stress conditions. Bad product units have hard defects or parametric deficiencies. These products are, in principle, detectable although they may escape the tests because practical tests are non-ideal and do not have 100% test coverage. It is important to note that an ideal test with 100% test

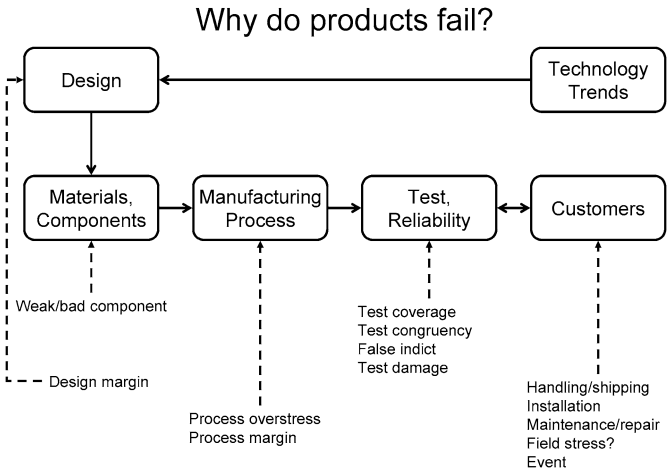


FIGURE 7.1. The causes of hardware failures may be categorized as problems in design, materials, manufacture, test and field use.

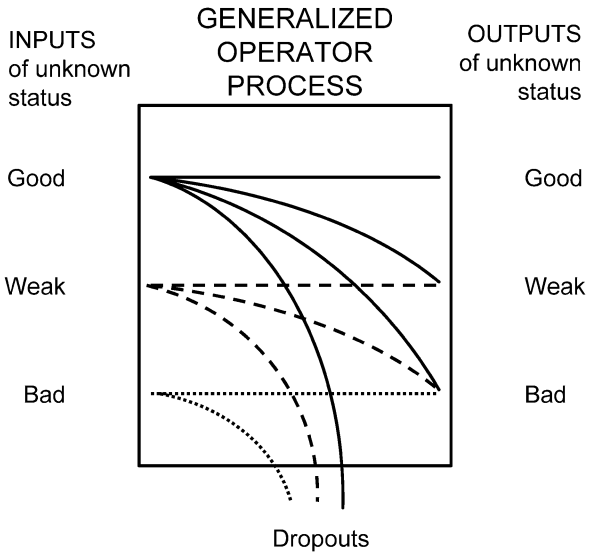


FIGURE 7.2. A model to categorize each product as in the state of good, weak, or bad in different stages during development and manufacture. Each stage is an operator that may change the state of the product.

coverage can only detect all the bad units but as will be explained later cannot detect the weak units.

Weak product units may be degradable or marginal. Marginal products may fail intermittently or only under certain stress conditions. Degradable products have latent defects and may degrade irreversibly into bad or marginal products. The degradation may be stimulated or accelerated by certain stresses.

The ideal stress operator will change the state of all weak units into the bad state during the stressing process.

In software reliability, a hard defect is a fault and a product weakness is a dormant fault.

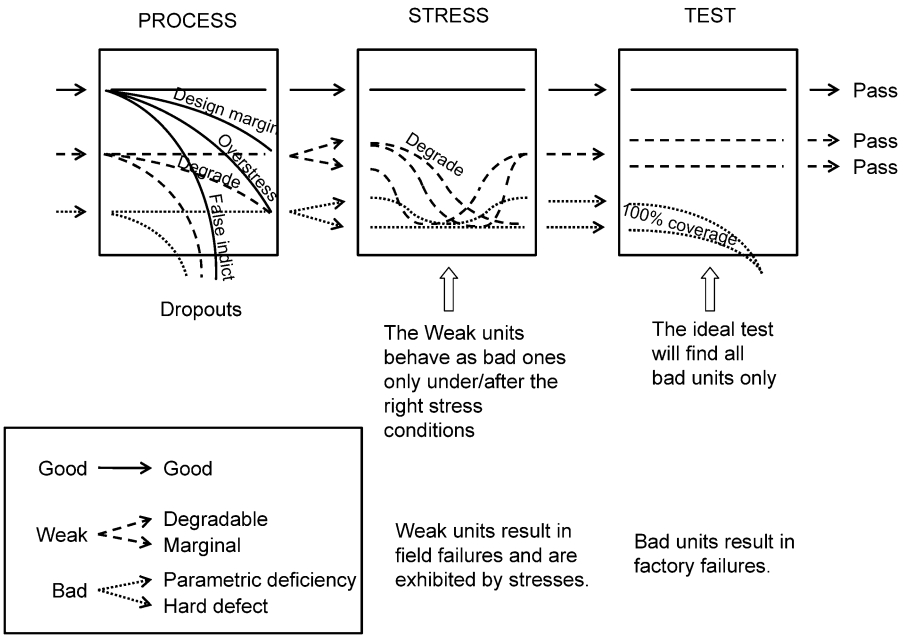


FIGURE 7.3. Three types of operators: process operator, ideal stress operator, and ideal test operator.

The process operator in Figure 7.3 may be any process such as design, development, or manufacture. It shows that both weak and bad products are produced with problems in design margin, manufacture, assembly, or material quality control.

Let us define an ideal fault-detection process as one that does not stress the product and yet achieves 100% test coverage. In the absence of stress, only the bad product units are detected. Weak product units are not detectable. They pass these tests and are shipped to customers, resulting in early failures in the field. Weak product units may degrade or exhibit failures only under certain stress conditions.

We also define an ideal stress process as one that degrades all degradable product units and shows up all marginal units without damaging the good ones. The exit from the ideal stress operator shows what happens to the units upon the removal of stresses. Here, the formerly degradable units remain in the bad state whereas the formerly marginal units return from the bad state to the weak state.

7.5.2. Stress Testing Principle

The principle of stress testing is shown in Figure 7.4. We have conceptually separated test into stressing and fault-detection with 2 separate roles. 100% test coverage refers to successful fault-detection of all bad units, whereas an ideal stressing process turns all hidden faults in the weak units into detectable faults, at least temporarily. The ideal stress-testing operator is simply the cascade of the ideal stressing and the ideal fault-detection operators. This cascaded operator shows that testing with the above stresses will find both weak and bad product units. Corrective actions may then be taken.

Notice that the cascaded process includes testing both during and after the stressing process. The weak product units that are marginal need to be tested while they are subjected

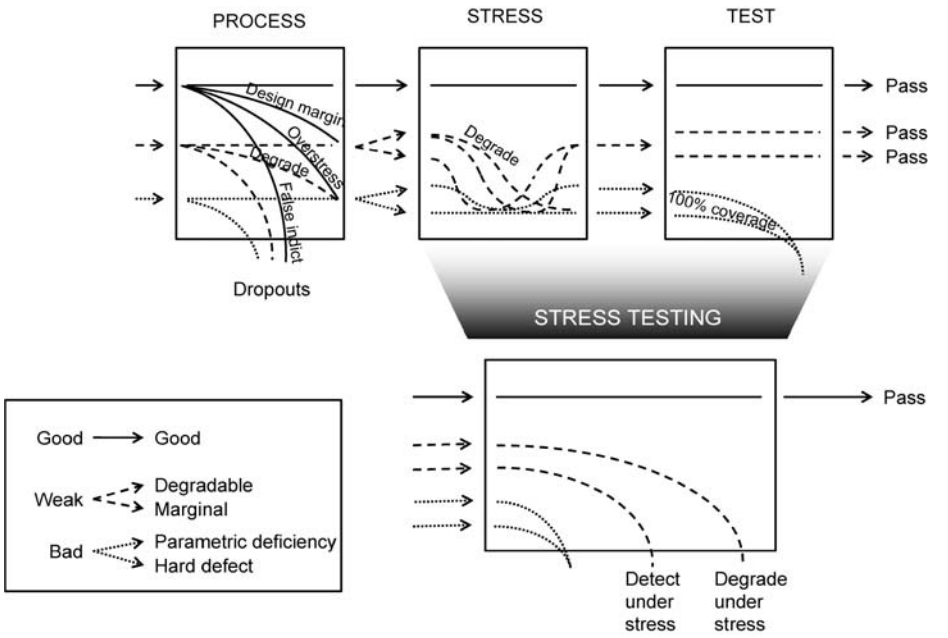


FIGURE 7.4. The principle of stress testing is shown by cascading the ideal stressing and the ideal fault-detection operators to obtain an ideal stress-testing operator.

to the proper stress conditions. One the other hand, the degradable product units may be tested after the stress conditions have been applied to degrade them into bad units.

7.5.2.1. Screening versus Corrective Actions Stress tests include screening without corrective actions and stress testing with corrective actions, shown in Figure 7.5.

Environmental Stress Screening (ESS). ESS applies stresses to stimulate observable failures for weak units. The stress level applied during screening needs to be carefully adjusted to achieve an efficient screening process. A mild test regimen may pass too many weak units whereas too much stress may damage good ones. Note that screening only screens out weak units. It cannot improve the strength of the rest of the units. Although one can reduce field failures by rejecting the weak units, a draw back of screening is that the weak units continue to be manufactured.

Accelerated Stress Testing with Corrective Actions (AST). In AST, stresses are used to identify the weaknesses of the product and the root causes of these weaknesses are investigated. Corrective actions are then taken to avoid producing these weaknesses through achieving robust design, excellent component quality, and robust processes. Once these objectives have been met, product reliability is assured without screening. Yet, AST on a sampling basis (Production-Sampling AST) to monitor and to maintain high quality may be required.

AST with corrective actions is generally most valuable during the design and qualification stage of a product (Design-Qualification AST). ESS may be applicable (Ongoing ESS) when a manufacturing process is new or has not been improved sufficiently. For example, some high performance VLSI IC components are manufactured using the newest technology, which is not yet mature, and tends to have low yield. In this case, ESS at the earliest possible level is useful to screen out the weak units.

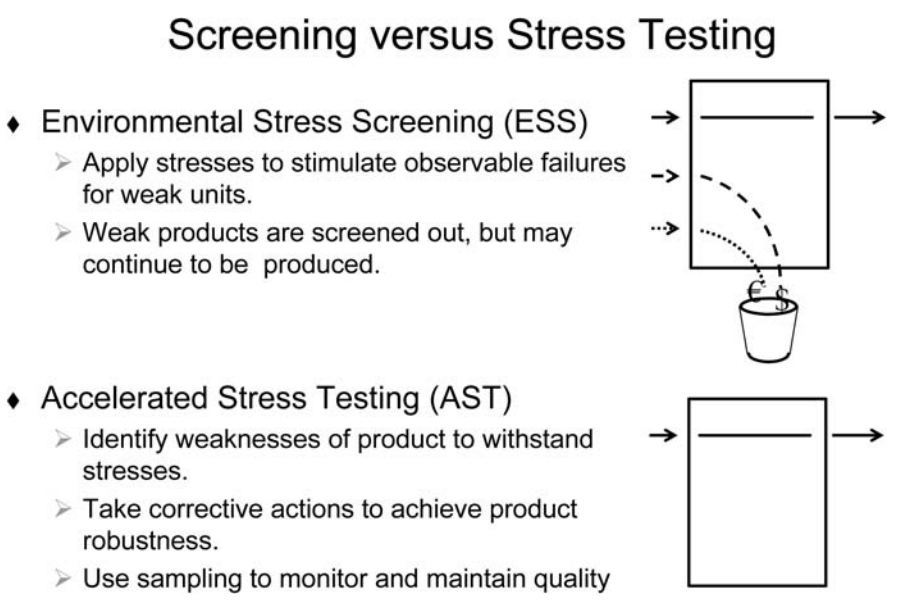


FIGURE 7.5. Comparison between stress screening and stress testing with corrective actions.

The concept of AST in hardware reliability may be extended to software reliability, as well as the reliability of a system where hardware weaknesses and software faults are interacting. Exceptions include the concept of screening, which is not applicable to software faults alone when all the product units have the same copy of software or operating system. The purpose of screening in this case will be to look for hardware and system weaknesses that show up under certain combinations of hardware and software stress conditions.

7.6. STRESS TESTING FORMULATION

7.6.1. Threshold and Cumulative Stress Failures

Products may fail in different ways when they are subjected to stress. The appropriate way to perform AST on a product depends on the nature of the failure modes under consideration. There are many failure modes, but in terms of stress-failure, there are three principal types.

7.6.1.1. Threshold Stress Failure A product typically exhibits a statistically distributed strength to withstand a given stress. In the simplest case, a failure occurs in a specific product sample when the stress level exceeds the threshold strength of the sample. A simple example of the threshold strength is the mechanical strength of a brittle material. Other examples are the threshold power to burn a device and the threshold electric field for avalanche breakdown to occur. A product with low-threshold strength may fail when a high maximum stress occurs during assembly, handling, transportation, installation, or field operation. On the other hand, a product with high-threshold strength is robust to resist these stresses. A screening process at a certain stress level may sort out the weak product units without damaging or wearing out the good ones, which possess threshold strength higher than the screening stress level. Yet, the product yield in screening is an also important consideration.

7.6.1.2. Cumulative-Stress Failure Some failures occur under the cumulative effects of repeated low-level stresses. Although these stresses may be well below the level that will cause immediate product failure, each time they are applied they produce some irreversible changes in the product. Thus, over repeated exposure to these stresses, the cumulative change grows to a level that results in product failure. Well-known examples of cumulative-stress failure are electromigration in a device under high-current density, and fatigue failure of solder joints due to extended thermal cycling. Screening weak products by applying stresses is not always appropriate for these failure modes because the cumulative effect of the damage produced in the units during screening will also shorten the product life of the good units.

7.6.1.3. Combined Threshold—Cumulative Stress Failure A combination of threshold stress and cumulative stresses may also stimulate certain failure modes. Here, a high maximum stress starts an incipient failure site that is later driven to a hard failure by cumulative stresses. For example, a device with a cracked package is more vulnerable to corrosion. Use of high stress to screen out such weak products needs caution. Exceeding the threshold stress alone may not show an immediately detectable failure. If the incipient failure site induced by exceeding the threshold strength is not driven to an observable failure, early field failure may occur.

A product that is robust to a large threshold stress, e.g., a big mechanical shock, may also be weakened by cumulative stresses, e.g., vibrations. After that, it may become vulnerable to a relatively smaller shock.

7.6.2. Stress Stimuli and Flaws

A variety of stress stimuli can be applied to a product as part of an AST regimen. We will review the most commonly used stimuli here, and note some of the types of product deficiencies they are likely to stimulate to failure.

Elevated temperature: testing a product for an extended period at an elevated temperature, or burn-in, is probably the most common form of stress testing. Marginal product designs often exhibit a temperature threshold above which the product will not function satisfactorily, and failure modes that involve chemical or diffusion processes can often be effectively accelerated at elevated temperature. However, for many electronic products there is a tendency to only test the product to the nominally specified upper temperature limit. By testing beyond the nominal temperature limit, one can better assess the robustness of the product.

Power cycling: turning a product on and off is another common form of stress testing; it is commonly done in conjunction with other types of stress testing. The temperature transients that occur during power-up can often stimulate thermal-mechanical defects. For electronic systems, the possibly variable conditions resulting from an abrupt shut-down or from on/off powering may also reveal design deficiencies.

Temperature cycling: a number of interconnection and packaging failure modes may best be stimulated by temperature cycling. The higher transient temperature conditions that occur during temperature cycling can also reveal design deficiencies not normally found during the slow transients normally associated with getting up to burn-in temperatures. Cycling to low temperatures can also reveal temperature threshold problems not covered by traditional burn-in.

Voltage variations: varying the voltage supplied to an electronic system can reveal design margin problems and marginal performance of specific components. This type of

stressing is often combined with testing at temperature limits to increase the detection of marginal conditions.

Clock variations: varying the clock rate can reveal timing margin problems. This type of stressing is often difficult to achieve because of the high degree of integration of clocks in the product circuitry, and special design provisions must be made to make it feasible.

Vibration and mechanical shock: vibration or mechanical shock stressing has traditionally been used to reveal structural support problems, problems in the securing of specific (often large and/or heavy) components, and connectorization or cabling problems. In addition, problems with surface mount solder joints are also increasingly being addressed using vibration testing.

Elevated humidity: elevated humidity testing is usually done in conjunction with high temperature testing to reveal problems with corrosion or high voltage isolation breakdown. Extended testing is often required to get results.

Electrostatic discharge (ESD), power surge: specialized testing is often done for ESD and power surges to check the adequacy of isolation and grounding designs. Such testing is generally not performed on product units that are shipped.

Electromagnetic interference (EMI) susceptibility: EMI testing is also often done to verify design robustness, though often only at ambient conditions. However, EMI testing done in conjunction with temperature cycling can reveal additional problems with leakage paths not found at static conditions.

Software stresses: an excess load may be applied to a system while new software is being tested. Servers and database systems may be tested at high network usage situations that may overload the system. Besides extreme values, out-of-bound or invalid data may also be input.

Combination and order of stresses: combinations and interactions of any lists of stresses from above often produce new failure modes not observed with any single type of stress. The order of these stresses can also produce different modes. One way to pick the combinations and order of these stresses is to investigate the "Use Cases". Good questions are how and under what conditions the system will be used, and what can happen.

7.6.3. Modes of Stress Testing

We discuss several modes of doing AST, in the order of descending desirability.

Design-Qualification AST: Product is tested near the end of the design stage to see if it is robust with respect to stress levels in excess of those likely to be encountered in the use environment. If deficiencies are found, the root cause is determined and corrective actions are taken to fix the underlying causes of the problem. This is the most productive mode of doing AST because the benefits are realized over the whole life of the product.

Manufacturing-Qualification AST: A representative sample of product is subjected to AST during manufacturing ramp-up to identify deficiencies in component quality or manufacturing processes. In addition, design margin deficiencies that did not show up in design qualification AST may be found. The emphasis again is on Failure Analysis (FA) and corrective actions, so that deficiencies in the product can be quickly eliminated before production volumes become large.

Production-Sampling AST: For products requiring high reliability, it is useful to continue to perform AST on a sampling basis to monitor the production process for manufacturing or component quality variations, even after satisfactory qualification has been

achieved. The emphasis continues to be on determining the root cause of any problems found and taking corrective action to fix them.

Ongoing AST: It sometimes occurs that AST is performed with the intention of doing good FMA and correcting any problems found, but because of a lack of sufficient resources, quality problems persist. In this case, it may still be economically feasible to continue doing AST on an ongoing basis, with some degree of FMA and corrective action to achieve a high-reliability product. Yet, this mode is certainly less desirable than the approaches mentioned above.

Ongoing ESS: Traditionally called environmental stress screening (ESS), this approach is mostly directed at sorting the good product from the bad, which is then repaired, but with little attention to FMA and corrective action. Although this mode may be suitable for limited production volumes where there is little opportunity for product improvement, it is not nearly as productive for higher volume products as the approaches mentioned above.

7.6.3.1. Acronyms of Stress Testing The following names of stress testing are ordered according to the approximate date of their first use in literature. Historically, many names have been given to emphasize its different aspects. Yet, as different companies are learning from each other, their processes will become less and less distinguishable. While these acronyms may appear different and confusing at this point, they may eventually be treated as meaning the same thing.

ESS—Environmental Stress Screening: the emphasis of ESS is in screening in the early developments of theses process in military applications, but people have begun to do corrective actions while still calling it ESS.

HALT—Highly Accelerated Life Testing: step stress to look for the upper/lower operational stress levels, and also the upper/lower destructive stress levels. Perform FMA and corrective actions to raise these levels. Iterate to raise these limits as far as possible. This term is originally used by Greg Hobbs.

HASS—Highly Accelerated Stress Screening: production screening process using HALT to establish elevated stress levels, which are usually established through HALT. This term is originally used by Greg Hobbs.

STRIFE—Stress test and improve design to achieve product robustness with longer service life. This term is originally used in Hewlett-Packard.

EST—Environmental Stress Testing: this term is originally used in Lucent Technologies, formally AT&T. The name tries to emphasize that the process is different from the conventional ESS process.

AST—Accelerated Stress Testing: this non-company specific term is originally used in naming the annually IEEE Workshop on AST. It also emphases that stresses can be general and are not limited to environmental stresses.

7.6.4. Lifetime Failure Fraction

Customers' perspective.

Reliability has different meanings in the literature. We take the perspective of the end customer here.

Knowing only the mean-time-between-failures, MTBF, is not enough for many applications. Customers need to be protected from both excessive infant mortality (time $<$ T1) and early wear-out (time $>$ T2), shown in Figure 7.6(a).

Customers expect reliability throughout the product's service life. A product is expected to properly function throughout the time it is in service. This time span defines

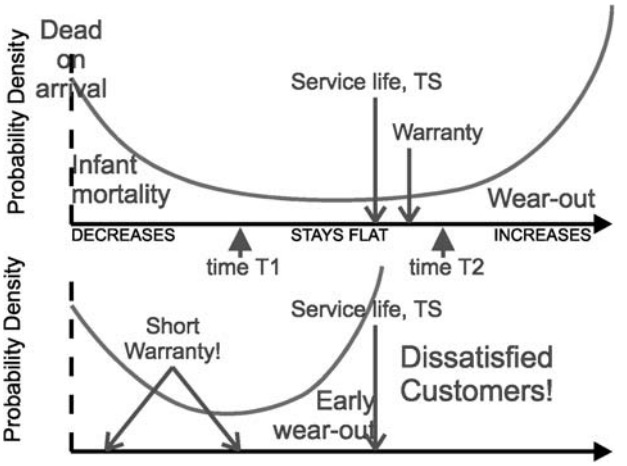


FIGURE 7.6. (a) Upper drawing: Example of a product life longer than service life. (b) Lower drawing: example of a product life shorter than service life.

the service life, *TS*, which may be up to ten years for some consumer electronics. Vendors sometimes only measure T2 only to compare with the warranty period, because of the warranty repair costs incurred by the vendor. However failures that occur after the warranty period continue to upset the customers. If a problematic product with a short T2 [Figure 7.6(b)] is marketed, the excessive failures right after the warranty period will simply turn future customers away. To prevent early wear-out, T2 needs to be large than the expected service life.

Reliability may be defined as the avoidance of failures, which customers see over the life of a product. The ability of a product to properly function just under the conditions specified by the vendor is not enough. If the product cannot withstand the various stresses that it may encounter in a customer's environment at any time during its service life, the customer will experience a product failure.

A convenient measure of reliability is the cumulative fraction failed over the product's service life. A traditional measure of a product's reliability is the (instantaneous) failure rate, which is the (instantaneous) probability density function $f(t)$ at age t divided by the fraction that has not failed. This failure rate is time-dependent and is complicated by infant mortality and early wear-out. For consumer electronics, all the product failures occurring from time $t = 0$ to $t = TS$ are important to the customer. Here, *TS* is counted from when a customer buys a product up to the time the customer replaces or forgets it. A simple measure of a product's reliability is the (time-independent) fraction failed over the product's service life, which is defined as

$$F(TS) \equiv \int_{t=0}^{t=TS} f(t)dt,$$

and is shown in Figure 7.7.

7.6.5. Robustness Against Maximum Service Life Stress

Product strength and lifetime maximum stress.

Examples of product strength distribution and lifetime maximum stress distribution are shown in Figure 7.8.

For threshold stress failures, each unit of a product needs to survive not simply the nominal stress but all the peak stresses during the product life. The highest peak stress encountered over the product life is defined here as the lifetime maximum stress, X. The maximum value of X that a specific product unit can withstand without failing is a measure of the robustness of that unit, and is called the product (yield) strength, Y.

For cumulative stress failures, one picture is to look at the instantaneous strength as being weakened over time by the (time dependent) instantaneous stress. Yet, it is desirable to skip the details of time dependence here and use the same unified formulation for both threshold and cumulative stress failures. This is achieved by defining the lifetime cumula-

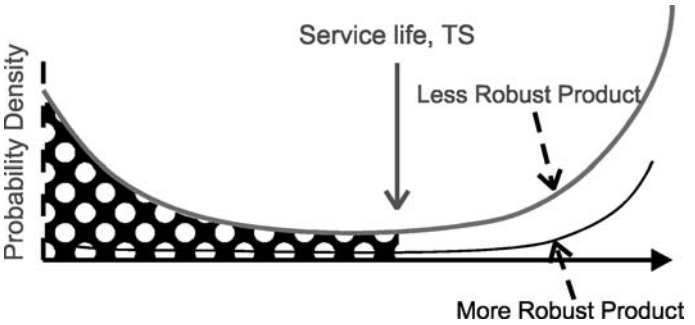


FIGURE 7.7. Fracture failed over service life is the area marked under the curve.

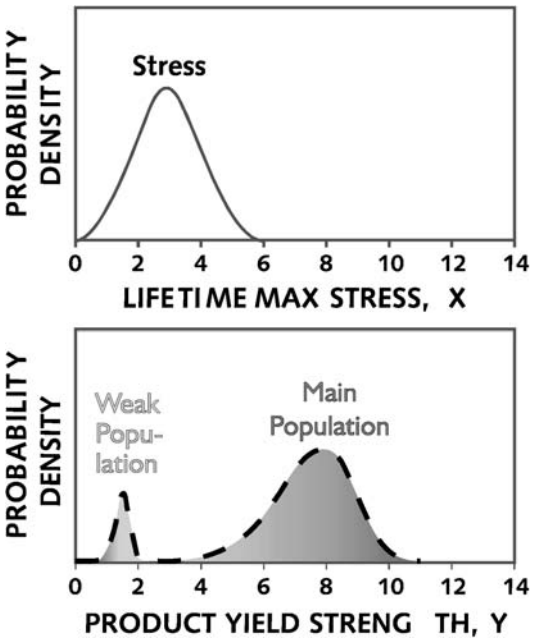


FIGURE 7.8. (a) Upper drawing of the probability density distribution of lifetime maximum stress. (b) Lower drawing of the probability density distribution of product yield strength.

tive stress, X, as an effective stress parameter that is proportional to the overall weakening of the instantaneous strength by a physical stress over the product life. The corresponding initial product strength against such an X is again denoted by Y. An example is 85°C, 85% humidity and under 5 V bias for 1000 hours.

The distribution of the lifetime maximum stress is determined by the customer's environment, whereas the product strength is a statistical distribution of the product units.

The product strength distribution usually has one or more weak sub-populations in addition to a main population. The weak sub-population is generally a main contributor to freak failures and infant mortality, which show up in a typical bathtub curve. Most units in the main population possess enough design margins to withstand incurred stresses. Therefore, for the case of threshold stress failure, they generally fail only under extreme stress conditions such as lightning-surge or electrostatic discharge (ESD). For the case of cumulative stress failure, they may fail only under true long-term wear-out. Yet, those falling in the low strength tail of the main population may also contribute to early wear-out.

The presence of weak sub-populations separated from the main population is consistent with the bathtub curve commonly observed in many products, and may model many hardware systems. Yet a wide distribution of the main population can also give arise to higher failure rate in the infant mortality stage. This latter category may be more appropriate for software weaknesses.

7.6.6. Stress–Strength Contour

The occurrence of field failures is determined by the distributions of both the stress X and strength Y. It is therefore convenient to look at the contour map of this joint probability distribution of X and Y (Figure 7.9).

In this contour map, the population lying in the $Y < X$ region will result in field failures whereas those in the $Y > X$ region will not fail in the field. An example of a weak population is one that lies mostly in the $Y < X$ region, while a tiny fraction on its left

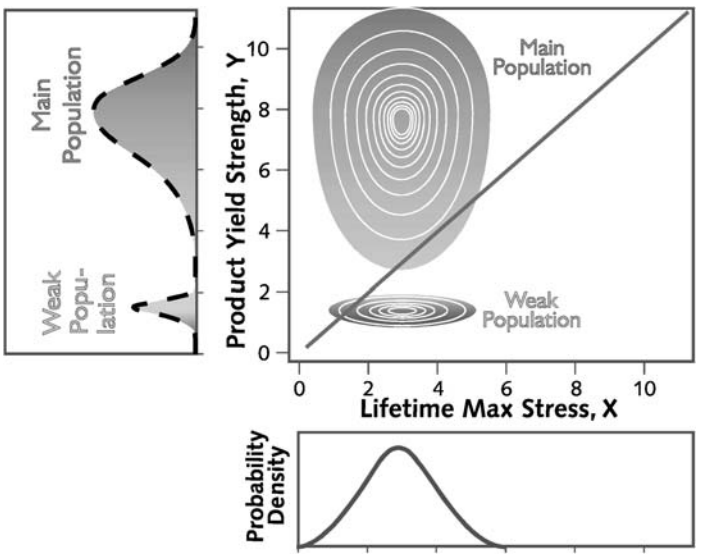


FIGURE 7.9. Joint probability density distribution of lifetime maximum stress and product yield strength.

does survive a benign stress environment. An example of a main population is one that lies mostly in the $Y > X$ region, but its lower-right part still falls into the $Y < X$ region. Then, even though the nominal strength of this main population may seem high enough compared to the nominal maximum stress, a significant fraction may still fail. These failures are owing to the statistical spread in the distributions of both the stress and the strength.

7.6.7. Common Issues

7.6.7.1. How Does Stress Testing Affect the Product? We define a maximum AST stress, X^{ST}, which is analogous to the definition of lifetime maximum stress. For threshold stress failure, it is the maximum stress applied to a unit during ESS or AST. For cumulative stress failure, it is the effective stress parameter proportional to the overall irreversible change that the stress testing process has made on a unit. Then the effects of stress testing at a given maximum stress level, X^{ST}, are shown by a dividing line on the product strength into a region of stress test failure below this line and a region of stress test pass above it.

In the stress-strength contour shown Figure 7.10, the solid horizontal line shows an X^{ST} level that catches all the weak population. Yet, it still does not catch the lower-right part of the main population. This part will fail in the field because it happens to experience a higher level of lifetime maximum stress. We may eliminate this part of the population if we raise X^{ST} to the level shown by the dashed line. Yet we would then also fail a significant portion of the main population in the lower-left part. For ESS, the yield will then be too low so that the screening process is not economical. Thus, the separation between the weak population and the main population must be large enough for a working window for screening to be feasible. For AST during the design stage of a product, the purpose of AST is to find weaknesses in design and manufacturing and to take corrective actions.

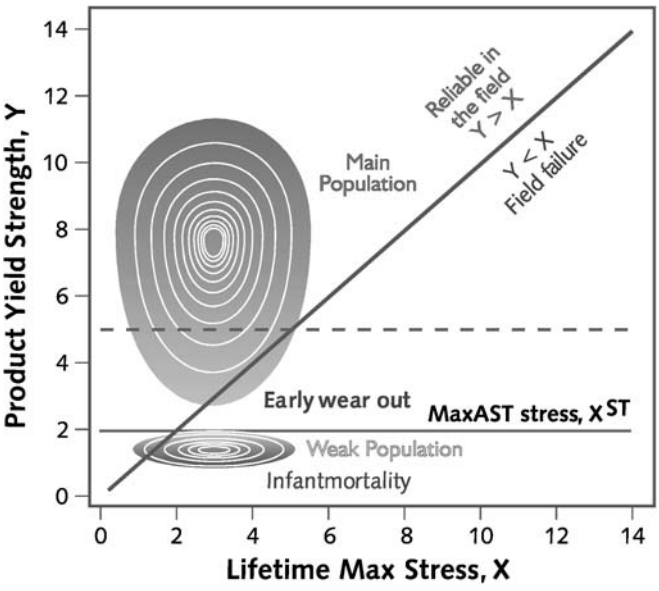


FIGURE 7.10. The effects of stress testing at a given maximum stress level, X^{ST}, are shown by a dividing line on the product strength into the lower region of stress test failure from the upper region of stress test pass.

Some failure modes may have a somewhat low field failure rate but are still not acceptable for a reliable product. The use of moderate stress during AST will not be an effective way to find them. A higher level of such as the one shown as the dashed line in Figure 7.10 is a more effective way to find these weaknesses. This holds if those failure modes in the field will also occur with the higher X^{ST}.

7.6.7.2. Will Stress Testing Damage Good Products? This is usually a concern for ESS only, because the purpose of Design and Manufacturing Qualification AST is to effectively identify and correct potential problems. The corrective actions are essential to achieve the robustness such as the one shown in the contour in Figure 7.11, where the robust main population is safe against stress testing, including Production Sampling AST. The prerequisite of having a robust main population is important. Indeed, incorrect application of screening without first meeting this requirement may damage more weak units to catch the weaker units. The result from such an improper stress-testing program may mislead people and cause them to step back to the use of mild stresses for all stress testing programs. Figure 7.10 shows such a non-robust product, where screening with either a mild stress or an elevated stress cannot improve its robustness. Corrective actions are essential here.

For threshold stress failures, stress testing does not affect the good products, even for Ongoing ESS. When a stress level X^{ST} is used in stress testing, the weak units whose threshold strength is below X^{ST} are detected. The good units whose threshold strengths are above X^{ST} are not weakened by the stress testing process.

For cumulative stress failures, the product must possess a robust main population that is well separated from the weak populations before screening may be applicable. For example, consider a product that has a main population with a robustness of 1000 stress

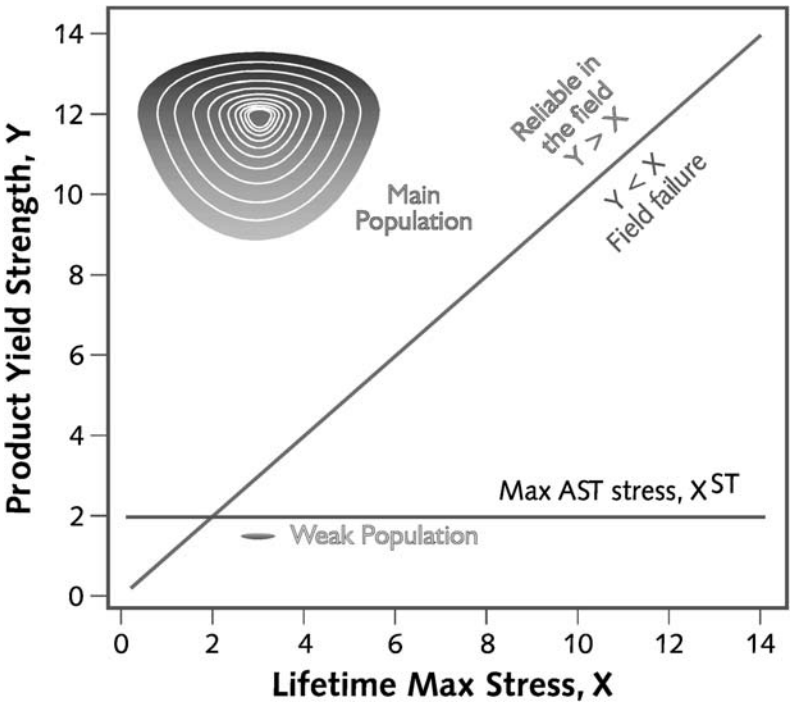


FIGURE 7.11. A more robust product than that represented by Figure 7.10.

units, (Ux), but its weak population is mostly below 10 Ux in strength. An X^{ST} of 10 stress units will screen out most of the weak units but the strengths of all the units are then weakened by 10 Ux after screening (Figure 7.11). Suppose there is a weak unit that has strength in the 100–110 Ux range before screening and will experience an X^{ST} of 100 Ux in a certain customer's environment. The dilemma is that this weak unit will encounter field failure but it could have escaped from it if it had not been screened. The answer to this dilemma is to compare the $F(TS)$ for the units that have been screened to the $F(TS)$ for those that have not been screened. When the main population is very robust, the weak populations may be screened out at the expense of a small decrease in the useful life of the product. Because there are far fewer units in the 100–110 Ux range than in the 0–10 Ux range, screening will decrease the cumulative failure fraction over the product life $F(TS)$ in the field for all the units.

7.6.7.3. Safety Testing AST should precipitate flaws in marginal products before they are shipped from the factory, but should not induce flaws or failure modes that normally would not be present in good products. The useful life of a product also should not be diminished by AST. To prevent this from happening, safety testing should be applied to the candidate regimen.

The preferred safety test method is to repeatedly apply a candidate stress testing regimen X^{ST} to a product until failure occurs at some level, X^{ST}. For example, if a candidate regimen of 10 thermal cycles is proposed and it is observed that on the order of 1000 thermal cycles is needed to eventually break the product, one may reasonably conclude that since $X^{ST} = 1000$ is $\gg$ than $X^{ST} = 10$, the regimen will not significantly reduce the useful life of good products.

In using highly accelerated stress testing, it is also necessary to perform sufficient safety testing on potential combined threshold-cumulative failure modes to be sure that the elevated stress levels are not causing incipient damage that can lead to later failures due to the cumulative effect of a secondary stress. As mentioned earlier, the corrosion

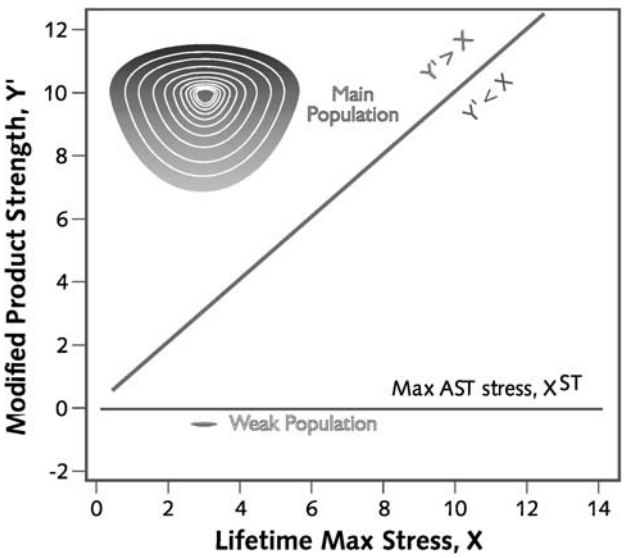


FIGURE 7.12. A more robust product after subjecting to cumulative stress.

of metalization within a device package cracked by an elevated mechanical stress is an example of such a concern.

The burden of doing safety testing is greatly diminished as one gains experience, since the results from safety tests for a technology used on one product may later be applied to other products using a similar technology. It is important to develop a mechanistic understanding of the relationship between applied stress conditions and observed failures so that one can properly judge when it is appropriate to extrapolate from previous safety testing results and when additional testing is required.

Another requirement of using a much higher stress level than those encountered in the nominal operating conditions is that the higher stress level should only increase the probability of the same failures that occur under the lifetime maximum stress conditions. The failures that take place at different values X^{ST} should be of the same failure mode although at different probabilities as evident in the strength distribution. The limit on elevating the stress used is that new physical phenomenon that will not occur under the lifetime maximum stress conditions should not occur under X^{ST}.

7.6.7.4. Are Failures from AST Related to Field Failures? A common question in reliability is whether a suspected product weakness will result in real field failures. Therefore, a stress testing program should choose the types of stresses that are likely to show the same types of failures found in the field environment. Whether these stresses should be thermal, mechanical, thermal-mechanical or chemical do not have to relate to the field stresses. Rather, the considerations are whether these stresses are relevant to the types of field failure modes that are likely in the field environment.

7.7. FURTHER READING

Further readings for Sections 7.1, 7.4, 7.5, and 7.6 may be found in H.A. Chan and P. Englert, "Stress Testing Handbook for Quality Products with Case Studies in Telecommunication and Computer Products," ISBN 0-7803-6025-7, IEEE Press and John Wiley & Sons, 2001. There is a bibliography of over 300 references there. Further reading for Section 7.2 may be found in most books on reliability.

8

How to Make a Device into a Product: *Accelerated Life Testing (ALT), Its Role, Attributes, Challenges, Pitfalls, and Interaction with Qualification Tests*

E. Suhir

University of California, Santa Cruz, CA, University of Maryland, College Park, MD, and ERS/Siloptix Co., Los Altos, CA, USA

"*You can see a lot by observing*"
Yogi Berra, American Baseball Player

"*It is easy to see, it is hard to foresee*"
Benjamin Franklin, American Scientist and Politician

8.1. INTRODUCTION

"*Vision without action is a daydream. Action without vision is a nightmare*"
Japanese Saying

Accelerated life tests (ALTs) are aimed at the revealing and understanding the physics of expected or occurred failures. Another objective of the ALTs is to accumulate representative failure statistics. Thus, ALTs are able to both detect the possible failure modes and mechanisms and to quantitatively evaluate the roles of various phenomena and processes that might lead to failures. Adequately designed, carefully conducted, and properly interpreted ALTs provide a consistent basis for obtaining the ultimate information of the reliability of a product—the probability of failure.

ALTs can dramatically facilitate the solutions to the problems of **cost effectiveness and time-to-market**. Because these tests can help a manufacturer to make his device into a product, they should play an important role in the evaluation, prediction and assurance of the reliability of micro- and opto-electronic devices and systems.

In the majority of cases, ALTs should be conducted in addition to the qualification tests required by the existing standards. There might be also situations, when ALTs can be used as an effective substitution for qualification tests, especially for new products, for

which qualification requirements have not been developed yet. Whenever possible, ALTs should be used as a consistent basis for the improvement of the existing qualification specifications.

In this chapter, we discuss the role, objectives, attributes, challenges and pitfalls, associated with the use of ALTs, as well as their interaction with qualification tests. The emphasis is on the **role that ALTs** should play in the development, design, qualification and manufacturing of micro-, opto-electronic and photonic components and devices.

8.2. SOME MAJOR DEFINITIONS

"*One should always to say 'tables, chairs, glasses of beer,' instead of 'points, straight lines and planes'* "
David Gilbert, German Mathematician

The following major definitions are used in engineering reliability .

Failure mode identifies what happened (or might happen), the objective effect by which a failure is observed. Failure mode is what has been detected, observed and/or reported as a failure, whether functional (electrical, optical, thermal, etc.), mechanical (structural, "physical") or environmental (high- or low-temperature induced, high-humidity induced, radiation induced, etc.) failure. The failure mode is the evidence, manifestation, by which the failure is observed. Examples are: shorts and opens; low or distorted output signal; high optical losses; low coupling efficiency; material's failure; loss of structural integrity; brittle fracture, etc.

Failure mechanism identifies what phenomenon or process resulted in a failure. Such a phenomenon or process could be of physical, chemical, mechanical, thermal, or technological nature. Examples are: voltage breakdown, corrosion, fatigue, material's aging, electro-migration, excessive heat, elevated stress, high level of current, initiation and propagation of cracks, excessive displacement, division by zero, etc.

Failure site identifies the location of the failure, i.e., where the failure has occurred.

Load is the mechanical, electrical, thermal, chemical, or physical condition that is able to precipitate failure.

Stress is the level (intensity) of the applied load at the given failure site. Stress does not have to be mechanical, but could be due to an electrical, optical, thermal or other phenomenon.

"Root" cause identifies why a particular failure occurred. Examples are: poor design, selection of an inappropriate material, overstress, use of an inadequate technology, manufacturing deficiencies, misuse or abuse of the employed equipment, human error, etc.

One is supposed to possess a "gut feeling" on what can possibly "go wrong" and should make a preliminary decision on how to detect if something is "going wrong" indeed (in terms of the methodologies and measuring equipment used, qualification of the personnel, available resources, etc.). The factors that are being considered in the experimental design effort can and should change with the changes in the materials and technologies, as well as with the changes in the market demand and competition.

8.3. ENGINEERING RELIABILITY

"*Reliability it is when the customer comes back, not the product*"
Unknown Reliability Manager

Reliability is the ability of a product to be consistently good in performance, and so to elicit trust of both the manufacturer and the customer. Reliability Engineering deals

with failure modes and mechanisms, "root" causes of occurrence of various failures, and methods to estimate and prevent failures. In products, for which a certain failure rate is considered acceptable, Reliability Engineering examines ways of bringing this rate down to an allowable level. A reliable item or a system are able to survive and to satisfactorily perform a required function, without failures or breakdowns, for a specific envisaged period of time under the stated operation and maintenance conditions. In **probability-based reliability engineering**, reliability is defined as the probability of an item or a system to function in the above indicated way.

If reliability is not defined and taken care of, the device will never be able to operate in accordance with the customer expectations and specifications, and the manufacturer will never be able to assess if/which state-of-the-art technologies would enable him to fulfill the customer's reliability and quality requirements.

It goes without saying that, in a sense, reliability related failures are similar to the consequences of a fire: it is easier to determine them in advance and to take measures to prevent them, rather than to eliminate their consequences and fix the design. If the reliability related bottlenecks of a particular design are anticipated and assessed well in advance, then the manufacturer would be able to compare various competing designs, manufacturing technologies and available metrological means from the viewpoint of the product's reliability and cost, and to establish the most feasible trade-offs between the reliability and guaranteed warranty.

8.4. FIELD FAILURES

"*Nothing is impossible. It is often merely for an excuse that we say things are impossible*"
Francois de la Rochefoucould, French Philosopher

The information of a failure could be obtained at different times, by using different means and at different locations. As far as "when" is concerned, the failures could be detected (observed) during the fulfillment of qualification tests, during screening tests that are carried out in the process of manufacturing, during burn-in tests, etc. As to "where" the information is obtained, it could be carried out at the vendor's site or at the customer's site, in the laboratory or in the field.

Field failures play a special and an important role in reliability engineering. It is "life itself," usually not distorted by accelerated test conditions. Field failure analysis is carried out on products recovered from the field after failure. It is the information from the field failure that is the most valuable in terms of the ability to accurately predict the likelihood of failure in the field, i.e., in actual use conditions.

It would not be an exaggeration to say that while the reliability engineer "supposes" what, when and why might fail, the field "disposes," and therefore a reliability engineer can often learn more from what happened in the field than from what has been observed in the lab or at the manufacturing site. An ironical thing is that although the reliability engineer does not want that field failures happen at all, he/she could learn a lot from an actually occurred field failure.

Field failure analyses require close cooperation between the customer and the vendor, as well as between designers, part suppliers, assemble line, and all those who in one way or another are involved in the product manufacturing and supply. Human error is often a cause of a field failure, and therefore it is often the human ego that is an obstacle to the

analysis of a field failure. Field failure analysis requires the creation and availability of a user-friendly flexible and informative data base.

Mechanical failures are the most common type of failure in micro- and opto-electronic and photonic devices and equipment. The overwhelming majority of field failures are often related to a particular failure mechanism (the "weakest link") and to a small number of insufficiently reliable components.

8.5. RELIABILITY IS A COMPLEX PROPERTY

"*Truth is rarely pure and never simple*"
Oscar Wilde, The Importance of Being Earnest

Reliability is a complex property. It includes the item's (system's) **dependability**, **durability**, **maintainability**, **reparability**, **availability**, **testability**, etc. Each of these qualities can be of a greater or lesser importance depending on the particular function and operation conditions of the item or the system. In this chapter, we include quality, which is typically associated with manufacturing and production, into reliability, which is typically associated with materials and design only, but could be treated in a much broader sense, if one considers also the consistency (predictability, stability, repeatability) of the manufacturing processes, reliability of software, etc. In other words, reliability can be defined as the probability of a certain specified and unacceptable deviation from the pre-established functional, mechanical, environmental or some other type of performance.

8.6. THREE MAJOR CLASSES OF ENGINEERING PRODUCTS AND MARKET DEMANDS

"*Plus usus sine doctrina, quam citra usum doctrina valet*"
("*Practice without theory is more valuable than theory without practice*")
Latin Proverb

The following three major classes of engineering products could be distinguished, as far as their objectives and requirements for their reliability are concerned:

Class I. The product has to be made as reliable as possible, and failure should not be permitted. Such products are being typically manufactured for the military and similar "markets," when cost is not viewed as the most important factor, and the object has to be made reliable by all means. Such "products" (some warfare, military aircraft, battleships, space apparatus, etc.) are seldom manufactured in large quantities, and their failure is viewed as a catastrophe. The consequences of failure of the Class I products are the most severe ones, and can be associated with bad publicity, legal actions and even with the country's security and/or prestige. These "products" usually have a single customer, such as the government or a big firm. Traditionally, the reliability requirements for these products are defined in the form of government standards, such as US Military Standards (MIL-STDs). These standards not only formulate the reliability requirements for the product, but also specify the methods that are to be used to prove (demonstrate) the reliability, and even prescribe how the product must be manufactured, tested and screened. It is the customer, not the manufacturer, who sets such reliability standards.

Class II. The product has to be made as reliable as possible, but only for a certain level of demand (load). If the actual demand happens to be larger than the design demand, the product might fail, although the probability of such a failure should be (made) very small. Examples are civil engineering structures, passenger elevators, ocean-going vessels, offshore structures, commercial aircraft, railroad carriages, some medical equipment, etc. These are typically highly expensive products, which, at the same time, are produced in large quantities, and, therefore, application of Class I requirements to such products, important as these products might be, will lead to unjustifiable and unacceptable expenses, which are, for this reason, deemed to be economically unfeasible. The failure of products of Class II is often associated with loss of human lives, and, like the Class I product failures, is viewed as a catastrophe.

The products of the Class II are typically intended for an industrial market, rather than government or individual consumers. This market is characterized by relatively high volume of production items (buildings, bridges, commercial ships, commercial aircraft, telecommunication networks), but also by fewer and more sophisticated customers than in the commercial market (see Class III below). The reliability standards/specifications for the Class II products come as industrial standards, such as Telcordia, JEDEC, ASTM, etc. These standards often include some MIL-STDs or MIL-STDs requirements as their constitutive parts. The vendor and the customer usually negotiate some form of a reliability-and-quality contract for the Class II products. This contract typically includes both the appropriate industrial specification requirements and, in addition, the requirements of a particular customer.

If the device/component/subsystem passes the required (qualification) functional, mechanical and environmental tests, it becomes a "product," and is "qualified" to be shipped to the customer and to be installed into the customer's equipment. For some types of the Class II products, a low number of field failures might be considered acceptable, and could be even specified beforehand in the contract. For the Class II products, like for the Class I products, it is the customer, rather than the manufacturer who sets the reliability requirements and standards. These, however, could be based on the agreeable and generally acceptable industry standards.

Class III. The reliability does not have to be very high. Failures are permitted, provided that their level is not too high. The demand for the product is driven more by the cost of the product, than by its reliability. The product is relatively inexpensive, produced in massive quantities, and its failure is not viewed as a catastrophe, i.e., a certain level of failures during normal operation is considered acceptable. Examples are various household items, consumer products, agricultural equipment, etc. The typical market for these products is the consumer market. An individual consumer is a very small part of the total consumer base. Consumer intended micro- and opto-electronic products is a typical example of such a market.

Field failures are allowed and are expected to occur, as long as the failure rate is within the anticipated/expected range. The reliability testing is limited, and the improvements are implemented based on the field and market feedback. No special reliability standards are often followed, and it is the customer satisfaction (on the statistical basis), which is the major criterion of the reliability and quality of the product. It is typically the manufacturer, not the consumer, who sets the reliability standards for the product. Relatively simple and innovative Class III consumer products, which have a high degree of customer appeal and are therefore in significant demand, may be able to prosper, at least for a short period of time, even if they are not very reliable.

8.7. RELIABILITY, COST AND TIME-TO-MARKET

"*Be grateful for luck, but do not depend on it*"
William Feather, American Politician

For many Class II and all the Class III products, **cost and time-to-market** are key issues in competing in the global market place. For Class II products, developing reliable products that cost less is the primary goal of the industry. It is equally important that the product is delivered to the market on time. Reliability, cost and time-to-market considerations play an important role in the design, materials selection and manufacturing decisions. A company cannot be successful, if its products are not cost effective, or do not have a worthwhile lifetime and service reliability to match the expectations of the customer. Product failures have an immediate, and often dramatic, effect on the profitability and even the very existence of a company. This is even true for the Class III products, not to mention the Class II products.

Failure to provide adequate reliability can be costly to a business through the cost of reworking or scrapping of products during manufacturing, as well as through the cost of additional inspections and testing. Warranty repairs after the product is shipped may not only be expensive to the manufacturer, but might be even more costly to the customer in terms of loss of service and/or increased maintenance cost. Profits decrease as the failure rate increases. This is due not only to the increase in the cost of replacing or repairing parts, but, more importantly, to the losses associated with the interruption in service, not to mention the "moral losses." Such "moral losses" make obvious dents in the company's reputation and, as the consequence of that affect its future sales. Too low a reliability can lead to a total loss of business. Many small and large companies that are failing today fail because of insufficient reliability of their products.

On the other hand, there is a permanent "struggle" between the recognition of the industry that high and well-predicted product reliability is a "must" and a strong business pressure that tends to compromise product's reliability in order to shorten the time-to-market and to reduce manufacturing costs. In response to the time and cost pressures of the markets and shareholders (investors), businesses frequently take a much lower approach to reliability than they would have taken otherwise. Businesses attempt to establish a minimum level of product testing or inspection that will provide a level of reliability, which they feel is adequate for the market they serve. Certainly, it is always a challenge to establish, for each particular product and a particular situation, the most reasonable balance between the level of reliability and market demands, in terms of schedule and cost. In the past, it used to be said that of quality, schedule and price, the customer could have any two. Today, none of these items could be compromised for the other two, and the best engineering and business decisions should consider the best trade-off among them.

8.8. RELIABILITY COSTS MONEY

"*Be thankful for problems. If they were less difficult, someone with less ability might have your job*"
Unknown Reliability Engineer

It is common knowledge that "reliability costs money." Conducting reliability evaluations is not cheap, and the cost of the subsequent failure mode analysis and corrective actions may be even more expensive. It is very undesirable, of course, that a business incurs excessive expenses pursuing high reliability standards with very little payback. The cost of

improving or maintaining a certain level of product reliability must always be weighed against the benefits obtained. From the cost and business point of view, there is always an adequate, though less than perfect, level of reliability appropriate for the given product or system. This is always the case for the Class II and Class III products.

Relatively simple and innovative Class III consumer products, which have a high degree of customer appeal and are of significant demand, may be able to prosper, at least for some time, even if they are not very reliable.

A business must understand the cost of reliability, both "direct" cost, i.e., the cost of its own operations, and the "indirect" cost, i.e., the cost to its customers and their willingness to make future purchases and to pay more for more reliable products. Having this in mind, each business, whether small or large, should try to optimize its overall approach to quality and reliability. He/she should also have in mind that the time to develop and time to produce products is rapidly decreasing. This circumstance places a significant pressure on reliability engineers. They are supposed to come up with a reliable product and to confirm its long-term reliability in a short period of time to make their device into a product and to make this product successful in the marketplace.

8.9. RELIABILITY SHOULD BE TAKEN CARE OF ON A PERMANENT BASIS

"*The probability of anything happening is in inverse ratio to its desirability*"
John W. Hazard, American Writer

There is a story about a young couple who had a newborn baby and asked George Bernard Shaw, who was famous of his wisdom, for an advice. "Our baby is four months old. When should we start bringing it up?" "You are four months late" was the answer. This is true also about when to start being concerned about reliability.

In order that a product is successful in the market place, the manufacturer must understand the physics of failures of his/her product(s). He/she should be able also to design and manufacture a product with the predicted and sufficiently low probability of failure. In other words, he/she should know the ways, in which the useful service life of a material, device, structure, or a system can be predicted and, if necessary, improved, without bringing the product's cost up or postponing its delivery. A reliability engineer should develop effective methods to predict failures, to measure/detect them, to develop reliable methodologies for the prediction of the probability of failure, and, on this basis, to develop methods to minimize and/or to prevent failures at all the stages of the product design, manufacturing, testing and production. The reliability evaluation and assurance cannot be delayed until the device is made (although it is often the case in many actual industries). Reliability of a product should be

- "conceived" at the early stages of its design (a reliability and optical engineers should start working together from the very beginning of the optical device engineering),
- implemented during manufacturing (quality control is certainly an important part of a manufacturing process),
- qualified and evaluated by electrical, optical, environmental and mechanical testing (both the customer requirements and the general qualification requirements are to be considered),
- thoroughly checked (screened) during production, and, if necessary,
- maintained in the field during the product's operation, especially at the early stages of the product's use.

New products present natural and particular reliability concerns, as well as significant challenges at all the stages of their design, manufacture and use. These concerns and challenges have to do with the evaluation and assurance of both the functional (electrical and optical) performance and the structural (mechanical) reliability of the product. One of the major challenges, associated with new product development and reliability, is design and implementation of the adequate accelerated qualification tests and accelerated life test (ALT) approaches, methodologies and procedures [1–8].

A key bottleneck to meet the cost and time-to-market objectives is the product qualification and quality assurance. The required level of reliability is being typically proven based on the standardized qualification tests (QTs) and specifications (acceptance criteria). It is primarily the QTs that make a photonic device into a product. But it is the ALTs that enable a reliability specialist to understand the engineering and science behind the product. It is also the ALTs that enable him/her to create, on the basis of the developed understanding, a viable and a reliable product with the predicted and sufficiently low probability of failure.

8.10. WAYS TO PREVENT AND ACCOMMODATE FAILURES

"*It is common sense to take a method and try it. If it fails, admit it frankly, and try another. But above all, try something*"
Franklin D. Roosevelt, American President

The best way to prevent failures is to understand well the physics of failure, to **anticipate the failure modes** that might occur in a particular system, and to design this system in such a way that the likelihood that these failures occur be sufficiently low. In order to achieve this one should be able to

- develop an in-depth understanding of the possible modes and mechanisms of failure in his/her design,
- understand and to distinguish between operational (functional), structural/mechanical (caused by mechanical loading) and environmental (caused by harsh environmental conditions) failures,
- assess the likelihood (the probability) that the anticipated modes and mechanisms might occur in service conditions,
- distinguish between the materials and structural reliability,
- assess the effect of the mechanical and environmental behavior of the materials and structures in his/her design on the functional performance of the product,
- understand the difference between the requirements of the qualification specifications and standards, and the actual operation conditions,
- understand well the qualification test conditions and to design the product not only that it would be able to withstand the operation conditions on the short- and long-term basis, but also to pass the qualification tests,
- control, if necessary, the product's operation and operating environment.

One should have in mind that no failure statistics, nor the most effective ways to accommodate failures, can replace good (robust) physical design. Nonetheless, some proactive measures can be very helpful and can minimize considerably the likelihood of a failure, provided that the best materials are selected and a good design is carried out.

8.11. REDUNDANCY

"*It is tough to make predictions, especially for the future*"
Yogi Berra, American Baseball Player

The most effective method to increase the reliability of a system comprised of not-very-reliable components is redundancy. The number of the redundant components does not have to be very large to build a reliable system out of relatively unreliable components. For instance, if one wants to design a system whose reliability (probability of non-failure) is as high as 99%, while the reliability (dependability) of the components that the system is built of is only 80%, one can employ just four redundant components (in parallel) to build such a system. If one, two, three or even four components fail, the system will still operate. Note that if the same components were arranged in series, the overall reliability (the probability of non-failure) of the system would be as low as about 33%. If the system has a good enough reparability, the customer will never know that there was failure in the system, because the system will always be available to him/her.

That is why it is the availability, and not the dependability, which is the appropriate reliability characteristic of the system. High **availability** (i.e., high probability that the system is available to the user when needed) can be achieved even with a not-very-high dependability (i.e., probability of non-failure) of its components, as long as the reparability level (i.e., the probability that the system's workability is restored within the given and short enough period of time) of the system is sufficiently high.

In some cases, a system can be designed in such a way that, if one or more of its parts fail, the system can still operate, with its capabilities impaired to a greater or lesser extent. A two-engine aircraft can still operate, if one of its engines fails. A passenger ship will still not sink, even if two adjacent compartments at her fore- or after-body are flooded.

8.12. MAINTENANCE AND WARRANTY

"*Only life insurance policy is able to provide a 100% warranty*"
Unknown Insurance Agent

Maintenance is another failure accommodation method. There are two extreme approaches to accommodate failures, using appropriate maintenance: (1) preventive maintenance and (2) reactive maintenance. When preventive maintenance is used, items are checked and, if necessary, replaced (even if they are still good) in accordance with some more or less well-justified schedule. When reactive maintenance is used, the faulty items are replaced when they fail. In the case of preventive maintenance, one relies on routine procedures for checking and replacing parts. In the case of reactive maintenance, one relies on good technical diagnostics and keeps a highly qualified, highly flexible and highly mobile "rescue squad" to find and to fix the occurred problem. The reactive maintenance approach is more risky, but might be much less costly.

In the Class III systems, a widely used way to mitigate the consequences of failure is to provide a warranty, i.e., a guarantee that the manufacturer will repair or replace, when necessary, the faulty item at no cost to the customer. Still, the manufacturer should bring this practice to the minimum, because nothing can replace the customer's inconvenience, time, irritation, and, hence, dissatisfaction.

8.13. TEST TYPES

"*Well done is better than well said*"
Benjamin Franklin, American Scientist and Politician

The integrated test program usually includes the following types of tests:

1. Functional (optical, electrical) testing;
2. Mechanical testing;
3. Environmental testing;
4. Safety testing.

A crucial component of these tests is the adequate definition of failure criteria. These can be established, based on the customer requirements, qualification standards, state-of-the-art in the given area of engineering, etc. The peculiarities of a particular test program depend on the resources available, reliability requirements, product application, qualification of the personnel, allocation of facilities and equipment, priorities, etc.

8.14. ACCELERATED TESTS

"*The golden rule of an experiment: the duration of the experiment should not exceed the lifetime of the experimentalist*"
Unknown Physicist

Shortening of product design and product development time does not allow for time-consuming reliability evaluations. To get maximum information and maximum reliability-and-quality in minimum time and at minimum cost is the major goal of a manufacturer. One certainly wishes to have guidelines/methodologies that would enable him/her to quickly and economically evaluate the reliability of a product, and to afford an opportunity to fix reliability problems long before they lead to major losses. It is impractical and uneconomical to wait for failures, when the mean-time-to-failure for a typical today's micro- or opto-electronic device (equipment) is on the order of hundreds of thousands of hours.

Accelerated tests use elevated stress level and/or higher stress-cycle frequency to precipitate failures over a much shorter time frame. As has been mentioned above, the "stress" does not necessarily have to be a mechanical or a thermo-mechanical one: it can be electrical current or voltage, high (or low) temperature, high humidity, high frequency, high pressure or vacuum, cycling rate, or any other factor responsible for the reliability of the device or the equipment. In order to accelerate the material's (device's) degradation and/or failure, one has to deliberately "distort" one or more parameters (temperature, humidity, load, current, voltage, etc.) affecting the device's functional and/or mechanical performance.

Accelerated tests enable one to gain greater control over the reliability of a product. They have become a powerful means in improving reliability. In accelerated tests one applies a high level of stress over a short period of time to a device/product presuming/assuming that there will be no "shift" in the failure modes and mechanisms. This is true regardless of whether failures will actually occur during the tests (Accelerated Life Tests, which are aimed at "testing to fail") or not (Qualification Tests, which are aimed at "testing to pass"). The accelerated tests must be specifically designed for the product under test. The experimental design should consider the anticipated failure modes and mechanisms, typical use conditions, and the required or available test resources, approaches and techniques.

8.15. ACCELERATED TEST LEVELS

"*If you do not raise your eyes, you will think that you are at the highest point*"
Antonio Porchia, Italian Poet

Accelerated tests can be performed at the part level, at the component level, at the module level, at the equipment level and even at the system level. In each particular case, the decision should be made on how to break down the equipment of interest, so that the number of failure modes of the object under testing would not be very large. For this reason, accelerated testing is usually conducted at the part (assembly) or at the component (device) level. If the reliability characteristics of all the components are established, then the reliability characteristics ("indices") of the equipment or the system can be evaluated theoretically, using methods of probabilistic (statistical) analyses. In this connection it should be pointed out that different reliability criteria are (and should be) used depending on whether it is an assembly, a component, a subsystem, a piece of equipment or a large system.

While the probability of failure (dependability) might be the right criterion for a non-reparable component, a piece of equipment should be characterized by its availability, i.e., the probability that this piece of equipment will be available to the user, when it is needed. As to a large and a complex system (say, a switching system or a highly complex communication/transmission system, in which its "end-to-end reliability" is important, including the "reliability" of software), it is the "operational availability" that is of importance. This can be defined as the probability that the system is available "today" and will remain available to the user for the given period of time "tomorrow." What this, actually, means is that the system performs as expected every time the customer accesses it or needs it, whether it is 300 million voice attempts a day or 675 trillion bytes of data a network carries each day. To achieve that one does not have to necessarily keep the dependability of a particular component or even of a subsystem at a very high level. He/she can run a highly available system by achieving high reparability, reasonable redundancy, high-level of trouble shooting, etc.

Because of that, there is a rather wide spectrum of reliability requirements, ranging from very high requirements for large and complex systems, in which a failure is considered a catastrophe, down to simple consumer products, for which the consequences of failure are not as catastrophic as they are for large systems. The reliability (availability) of the contemporary communication networks is as high as 0.999. For consumer products, however, it is the cost and time-to-market that are the major driving forces, and their reliability (typically, dependability) should only be adequate for customer acceptance and reasonable satisfaction. No wonder that in reliability communities one can find a variety of opinions, attitudes and approaches to, and actual practices in, reliability assurance. It depends on the driving market forces and a particular business, whether it is component/device making, equipment manufacturing, or service provision.

8.16. QUALIFICATION STANDARDS

"*By asking impossible obtain the best possible*"
Italian Saying

The today's qualification standards and specifications (such as, say, Telcordia requirements for photonics equipment) enable one to "reduce to a common denominator" different products, as well as similar products, but produced by different manufacturers. These standards reflect, to a great extent, the state-of-the-art in a particular field of engineering, as well as more or less typical requirements for the performance of a product

intended for a particular application. Industry cannot do without accelerated qualification tests and qualification standards.

However, qualification standards and requirements are only good for what they are intended—to confirm that the given product (provided that it passed the tests) is indeed "qualified" to serve in a particular capacity. In some cases, especially for new products and new technologies, when no experience has been yet accumulated, the general qualification standards, based on the previous generations of the device or on other, "similar," devices and components, might be too stringent. An unreasonable ("torture"/"sledgehammer") qualification test that does not reflect the actual field conditions might result in a rejection of a good product, i.e., of a product that would be able to perform successfully in the field for many years ("supplier's/vendor's risk"). In other cases, the qualification specifications might not be stringent enough for a particular application or particular use conditions, and a product with a not high enough reliability level might be shipped to the customer ("consumer's/customer's risk").

If a product passed the standardized qualification tests, it is not always clear why this product was good, and if the product failed, it is equally unclear what could be done to improve its reliability. Since qualification tests are not supposed to be destructive, they are unable to provide the most important ultimate information about the reliability of the product—the information about the probability of its failure after the given time in service under the given conditions of operation. If a product passed the qualification tests, it does not mean that there will be no failures in the field, and it is unclear how likely or unlikely these failures might be, nor what could be done to improve the product's reliability.

8.17. ACCELERATED LIFE TESTS (ALTS)

"*In a long run we are all dead*"
John Maynard Keynes, British Economist

The body of knowledge in the accelerated life tests (ALTs) has come a quite long way in a rather short time. ALTs are aimed at the revealing and understanding the physics of the expected or occurred failures. Unlike QTs, ALTs are able to detect the possible failure modes and mechanisms. Another objective of the ALT's is to accumulate sufficiently representative reliability/failure statistics. Thus, ALT's deal with the two major areas of Reliability Engineering—physics and statistics of failure. ALT's should be planned, designed and conducted depending on the projected lifetime of the product, the expected operational and non-operational loading conditions and environment, the frequency and duration of such loading and environmental conditions, etc.

Adequately planned, carefully conducted, and properly interpreted ALTs provide a consistent basis for the prediction of the probability of failure after the given time in service. This information can be extremely helpful in understanding of what and how should be changed in order to design a viable and reliable product. Indeed, any structural, materials and/or technological improvement can be "translated," using the ALTs data, into a reduced probability of failure for the given duration of operation under the given service (environmental) conditions. This is, in effect, the substance of a probabilistic approach to physical (structural) and functional (electrical or optical) design of a component or a device [11,12].

Well-designed and thoroughly implemented ALTs can dramatically facilitate the solutions to many business-related problems, associated with the cost effectiveness and time-to-market. Therefore ALTs, along with the (accelerated) product development/verification

tests (PDTs) and qualification tests (QTs), play an important role in understanding and predicting the short- and long-term reliability of microelectronic and photonic equipment and devices.

In the majority of cases, various ALTs should be conducted in addition to, and, preferably, long before (or, at least, concurrently with) the qualification tests. There might be also situations, when accelerated testing can be used as an effective substitution for the qualification tests and standards, especially for very new products, when "reliable" (widely acceptable) qualification standards do not yet exist. This might result in a better understanding of the modes and mechanisms of failure, in the reduced cost of the product and in a shorter time to market, without compromising the product's reliability.

Unfortunately, quite often different manufacturers have to run the same ALTs and quite often learn reliability lessons from their own mistakes. This is because ALTs methodologies, studies, and, especially, test data are generally considered highly proprietary information, which is seldom published.

8.18. ACCELERATED TEST CONDITIONS

"*If a man will begin with certainties, he will end with doubts*;
but if he will be content to begin with doubts, he shall end in certainties"
Francis Bacon, French Philosopher

The accelerated test conditions are selected based on

- the expected failure modes and mechanisms,
- the most likely use conditions,
- anticipated environmental conditions,
- possible mechanical loadings, and
- qualification test conditions and requirements.

The most common accelerated test conditions (in any type of accelerated tests) are:

- High Temperature (Steady-State) Soaking/Storage/ Baking/Aging/ Dwell,
- Low Temperature Storage,
- Temperature (Thermal) Cycling,
- Power Cycling,
- Power Input and Output,
- Thermal Shock,
- Thermal Gradients,
- Fatigue (Crack Initiation and Propagation) Tests,
- Mechanical Shock,
- Drop Shock (Tests),
- Sinusoidal Vibration Tests (with the given or variable frequency),
- Random Vibration Tests,
- Creep/Stress-Relaxation Tests,
- Electrical Current Extremes,
- Voltage Extremes,
- High Humidity,
- Radiation (UV, cosmic, X-rays),
- Altitude,
- Space Vacuum,

- Industrial Pollution,
- Salt Spray,
- Fungus,
- Dirt,
- High Intensity Noise.

Some of the existing accelerated test equipment enables one to carry out also a combination of these tests.

This is done to detect and evaluate certain types of failure modes. Examples are: temperature/humidity bias (typically, 85°C/85%RH), fatigue tests at elevated temperature conditions, vibration tests at elevated temperature conditions, temperature cycling with voltage variations, etc.

If one cannot define the appropriate test condition with sufficient certainty, it is always advisable to assess this condition in an approximate fashion, probably, with a certain "margin of safety," rather than to ignore a particular test condition at all. If the customer, in the case of Class I or Class II products, does not define a particular test condition, it is the manufacturer who should do that.

8.19. ACCELERATION FACTOR

"*A theory without an experiment is dead. An experiment without a theory is blind*"
Unknown Reliability Engineer

Once relevant accelerated stress conditions ("stimuli") are selected, appropriate stress levels must be determined. These levels are product and application specific. For a PCB, for instance, operated in an environment at the temperatures between zero and 50°C, the qualification test design margins of 10°C, 20°C and 30°C above the specified limit are considered "marginally robust," "robust" (acceptable) and very robust ("excellent"), respectively [20]. Another approach [21] suggests that, in order to establish the appropriate stress levels for a particular product, the stress levels should be incrementally increased until a significant percentage (say, larger than 50%) of the sample size no longer functions.

The degree of stress acceleration is described by an acceleration factor. This factor is defined as the ratio of the lifetime (cycles) under normal use (field) conditions to the lifetime (cycles) under the accelerated conditions. The acceleration factor can be interpreted as the number of times the particular failure mechanism has been accelerated during the tests, because of making the conditions more severe than those anticipated in the actual service. The acceleration factor can be established after an appropriate predictive model is agreed upon. It is presumed that such a model holds for both field and test conditions.

The design of the ALTs should consider all the possible failure mechanisms caused by a particular stressing environment. In light emitting diodes (LEDs) and lasers, for instance, the ambient temperatures, the magnitude of the injected current and the light output power level are generally used as acceleration factors. Elevated humidity, temperature cycling and mechanical vibrations can also be used to stimulate failures.

It should be pointed out, however, that high acceleration couldn't always be applied to optical devices. For instance, the internal quantum efficiency of LEDs and lasers (i.e., the efficiency of converting the injected current into light) is very sensitive to the ambient temperature. Most lasers stop lasing at around 100°C. For this reason, extrapolation with a rate of degradation is usually used to estimate the lifetime of an active optical device.

Because no device failure can be typically observed through the time of testing, random failure rate does not occur and therefore failure rate or the probability of failure cannot be used as statistical characteristics of functional failures.

8.20. ACCELERATED STRESS CATEGORIES

"*Say not 'I have found the truth,' but rather 'I have found a truth*' "
Kahlil Gibran, Lebanese Poet and Artist

Accelerated tests can be divided, from the standpoint of their objectives, into the following three major types (categories):

- Product development/verification tests (PDTs), or design testing,
- Qualification ("screening") tests (QTs), or production testing,
- Accelerated life tests (ALTs), and highly accelerated life tests (HALTs).

All these tests use harsh environment (elevated stresses) to accelerate the precipitation of dormant defects and potential failures. The tests differ by their objectives, end points, success/failure criteria, and the subsequent action of the human analyst to detect failures (Table 8.1) [7].

The objective of the **product development/verification tests** (PDTs) is to obtain information on the product reliability during design, development, and early manufacturing stages. Many reliability problems are caused by inadequate design margins or variations in manufacturing processes or component quality. To create a reliable product, one must achieve robust (not very stringent) functional design margins and tighten the control of materials and structural variations. The PDTs are supposed to pinpoint the weaknesses and limitations of the design, materials, and the manufacturing technology or process. These tests are used also to evaluate new designs, new processes, the appropriate correction actions, or to compare different products from the standpoint of their reliability. This type of

TABLE 8.1.
Accelerated test types (categories).

Accelerated test type (category)	Product development (verification) tests (PDT)	Qualification ("screening") tests	Accelerated life tests (ALT) and highly accelerated life tests (HALT)
Objective	Technical feedback to make sure that the taken design approach is viable/acceptable	Proof of reliability; demonstration that the product is qualified to serve in the given capacity	Understand the modes and mechanisms of failure and to accumulate failure statistics
End Point	Time, type, level, or number of failures	Predetermined time, or the number of cycles, or the excessive (unexpected) number of failures	Predetermined number (or percent) of failures
Follow-up Activity	Failure analysis; design decision	Pass/fail decision	Failure analysis and statistical analysis of the test results
The Perfect/Ideal Test	Specific definitions	No failure in a long time	Numerous failures in a short time

testing is often limited by time (when almost no failure occurs), and has to be followed by an analysis of the observed failures, or by another in-depth ("independent") investigation. PDTs are, as a rule, destructive. Shear-off tests are a typical example of PDTs aimed at the selection and evaluation of the adequate bonding material.

The objective of the **qualification tests** (QTs) is to prove that the reliability of the product-under-test is above a specified level. This level is usually measured by the percentage of failures per lot and/or by the number of failures per unit time (failure rate). Testing is time limited. The analyst of the test results usually hopes to get as few failures as possible. The pass/fail decision is based on a go/no-go criterion. The typical requirements are no more than a few percent failing parts out of the total lot (population). Although the QTs are unable (and are not supposed) to evaluate the failure rate, their results can be, nonetheless, sometime used to suggest that the actual failure rate is at least not higher than a certain value. This can be done, in a very tentative way, on the basis of the observed (or anticipated) percent defective in the lot. Qualification tests, in the best case scenario, are nondestructive, but some level of failures is acceptable.

8.21. ACCELERATED LIFE TESTS (ALTS) AND HIGHLY ACCELERATED LIFE TESTS (HALTS)

"*If you come to a fork, take it*"
Yogi Berra, American Baseball Player

The objective of the **accelerated life tests** (ALTs and HALTs) is to reveal the physics of failure, i.e., to establish/reveal the modes and mechanisms of failure, to identify parametric degradation of the materials and structures under test, and the longer-term failure mechanisms. In addition, ALTs are supposed to collect (accumulate) sufficiently representative statistical information about the product-under-test through its failures [13–16]. Qualification tests (QTs) give no indication on the probability of failure. ALTs do.

The ALT and HALT analyst needs to generate as many failures as feasible and as fast as possible. The ALTs are terminated, when the modes and mechanisms of failure are established, and enough failure statistics is collected. The typical acceptable failure ratio is 50%. ALT's and HALT's are destructive tests.

The difference between the accelerated life tests (ALTs) and highly accelerated life tests (HALT's) is that the HALTs are carried out to obtain, as soon as possible, the preliminary information about the reliability of the products, and the principal physics of their failures. The goal of the HALT's is to determine the "weakest links," "bottlenecks" of the design, and to obtain the preliminary information about the major modes and mechanisms of failure. The HALT's are conducted with a smaller number of samples and at higher acceleration factors than the ALTs, so that the duration of tests could be made short enough.

Typically, the duration of HALTs does not exceed two or three months. ALTs, on the other hand, enable one to obtain more realistic information about the product's failure. ALTs are conducted with a larger number of samples and for a longer time than HALT's. It is on the basis of the ALTs (not HALTs) that a reliability engineer can accumulate sufficiently representative failure statistics and to establish the probability of failure in the filed conditions after the given time of operation.

New products should be evaluated, based on both categories of the accelerated life tests, since these products are leading edge technology, are often rather complex and relatively high-cost, and might have long-term failure modes that can be time-consuming to

isolate and resolve. Both types of accelerated life testing are usually required to ensure customer satisfaction. The cost of such tests, compared with the likelihood of encountering a field problem months after a large deployment of the product to the field, is usually considered worth the investment. An early warning of a potential failure can easily pay off the test cost, especially when the product ramp rates are steep. The ALTs give both the supplier and the customer an indication of the actual reliability of the product and its components.

One should always have in mind that a field failure might occur even if the product passed all the QTs. QTs are not supposed to reflect the actual use conditions. ALTs are.

8.22. FAILURE MECHANISMS AND ACCELERATED STRESSES

"*All life is an experiment. The more experiments you make the better*"
Ralph Waldo Emerson, American Poet and Philosopher

Typically, there is a predominant stress leading to a particular failure mechanism. Some of the failure mechanisms and the corresponding predominant accelerated stresses are summarized in Table 8.2 [7].

8.23. ALTS: PITFALLS AND CHALLENGES

"*In every big cause one should always leave something to a chance*"
Napoleon, French Emperor

Sometimes, accelerated test conditions may hasten failure mechanisms that are different from those that could be actually observed in service conditions. Examples are: change in materials properties at high or low temperatures, time-dependent strain due to diffusion, creep at elevated temperatures, occurrence and movement of dislocations caused by an elevated stress, etc. Because of the existence of such a "pitfall," it is always necessary to correctly identify the expected failure modes and mechanisms, and to establish the

TABLE 8.2.
Failure mechanisms and the corresponding accelerated stresses.

Failure mechanisms	Accelerated stresses and parameters
Corrosion (electrochemical, gaseous, galvanic, diffusion-controlled, in the presence of polymer coatings, nonelectrolyte, etc.)	Corrosive atmosphere, temperature, relative humidity
Creep and stress relaxation (static, cyclic)	Mechanical stress, temperature
Delamination	Temperature cycling, relative humidity, frequency
Dendrite growth and/or intermetallics formation	Voltage, humidity
Diffusion	Temperature, concentration gradient
Electromigration and thermomigration (forced diffusion due to electric potential or thermal gradients)	Current density, temperature
Fatigue (high- or low-cycle) crack initiation & propagation	Mechanical stress range, cyclic temperature range, frequency
Interdiffusion	Temperature
Radiation damage (radiation induced embrittlement, charge trapping in oxides, etc.)	Intensity of radiation, total dose of radiation
Stress corrosion cracking	Mechanical stress, temperature, relative humidity
Contacts' wear	Contact force, frequency, relative sliding velocity

appropriate stress/temperature limits, in order to prevent the distortion of ("shift" in) the original (actual) dominant failure mechanism. If, for one reason or another, such a situation cannot be avoided, it should be well understood and adequately interpreted, so that the ALTs do not lead to an erroneous conclusion. In this connection, it should be pointed out that different failure mechanisms are characterized by different activation energies (if, say, Boltzmann-Arrhenius type of equation is used to extrapolate the test data for the use conditions). A simple superposition of the effects of two mechanisms can result in erroneous reliability projections and, as a rule, should not be used.

Another pitfall has to do with the situation, when the accelerated test conditions lead to a bimodal distribution of failures, i.e., to a situation when a dual mechanism of failure takes place. Particularly, infant mortality ("early") failures might occur concurrently with the anticipated ("operational") failures. It is important to make sure that the "early" and "operational" failures are well separated in the tests. Infant mortality failures are usually due to the shortcomings of the manufacturing process and, although should be viewed as "atypical," are, in effect, inevitable. The most common infant mortality failures in microelectronic and photonic structures are: weak boundaries and delaminations, inclusions and voids, imperfections in geometry and materials (leading to elevated stress concentration), uneven coatings and nonuniform adhesive layers, current leakage, etc.

8.24. BURN-INS

"*There is no such thing as failed experiment. There are only experiments with unpredictable outcomes*"
Unknown Reliability Engineer

Burn-in ("screening") tests are widely implemented to detect and eliminate infant mortality failures. The rationale behind the burn-in tests is based on a concept that mass production of devices generates two categories of products that pass qualification specifications: robust ("strong") components that are not expected to fail in the field and relatively unreliable ("week") components ("freaks") that most likely will fail in the field in some future time, if shipped to the customer.

Burn-ins are supposed to stimulate failures in defective devices by accelerating the stresses that will cause defective items to fail without damaging good items. Burn-ins are needed to stabilize the performance of the device in use. Burn-ins can be based on high temperatures, thermal cycling, voltage, current density, high humidity, etc. In burn-ins the stress is highly enhanced to generate failure of the "weakest link"/weakest-element in a very short time. These tests strongly accelerate the failure mechanisms' kinetics and cause defective parts to fail, thereby, supposedly, excluding the risk of their failure in the field. In other words, burn-in tests are intended to eliminate the infant mortality portion of the bathtub curve.

For products that will be shipped out to the customer, burn-ins are nondestructive tests. Burn-ins are mandatory on most high-reliability procurement contracts, such as defense, space, and telecommunication systems. In the today's practice burn-ins are often used for consumer products as well. For military applications the burn-ins can last as long as a week (168 hours). For commercial applications burn-ins typically do not last longer than two days (48 hours).

Optimum burn-in conditions can be established by assessment of the main expected failure modes and their activation energies, and from the analysis of the failure statistics during burn-in. Burn-ins are performed by either manufacturer or by an independent test

house. Burn-in is a costly process, and therefore its application must be thoroughly monitored. Special investigations are usually required, if one wishes to ensure that cost-effective burn-in of smaller quantities is acceptable. Another cost-effective simplification can be achieved, if burn-in is applied to the complete equipment (assembly or subassembly), rather than to an individual component, unless it is a large system made up of several separately testable assemblies.

Although there is always a possibility that some defects might escape the burn-in tests, it is more likely that burn-in will introduce some damage to the "healthy" structure, i.e., will "consume" a certain portion of the useful service life of the product. This is because burn-ins not only "fight" the infant mortality, but accelerate the very degradation process that takes place in the actual operation conditions, unless the defectives have a much shorter lifetime than the "healthy" product and have a more narrow (more "deterministic," more "delta-like") probability-of-failure distribution density.

Some burn-in tests (high electric fields for dielectric breakdown screening, mechanical stresses below the fatigue limit, and some others) are harmless to the materials and structures under test, and do not lead to an appreciable "consumption" of the useful lifetime (field life loss). Others, although do not trigger any new failure mechanisms, might consume some small portions of the device lifetime. Therefore, when planning, conducting and evaluating the results of the burn-in tests, one should make sure that the stress applied by the burn-in tests is high enough to weed out infant mortality failures, but, at the same time, is low enough not to consume a significant portion of the product's lifetime, nor to introduce a permanent damage.

A natural concern, associated with the burn-in tests, is that there is always a jeopardy that burn-in might trigger some failure mechanisms that would not be possible in the actual use conditions and/or might affect the components that should not be viewed as defective ones.

8.25. WEAR-OUT FAILURES

"*The problem is not that old age comes.*
The problem is that young age passes"
Common Wisdom

The bathtub curve of a device that underwent burn-in is supposed to consist of a steady state and wear-out portions only. In lasers, the "steady-state" portion is, in effect, not a horizontal, but a slowly rising curve. Standard production burn-in tests should be combined for laser devices with the long-term life testing. Burn-in for laser devices is typically conducted in dark forced-air ovens at different combinations of constant temperature and current. Periodically parts are removed from the oven and dc tested (at room temperature). Failure can be defined, for instance, as a 2 dB reduction in the output power at the given current.

There is another pitfall associated with the wear-out failures. For a well-designed and adequately manufactured product, the were-out failures should occur at the late stages of operation and testing. If one observes that it is not the case (the steady-state portion of the "bathtub" curve is not long enough or does not exist at all), one should revisit the design and to choose different materials and/or different design solutions, and/or a different (more consistent) manufacturing process, etc. In photonics products the wear-out part of the bathtub curve can occupy a significant portion of the product's lifetime, and should be carefully analyzed.

8.26. NON-DESTRUCTIVE EVALUATIONS (NDE'S)

"*It is always better to be approximately right than precisely wrong*"
Unknown Reliability Engineer

Many nondestructive means of failure detection and evaluation (NDE) can be very useful: ultrasonic methods, X-raying, Moiré interferometry, IR defectometry, etc. In connection with the use of nondestructive methods, it is noteworthy that some observed defects should not be necessarily viewed as reliability concerns, but should be rather considered as "quality defects." One should have in mind that it is the size and location of a defect, and the loading (stress) conditions that should be considered when deciding if this defect should be tolerated or might cause a reliability problem. For instance, even a large void in the middle of a solder joint might be acceptable and should be viewed as a quality, rather than reliability, defect. However, even a small void (especially a number of "organized," "lined-up," small voids) at the interface (especially at the solder bump corner) can lead to the fatigue (and then brittle) crack initiation and propagation, and should be avoided.

Another aspect, associated with nondestructive evaluations, concerns the resolution (measurement accuracy) of the available/affordable equipment. If it is likely that the level of defectives that might escape inspection exceeds the tolerance limits for the given measuring device, then it is incumbent that burn-in is implemented. For instance, it is well known that the "accuracy" of the operation of a laser diode might very well exceed the accuracy of the equipment, which is used to measure its performance.

8.27. PREDICTIVE MODELING

"*Any equation longer than three inches is most likely wrong*"
Unknown Physicist

"*God created the world such that what is simple is true and what is complicated is false*"
Gregory Skovoroda, Ukrainian Philosopher

ALTs cannot do without simple and meaningful predictive models. It is on the basis of such models that a reliability engineer decides which parameter should be accelerated, how to process the experimental data and, most importantly, how to bridge the gap between what one "sees" as a result of the accelerated testing and what he/she will possibly "get" in the actual operation conditions [9,10].

For a manufacturer, the existing qualification standards for the Class I and Class II are "the bible," and, when implementing these standards, he/she can make his/her product qualified without even knowing the actual modes and mechanisms of failure. However, for an engineer who is developing qualification standards, predictive modeling is as important as the actual experimental data are. These models are supposed to provide meaningful relationships that clearly indicate "what affects what and what is responsible for what" and that are able to quantitatively describe these effects. These relationships may or may not include time. If the constitutive relationships do not include time, it usually means that they describe the "steady state" conditions that occur during the mid-portion of the product's lifetime. The relationship could be analytical or based on computer simulations, could be of deterministic or probabilistic nature, could be based on an apriori (probabilistic) analysis (prediction) or on a posteriori (statistical) processing of the obtained experimental data, etc.

Predictive modeling, both functional performance and materials reliability related, should be viewed as an important constituent part of the reliability evaluations. Computation of the expected reliability at conditions other than the actual or accelerated test envi-

ronment can provide important information about the device performance after a certain time in service or during accelerated testing at the given conditions. By considering the fundamental physics that might constrain the final design, predictive modeling can result in significant savings of time and expense.

Modeling can be very helpful, for instance, in optimizing the performance and lifetime of a device. For instance, the threshold current in an oxide-aperture VCSEL can be brought down by reducing the oxide aperture diameter. This, however, will result in a higher electrical resistance and higher thermal impedance, because the current must pass through a smaller constriction [17]. Since there are size-related trade-offs between the functional performance and structural/materials reliability, an optimal combination of the design possibilities (oxide thickness, vertical placement, aperture diameter, mirror and active region design, etc.) can possibly exist. Clearly, such a design optimization can be achieved only on the basis of predictive modeling.

As far as photonics applications are concerned, high precision in modeling is as important, as high precision in manufacturing. For instance, special effort should be taken to make the existing finite element programs accurate enough to be suitable for the evaluation of the stresses in, and the displacements of, the structural elements in a photonic device. Based on our recent experience, we suggest that analytical ("mathematical") modeling be more widely used to master the preprocessing models in finite element analyses, or, in some cases, even to carry out the entire modeling process [18]. This provides an obvious challenge for reliability and design engineers.

Another challenge, associated with predictive modeling in photonics reliability engineering, is the necessity for considering time-dependent behavior of a material or a structure. Creep and stress relaxation are crucial phenomena to be considered and, if possible, adequately modeled, when a photonic product is designed for a high long-term reliability. In lasers, significant temperature acceleration cannot be applied, since lasers stop lasing at about 100°C. Consequently, extrapolation with a degradation rate is usually employed to estimate the lifetime. The degradation rate is given by the change in the monitored characteristics as a function of aging.

It is the structure of a particular analytical model, and not the numerical values of the parameters, that makes it generic and, therefore, useful. Although in some situations a particular model might be inadequate for the given application or a new situation, it is important that it is amenable to updates and revisions, if necessary, and that it "reduces to the common denominator" the accumulated knowledge to provide continuity. A good predictive reliability model does not need to reflect all the possible situations, but rather should be simple, should clearly indicate "what affects what" in the given phenomenon or structure, and be suitable/flexible for new applications, with new environmental conditions and new technology developments.

8.28. SOME ACCELERATED LIFE TEST (ALT) MODELS

"*All the general theories stem from examination of specific problems*"
Richard Courant, German Mathematician

It is expected that an accelerated life test model is simple enough, yet meaningful, to be useful for the application in question. It does not have to be comprehensive, but has to be sufficiently generic, and should include all the major variables affecting the phenomenon (failure mode) of interest. As Einstein said, "a good model should be as simple as possible,

but not one bit simpler." In other words, a good model should contain all the most important parameters that are needed to describe and to characterize the phenomenon of interest, while parameters of the second rate of importance should not be included into the model. A good life test model should be suitable for the accumulation, on its basis, the reliability statistics and should be flexible enough to account for the role of materials, structures, loading (environmental) conditions, new designs, etc. The scope of the model depends on the type and the amount of information available. ALT models take inputs from various theoretical analyses, test data, field data, customer requirements, qualification spec requirements, state-of-the-art in the given field, consequences of failure for the given failure mode, etc.

Here are some major ALT models (constitutive equations, relationships) used in ALTs of microelectronic and photonic structures. They are all deterministic, and the majority of them apply to the steady state conditions only, i.e., do not consider time related effects. For this reason these relationships are not applicable to the infant-mortality and wear-out portions of the bathtub curve.

8.28.1. Power Law

For some failure mechanisms the analytical models that are used to predict reliability (as represented by the time-to-failure, or cycles-to-failure) have a power law structure:

$$T = C\sigma^n, \tag{8.1}$$

where σ is the stress parameter, and C and n are material parameters.

The power law is used, for instance, to describe degradation in lasers, when the injection current or the light output power are used as acceleration parameters. It is used also to describe the "static fatigue" (delayed fracture) of silica material in optical lightguides. In this case, T in the formula (8.1) is time-to-failure and the exponent n is negative: $n = -18 \to -20$.

8.28.2. Boltzmann-Arrhenius Equation

If Boltzmann-Arrhenius equation is used, the mean time-to-failure can be sought as

$$\tau = \tau_o \exp\left[\frac{U_a}{k(T - T_*)}\right], \tag{8.2}$$

where U_a, eV, is the activation energy, $k = 8.6174 \times 10^{-5}$ eV/K is Boltzmann's constant, T is the absolute temperature, T_* is the temperature sensitivity threshold (if any), and τ_o is the time constant. The equation was first obtained by the German physicist L. Boltzmann in the statistical theory of gases, and then applied by the Swedish chemist S. Arrhenius to describe the inversion of sucrose.

Boltzmann-Arrhenius equation is applicable, when the failure mechanisms are attributed to a combination of physical and chemical processes. Since the rates of many physical processes (such as, say, solid state diffusion, many semiconductor degradation mechanisms) and chemical reactions (such as, say, battery life) are temperature dependent, it is the temperature that is used as an acceleration parameter. The activation energy has been determined for many materials and failure mechanisms used in micro- and opto-electronics. For semiconductor device failure mechanisms the activation energy ranges from 0.3 to 0.6 eV; for intermetallic diffusion it is between 0.9 and 1.1 eV.

Activation energies for some typical failure mechanisms in semiconductor devices are [22]:

- for metal migration 1.8 eV
- for charge injection 1.3 eV
- for ionic contamination 1.1 eV
- for Au-Al intermetallic growth 1.0 eV
- for surface charge accumulation 1.0 eV
- for humidity-induced corrosion 0.8–1.0 eV
- for electromigration of Si in Al 0.9 eV
- for Si junction defects 0.8 eV
- for charge loss 0.6 eV
- for electromigration in Al 0.5 eV
- for metalization defects 0.5 eV

The Boltzmann-Arrhenius equation can be used to model temperature induced degradation in many electronic and photonic products, including lasers. It is presumed that the rate of degradation in lasers is due to diffusion, precipitation, oxidation and other temperature dependent phenomena, so that the degradation, D, rate can be described by the equation

$$dD/dt = A\exp(-U_0/kT). \tag{8.3}$$

Solid-state diffusion can form brittle intermetallic compounds, weaken local areas, cause high electrical impedance.

The effect of the relative humidity (RH) can be accounted for, if the relationship (8.2) is used, by multiplying the right part of this equation by the factor $1/(RH)^n$, where n is an empirical parameter. This relationship can be used, for instance, to describe the results of ALTs for planar lightwave circuit (PLC) devices.

The activation energy and the temperature sensitivity threshold should be established experimentally for a particular application. As to the time constant τ_o, it does not have to be determined, if it is the acceleration factor that is of interest.

8.28.3. Coffin-Manson Equation (Inverse Power Law)

The Coffin-Manson equation (inverse power law) is applicable when the lifetime of the material or a structure is inversely proportional to the applied stress [23,24]. In accordance with this equation, the median number-of-cycles-to-failure in the low-cycle fatigue conditions can be found as

$$N_f = C\sigma_r^{-m}, \tag{8.4}$$

where σ_r is the cyclic mechanical stress range ($\sigma_r = \Delta\sigma = \sigma_{\max} - \sigma_{\min}$) and C and m are material's constants. This formula was applied by many investigators to evaluate the life-time of solder joints in micro- and opto-electronics. W. Engelmaier suggested the following formula to predict the lifetime of solder joint interconnections

$$N_f = \frac{1}{2}\left(\frac{\varepsilon_r}{2\varepsilon_f}\right)^b, \tag{8.5}$$

where ε_r is the plastic strain, $\varepsilon_f = 0.325$ is the fatigue ductility coefficient [25],

$$b = \left[1.74 \times 10^{-2} \ln(1 + f) - 6 \times 10^{-4} T_s - 0.442\right]^{-1} \tag{8.6}$$

is the fatigue ductility exponent, f is the cyclic frequency ($1 \leq f \leq 1000$ cycles per day), and T_s is the mean cyclic temperature.

In random vibration tests, the mean-time-to-failure can be found in accordance with the Steinberg equation

$$\tau = C\sigma_r^{-m/2}, \tag{8.7}$$

where σ_r is the stress at the resonant frequency, and C and m are material's constants. The Equation (8.7) indicates that the mean-time-to-failure is proportional to the square root of the stress, induced at the resonance frequency.

Inverse power law is used also to model aging in lasers in the cases of current or power acceleration:

$$\tau = AI^{-n}, \quad \text{or} \quad \tau = AP^{-n}.$$

Inverse power law is used also to assess the lifetime of a silica material in optical fibers from the measured time-to-failure during accelerated stress.

8.28.4. Paris-Erdogan Equation

This equation establishes the relationship between the fatigue crack growth rate and the variation in the cyclic stress intensity factor:

$$\frac{da}{dN} = A(\Delta K)^{m_p}, \tag{8.8}$$

where da/dN is the crack growth rate, A and m_p are material constants,

$$K = G\sigma\sqrt{2\pi a} \tag{8.9}$$

is the stress intensity factor, a is the crack length, σ is the nominal stress, and the factor G is a function of geometry. The stress intensity factor range, ΔK, in the Equation (8.8) is

$$\Delta K = G\sigma_r\sqrt{2\pi a}, \tag{8.10}$$

where σ_r is the nominal stress range. The Equations (8.8)–(8.10) are applicable when the stress intensity factor range ΔK is larger than a certain threshold for the given material, below which no crack growth can occur, or below which the crack growth rate is very low. Generally, in most electronic and optoelectronic devices under normal use conditions, the initial cracks are very small, and so is the nominal stress range. It could be expected that in normal operating conditions, the ΔK value is smaller than, but not far below, the threshold value. In such a case, the fatigue life is dominated by crack initiation only. However, if the stress range increases, as it takes place in an accelerated test, then the stress intensity factor range ΔK my increase beyond the threshold value, and the failure mechanism might shift from crack initiation to crack propagation.

8.28.5. Bueche-Zhurkov Equation

Bueche-Zhurkov's equation contains not only the absolute temperature, but also the applied stress as an acceleration factor:

$$\tau = \tau_o \exp\left[\frac{U_a - \gamma\sigma}{kT}\right], \tag{8.11}$$

where γ is the stress sensitivity factor, which depends on the structure of the material and the degree of the accumulated damage. The experimentally found stress sensitivity factor for a non-oriented condition of a polyamide is about $\gamma = 1.3 \times 10^{-27}\text{m}^3$. For an oriented condition it can be significantly lower.

The Equation (8.11) underlies the kinetic approach to the evaluation of the strength of materials. In accordance with this approach, it is the random thermal fluctuations of particles (atoms) that are primarily responsible for the materials strength (failure), while the role of the external stress is reduced simply to lowering the activation energy. In many practical applications, it is only the governing relationships of the type (8.11) that is considered, while the numerical values of the parameters are evaluated experimentally for a particular application.

8.28.6. Eyring Equation

In the Equation (8.11) the effect of the external stress is considered indirectly, by reducing the level of the activation energy. This effect is considered directly in the Eyring equation:

$$\tau = A\sigma^{-1} \exp\left(\frac{U_a}{kT}\right). \tag{8.12}$$

Unlike in the Equation (8.11), the stress σ in the Eyring equation does not have to be necessarily a mechanical stress: it could be voltage, humidity, etc.

8.28.7. Peck and Black Equations

Peck's equation is, in effect, Eyring equation expanded and modified for modeling the time-to-failure in the temperature humidity bias conditions:

$$\tau = A(RH)^{-n} \exp\left(\frac{U_a}{kT}\right). \tag{8.13}$$

Here *RH* is the percent relative humidity. In Black's equation, the *RH* is substituted with the current density J, the A value is a constant related to the geometry of the conductor, and n is a parameter related to the current density, which accounts for the effects of current flow other than joule heating of the conductor.

8.28.8. Fatigue Damage Model (Miner's Rule)

Fatigue damage rule (the law of linear accumulation of damages) can be formulated as

$$D = \sum_{i=1}^{m} \frac{n_i}{N_i} \leq 1, \tag{8.14}$$

where D is the cumulative damage, n_i is the actual number of cycles applied at the i-th stress level, N_i is the number of cycles to failure under this stress, and m is the total number of different stress levels. The law of linear accumulation of damages is, generally speaking, applicable only for stresses, not exceeding the yield stress. This linear law is, strictly speaking, not applicable for the assessment of the low-cycle-fatigue lifetime. It is nonetheless often used to estimate the number-of-cycles-to-failure when a wide range of applied stresses, both below and above the yield point, are likely.

8.28.9. Creep Rate Equations

Assuming that the (static) creep rate is constant throughout the test, and that the phenomenon is dominated by the secondary stage, one can evaluate the strain rate due to creep as (Norton creep law)

$$\dot{\varepsilon} = A\sigma^n \exp\left(-\frac{U_a}{kT}\right), \tag{8.15}$$

where σ is the applied stress, and A and n are material's parameters. It is important that the Equation (8.15) is capable to represent (more or less) the entire creep curve. If the creep phenomenon is heavily dominated by the tertiary stage, the Equation (8.15) might not be adequate. Another widely used relationship for creep rate is Prandtl's law:

$$\dot{\varepsilon} = A[\sinh(B\sigma)]^n \exp\left(-\frac{U_a}{kT}\right).$$

The following Graham-Walles equation was suggested to represent all the creep stages:

$$\dot{\varepsilon} = A\sigma^n \exp\left(-\frac{U_a}{kT}\right)\left(a + bt^{-2/3} + ct^2\right), \tag{8.16}$$

where A, n, a, b and c are experimentally determined constants. It is noteworthy that if the stress σ is due to the thermal expansion mismatch of the dissimilar materials in the structure, it is not an independent variable, but is a function of the temperature T.

Creep tests are much easier to conduct than stress relaxation tests. On the other hand, phenomena associated with stress relaxation (time dependent stress for the given deformation) can be predicted, with sufficient accuracy, theoretically, if creep (time dependent deformation for the given stress) test data are available.

8.28.10. Weakest Link Models

The weakest link model assumes that the material (device) failure originates from the weakest point. This model is applicable, when the physics of the failure phenomenon

confirms that this is indeed the case. Failures due to crack generation and propagation, and dielectric breakdown are examples of weakest link failures.

8.28.11. Stress–Strength Models

These models are widely used in various problems of structural (physical) design [11]. In this model the interaction of the probability density functions for the strength and stress distributions is considered. In aerospace, civil, ocean and other structures, the probability density functions are steady state, i.e., do not change with time. In lasers, however, one can assume that the stress distribution function is indeed time independent, but the strength distribution function becomes broader and shifts toward the stress distribution function, when time progresses. At the initial moment of time the two functions are well separated, and the distance between their end points provides an appreciable margin of safety. At a certain moment of the lifetime, the right end of the strength distribution "touches" the left end of the stress distribution (the marginal state). When the time of operation exceeds the moment of time that corresponds to the marginal state, the two curves start to overlap, and the probability of failure is not zero anymore. It does not mean, however, that the device cannot be operated beyond the marginal point of time, provided that the probability of failure can be predicted with sufficient accuracy, and be made low enough for the required (specified) time of operation.

8.29. PROBABILITY OF FAILURE

"*If you bet on a horse, that's gambling. If you bet you can make three spades, that's entertainment. If you bet the device will survive for twenty years, that's engineering. See the difference?*"
Unknown Reliability Engineer

"*Probability is too important to be left to the mathematicians*"
Unknown Reliability Engineer

Based on the accelerated test data, one can predict the probability of failure at the end of the given time of the device operation. Different approaches can be used to evaluate such a probability (see, for instance, [11,12]). The most typical ("parametric") approach, used in engineering practice, is based on an assumption that not only the relationships of the previous section hold for both the accelerated and use conditions, but that the laws of the probability distributions for the parameter of interest do not change either. Recently, there were suggested several ("non-parametric") approaches, based on the extreme value distributions that enable one to successfully process the ALT data, even if the lifetime distribution is stress level dependent [19].

If, for instance, the Bueche-Zhurkov's Equation (8.11) is used, the probability of failure (in a long run) can be found as

$$P = \exp\left(-\frac{U_e - \gamma\sigma}{kT}\right). \tag{8.17}$$

If Engelmaier's Equation (8.5) is applied, then the formula

$$P = 1 - \exp\left[-2N_f\left(\frac{2\varepsilon_f}{\varepsilon_r}\right)^b\right] \tag{8.18}$$

can be used to evaluate the probability of failure after N_f cycles of loading.

It should be pointed out that many manufacturers are not familiar enough with the "mathematical" analyses underlying the "quantitative" part of the "probabilistic/statistical" reliability. For this reason, the processing of the experimental data is usually practiced by statisticians who are typically not very well familiar with and are not directly involved in the design and manufacturing of the product. In addition, their activity is associated only with what happened after (a posteriori assessment), and not prior to (apriori evaluations), the experiment. This is, of course, a significant shortcoming of the to-day's practices. This is also the reason why the probabilistic design, i.e., a design with the predicted probability of failure of the component or a device, is not even present in the to-day's micro- and opto-electronics industry.

8.30. CONCLUSIONS

"*Life is the art of drawing sufficient conclusions from insufficient premises*"
Samuel Butler

The following conclusions can be drawn from the above discussion:

- Accelerated life tests (ALTs) are aimed at the revealing and understanding the physics of the expected or occurred failures, and are able to detect the possible failure modes and mechanisms. Another objective of the ALTs is to accumulate sufficiently representative failure statistics.
- Adequately designed, carefully conducted, and properly interpreted ALTs provide a consistent basis for the prediction of the probability of failure of the product after the given time of service. Such tests can dramatically facilitate the solution to the cost effectiveness and time-to-market problems.
- ALTs should play an important role in the evaluation, prediction and assurance of the reliability of micro- and opto-electronics devices and systems. ALTs should be conducted in addition to (and, preferably, should start prior to) qualification tests required by the existing standards.

REFERENCES

1. G. Di Giacomo, Reliability of Electronic Packages and Semiconductor Devices, McGraw-Hill, New York, 1997.
2. M. Fukuda, Reliability and Degradation of Semiconductor Lasers and LEDs, Artech House, 1991.
3. O. Svelto and D.C. Hanna, Principles of Lasers, Plenum, 1998.
4. E. Suhir, M. Fukuda, and C.R. Kurkjian, Eds., reliability of photonic materials and structures, Materials Research Society Symposia Proceedings, Vol. 531, 1998.
5. E. Suhir, R.C. Cammarata, D.D.L. Chung, and M. Jono, Mechanical behavior of materials and structures in microelectronics, Materials Research Society Symposia Proceedings, Vol. 226, 1991.
6. A. Katz, M. Pecht, and E. Suhir, Accelerated testing in microelectronics: review, pitfalls and new developments, Proceedings of the International Symposium on Microelectronics and Packaging, IMAPS, Israel, 2000.
7. E. Suhir, Microelectronics and photonics-the future, Microelectronics Journal, 31(11-12) (2000).
8. E. Suhir, Analytical modeling in structural analysis for electronic packaging: its merits, shortcomings and interaction with experimental and numerical techniques, ASME Journal of Electronic Packaging, 111(2) (1989).
9. E. Suhir, Thermo-mechanical stress modeling in microelectronics and photonics, Electronic Cooling, 7(4) (2001).
10. E. Suhir, Applied Probability for Engineers and Scientists, McGraw-Hill, 1997.

11. E. Suhir and B. Poborets, Solder glass attachment in cerdip/cerquad packages: thermally induced stresses and mechanical reliability, Proc. of the 40th Elect. Comp. and Techn. Conf., Las Vegas, Nevada, May 1990; See also: ASME Journal of Electronic Packaging, 112(2) (1990).
12. E. Suhir, Analytical stress-strain modeling in photonics engineering: its role, attributes, challenges, and interaction with the finite-element method, Laser Focus World (May) (2002).
13. E.M. Baskin, Processing of the results of the accelerated life tests for the unspecified time-to-failure distribution function, Proceedings of the Academy of Sciences of the USSR, Technical Cybernetics, (3)(1988) (in Russian).
14. H.A. Chan and P.J. Englert, Eds., Accelerated Stress Testing Handbook, IEEE Press, 2001.
15. G.K. Hobbs, Development of stress screens, Proceedings of the Annual Reliability and Maintainability Symposium, Philadelphia, PA, January 1987.
16. R.A. Evans, Reliability engineering, ancient and modern, IEEE Transactions on Reliability, 47(3), p. 209 (1998).
17. L.W. Condra, Reliability Improvement with Design of Experiments, Marcel Dekker, Inc., 2001.
18. W. Weibull, Statistical design of fatigue experiments, ASME Journal of Applied Mechanics, (March) (1952).
19. D. Kececioglu and J. Jack, The Arrhenius, Eyring, inverse power law and combination models in accelerated life testing, Reliability Engineering, 8 (1984).
20. D.C. Peck and O.D. Trapp, Accelerated Testing Handbook, Technology Associates, Portola Valley, CA, 1987.
21. W. Nelson, Accelerated Testing, John Wiley and Sons, New York, 1990.
22. D.J. Klinger, On the notion of activation energy in reliability: Arrhenius, Eyring and thermodynamics, Proc. of the Reliability and Maintainability Symposium, 1991.
23. L.F. Coffin, Jr., A study on the effect of cyclic thermal stresses on a ductile metal, ASME Journal of Applied Mechanics, 76(5) (1954).
24. S.S. Manson, Fatigue: a complex subject—some simple approximations, Experimental Mechanics, 5(7) (1965).
25. W. Engelmaier, Fatigue life of leadless chip carrier solder joints during power cycling, IEEE CPMT Transactions, CHMT-6 (3) (1985).

9

Micro-Deformation Analysis and Reliability Estimation of Micro-Components by Means of NanoDAC Technique

Bernd Michel[a] and Jürgen Keller[b]

[a]*Fraunhofer MicroMaterials Center, Berlin, Germany*
[b]*Fraunhofer Institute for Reliability and Micro Integration, IZM, Berlin, Germany*

9.1. INTRODUCTION

The manufacturing of microscopic and nanoscopic objects requires the quantification of their properties. While the measurement of geometrical and size data is more easily accessible by Scanning Force Microscopy (SFM) and related methods, kinematic and mechanical characterization is a general problem for micro- and nanoobjects and devices. Displacements and their derivatives are two basic properties to be measured for mechanical description. Until now only a few methods exist to make accessible quantified field data for these tiny regions.

For that purpose different kinds of SFM imaging have been used [1–5]. Among the published quantitative approaches two techniques exist—Moiré [6–8] and image correlation based methods [9–11].

In contrast to classical Moiré measurements correlation type measurements base on higher pixel resolution SFM scans. They allow to measure displacements and strains with moderate spatial resolution within the SFM scan area [11,12]. Digital Image Correlation (DIC) is the technique currently used by most of authors measuring object deformations from SFM images.

Considering DIC techniques SFM images are captured subsequently for different object states. Mechanical and/or thermal loading is carried out by special loading stages developed for SFM and SEM application. Locally applied cross correlation algorithms are utilized to compute displacement fields and the corresponding first order derivatives from SFM images [16].

9.2. BASICS OF DIGITAL IMAGE CORRELATION

Correlation analysis on gray scale images can be realized as *field measuring and characterization method* making use of digital image or pattern acquisition and subsequent digital image processing. Then a complete set of very local image pattern is tracked between two or more object states represented by different digitized images. One of the most outstanding advantages of field measuring methods in connection with digital image processing is the possibility to obtain full two-dimensional field information instead of only point wise data. The method of correlation analysis on gray scale image pattern offers a lot of interesting new possibilities for applications in many areas of materials science and production technology. With the description of materials deformation due to thermal and mechanical loading elastic properties such as Poisson's ratio and the coefficient of thermal expansion (CTE) are obtained by correlation techniques [17]. Other fields of application are the evaluation of system response of complete structures and components including material interfaces. Compared to other displacement and strain measurement techniques such as laser interferometric or Moiré methods

Digital Image Correlation has several advantages:

- In many cases only relatively simple low-cost hardware is required (optical measurements) or already existing microscopic tools like SEM and SFM can be utilized without any changes.
- Once implemented in a well designed software code, the correlation analysis of gray scale images is user friendly and easy to understand in the measuring and post-processing process.
- For optical micrographs no special preparation of the objects under investigation is needed.
- According to its nature the method possesses an excellent downscaling capability. By using microscopic imaging principles, also very small objects can be investigated. Therefore, correlation analysis of gray scale images is predestined for qualitative and quantitative characterization of micromechanical and nanomaterial properties.

9.2.1. Cross Correlation Algorithms on Gray Scale Images

Digital image correlation methods on gray scale images were established by several research groups. Examples from different fields of applications can be found in various publications, e.g., in [9,16,18–23].

Modern SEM's allow to capture digital images and to apply correlation algorithms directly to them. This approach has been chosen by different research labs and is described in several publications [16,22,23]. The authors have developed and refined different tools and equipment in order to apply SEM images for deformation analysis on thermo-mechanically loaded electronics packages. The respective technique was established as *microDAC*, which means **micro D**eformation **A**nalysis by means of **C**orrelation algorithms [22].

The microDAC technique, by definition, is a method of digital image processing. Digitized micrographs of the analyzed objects in at least two or more different states (e.g., before and during mechanical or thermal loading) have to be obtained by means of an appropriate imaging technique. Generally, the utilized cross correlation algorithms can be applied to micrographs extracted from very different sources. Digitized photographs or video sequences but also images from e.g., optical microscopy, SEM, LSM or SPM are suitable

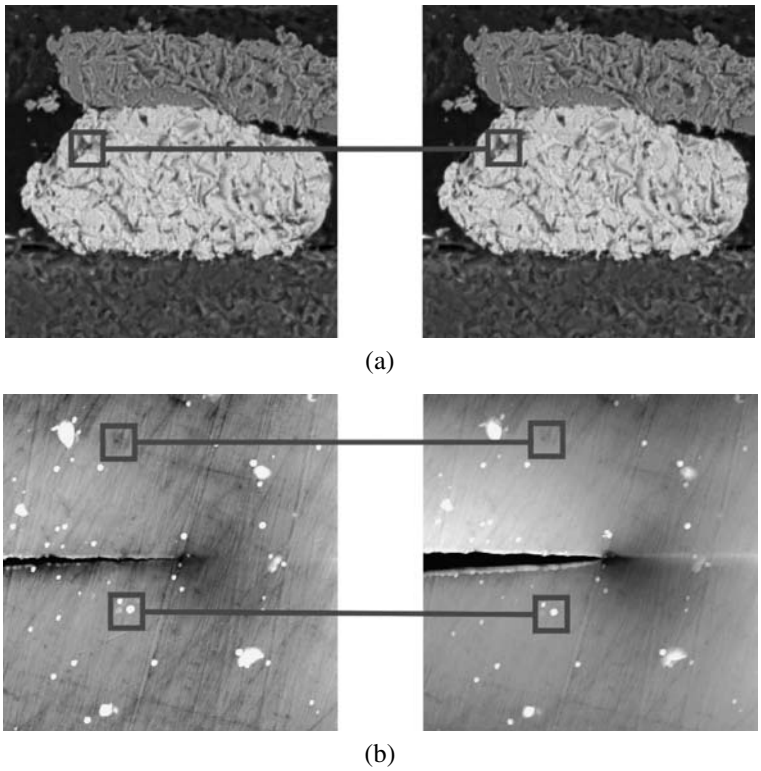


FIGURE 9.1. Appearance of local image structures (patterns) during specimen loading; (a) SEM images of flip chip gold bump; (left): at room temperature, (right): at 125°C; (b) SFM topography image of a crack in a thermoset polymer material for different crack opening displacements, scan size 15 × 15 μm.

for the application of digital image correlation. The basic idea of the underlying mathematical algorithms follows from the fact that images of different kinds commonly allow to record local and unique object patterns, within the more global object shape and structure. These pattern are maintained, if the objects are stressed by temperature or mechanically. Figure 9.1 shows two examples of images taken by SEM and SFM. Markers indicate typical local pattern of the images. In most cases, these patterns are of stable appearance, even if severe load is applied to the specimens. Just for strong plastic, viscoelastic or viscoplastic material deformation, local patterns can be recognized after loading, i.e., they can function as a local digital marker for the correlation algorithm.

The correlation approach is illustrated by Figure 9.2. Images of the object are obtained at a the reference load state 1 and at different second load state 2. Both images are compared with each other using a special cross correlation algorithm. In the image of load state 1 (reference) rectangular search structures (kernels) are defined around predefined grid nodes (Figure 9.2, left). These grid nodes represent the coordinates of the center of the kernels. The kernels themselves act as gray scale pattern from load state image 1 that have to be tracked, recognized and determined by their position in the load state image 2. In the calculation step the kernel window ($n \times n$ submatrix) is displaced inside the surrounding search window (search matrix) of the load state image 2 to find the best-match position (Figure 9.2, right).

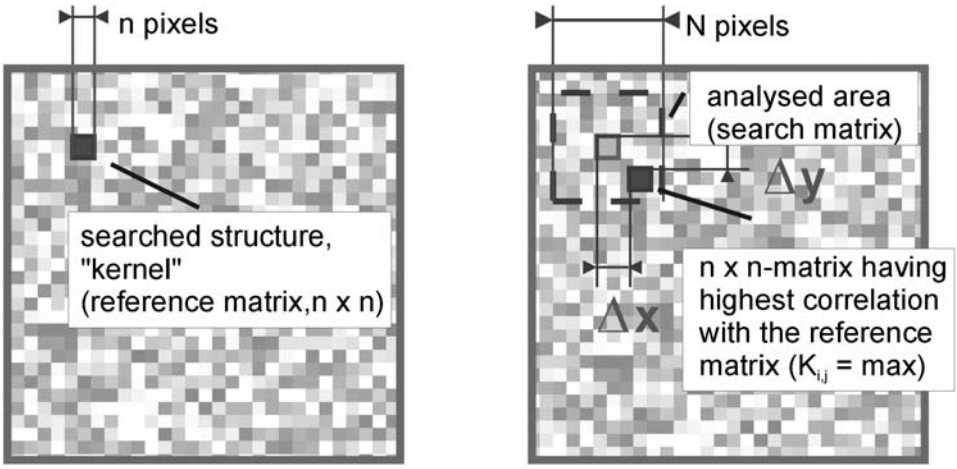


FIGURE 9.2. Displacement evaluation by cross correlation algorithm; (left) reference image at load state 1; (right) image at load state 2 used for comparison.

This position is determined by the maximum cross correlation coefficient, which can be obtained for all possible kernel displacements within the search matrix. The computed cross correlation coefficient K compares gray scale intensity pattern of load state images 1 and 2, which have the same size of the kernel. K is equal to:

$$K_{i',j'} = \frac{\sum_{i=i_0}^{i_0+n-1}\sum_{j=j_0}^{j_0+n-1}(I_1(i,j)-M_{I_1})(I_2(i+i',j+j')-M_{I_2})}{\sqrt{\sum_{i=i_0}^{i_0+n-1}\sum_{j=j_0}^{j_0+n-1}(I_1(i,j)-M_{I_1})^2\sum_{i=i_0}^{i_0+n-1}\sum_{j=j_0}^{i_0+n-1}(I_2(i+i',j+j')-M_{I_2})^2}}. \tag{9.1}$$

$I_{1,2}$ and $M_{I_{1,2}}$ are the intensity gray values of the pixel (i, j) in the load state images 1 and 2 and the average gray value over the kernel size, respectively. i' and j' indicate the kernel displacement within the search matrix of load state image 2. Assuming quadrangle kernel and search matrix sizes $K_{i',j'}$ values have to be determined for all displacements given by $-(N-n)/2 \le i', j' \le (N-n)/2$.

The described search algorithm leads to a two-dimensional discrete field of correlation coefficients defined at integer pixel coordinates (i', j'). The discrete field maximum is interpreted as the location, where the reference matrix has to be shifted from the first to the second image to find the best matching pattern. Figure 9.3 shows an example of the correlation coefficients inside a predefined search window.

With this calculated location of the best matching submatrix an integer value of the displacement vector is determined.

9.2.2. Subpixel Analysis for Enhanced Resolution

As described in the previous section the calculated displacements by the cross correlation algorithm are evaluated for integer pixel coordinates. For the calculation of the displacement field with higher accuracy the displacement evaluation has to be improved.

In the reported calculation codes applied by different research groups, the accuracy of the cross correlation technique is improved in a second calculation step using a special

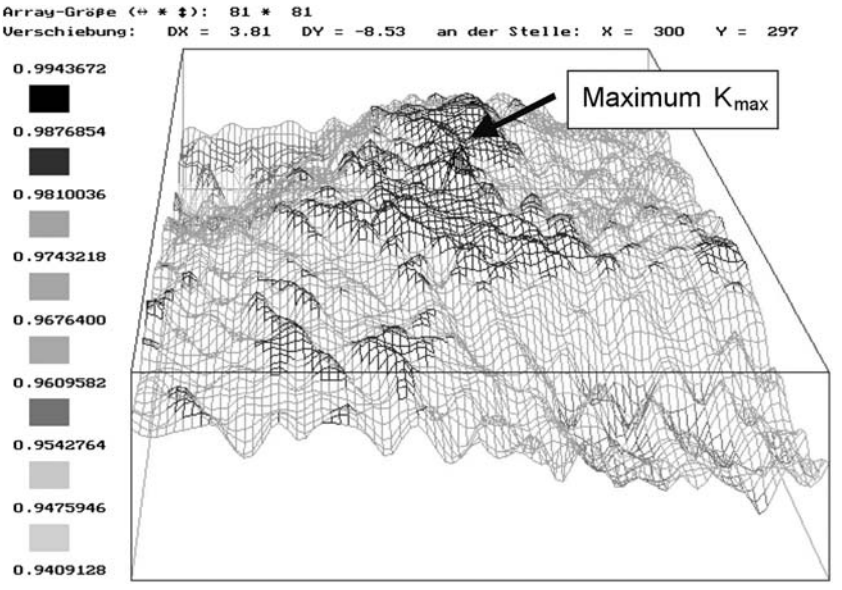


FIGURE 9.3. Discrete correlation function $K_{i',j'}$ defined at integer i', j' coordinates; the maximum of the coefficient of correlation is marked by an arrow.

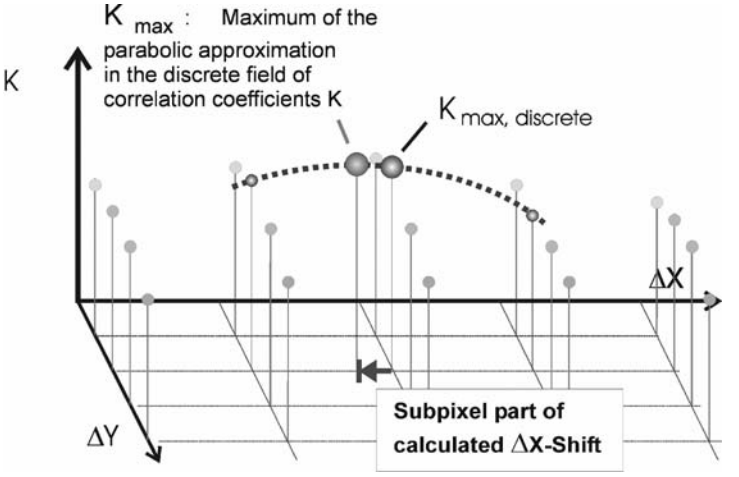


FIGURE 9.4. Principle of the parabolic subpixel algorithm.

subpixel algorithm. The presumably simplest and fastest procedure to find a value for the non-integer subpixel part of the displacement is realized in parabolic fitting. The algorithm searches for the maximum of a parabolic approximation of the discrete function of correlation coefficients in the close surrounding of the maximum coefficient $K_{max,discrete}$. The approximation process is illustrated in Figure 9.4.

The location of the maximum of the parabolic function defines the subpixel part of the displacement. This algorithm implemented quite often allows to get a subpixel accuracy of about 0.1 pixel. Even so it must be stated, that it can fail considerably and introduce

large systematic errors under some circumstances. More advanced algorithms are more accurate, allow to reach subpixel accuracies up to 0.01 . . . 0.02 pixel for common 8 bit depth digitizing, but demand sophisticated analysis and depend on the kind image sources and of data to be treated.

9.2.3. Results of Digital Image Correlation

The result of the two-dimensional cross correlation and subpixel analysis in the surroundings of a measuring point primarily gives the two components of the displacement vector. Applied to a set of measuring points (e.g., to a rectangular grid of points with user defined pitches), this method allows to extract the complete in-plane displacement field. These results can be displayed in the simplest way as a numerical list which can be post-processed using standard scientific software codes. Commonly, graphical representations such as vector plots, superimposed virtual deformation grids or color scale coded displacement plots are implemented in commercially available or in in-house software packages. Figure 9.5 shows two typical examples of graphical presentations for the results at an SFM image.

Finally, taking numerically derivatives of the obtained displacement fields $u_x(x, y)$ and $u_y(x, y)$ the in-plane strain components ε_{ab} and the local rotation angle ρ_{xy} are determined:

$$\varepsilon_{xx} = \frac{\partial u_x}{\partial x}, \quad \varepsilon_{yy} = \frac{\partial u_y}{\partial y}, \quad \varepsilon_{xy} = \frac{1}{2}\left(\frac{\partial u_x}{\partial y} + \frac{\partial u_y}{\partial x}\right), \quad \rho_{xy} = \frac{1}{2}\left(\frac{\partial u_x}{\partial y} - \frac{\partial u_y}{\partial x}\right). \tag{9.2}$$

Derivation is included in some of the available correlation software codes or can be performed subsequently with the help of graphics software packages.

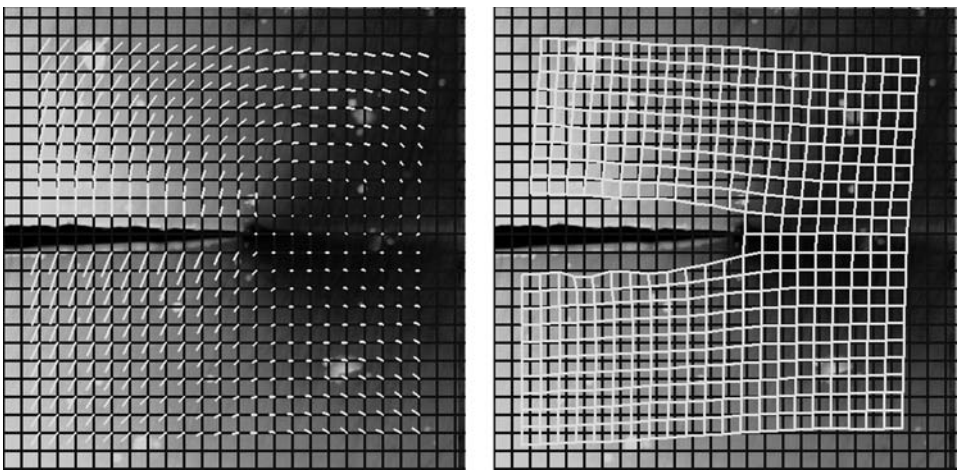


FIGURE 9.5. Digital image correlation results derived from SFM images of a crack tip, scan size [15 × 15 µm]; (left) image overlaid with user defined measurement grid and vector plot; (right) image overlaid with user defined measurement grid and deformed measurement grid, displacement vector and deformed grid presentation are enlarged with regard to the image magnification.

9.3. DISPLACEMENT AND STRAIN MEASUREMENTS ON SFM IMAGES

9.3.1. Digital Image Correlation under SPM Conditions

In comparison to DIC-based measurements treating optical or SEM micrographs, some more essential difficulties have to be overcome for SFM imaging. They correspond to the extreme magnification under SFM conditions. Because SFM image scans are taken over a time interval from one to several minutes smallest system drifts can cause significant artificial object deformations. Classifying different drift sources it can be distinguished between

- SFM scanner drifts, which are related to time dependent behavior of the piezo drives,
- relative movements between the scanner head and the sample fixture, mainly caused by temperature changes,
- drift of sample loading parameters (temperature, forces, load paths, etc.) within testing stages installed at the microscope, and
- incremental object deformations originating from viscous material behavior of test specimens, i.e., time dependent object deformations which take place even under constant loading parameters.

Substantial concerns regarding stability and reproducibility originate from drifts of the SFM scanner piezo and the thermo-mechanical loading parameters over time. As a consequence, the accurate selection of SFM equipment and loading stages is a crucial issue. Moreover, the development and implementation of methods of drift control and compensation may be a must for particular applications.

In the following drifts originating from SFM scanner drifts are considered. Figure 9.6 shows a typical result of a respective stability check carried out at a SFM equipment with activated feedback loop of the piezo scanner. For the stability measurement of Figure 9.6 a series of non-contact topography SFM scans have been picked up from an unloaded and stable mounted object. Topography data was extracted from one scan direction only, to suppress artifacts caused by scanner hysteresis. Displacement and strain values were computed for pairs of subsequent scans. The time period between the scans was negligible compared

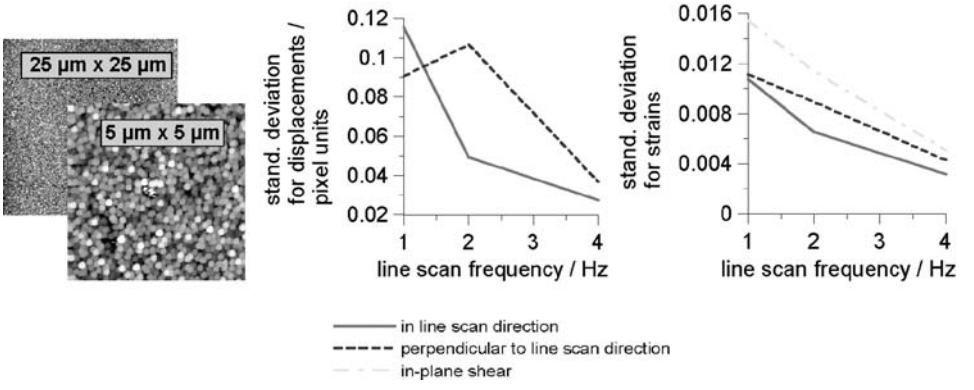


FIGURE 9.6. Estimation of measurement errors for nanoDAC by standard deviation of data determined over measurement points of a whole image, subsequent scans from an unloaded test specimen, non-contact scan mode; (left): scan from Si specimens (sample roughness approx. 10 . . . 20 nm); (right): displacement and strain standard deviation as a function of line scan frequency.

to the interval of scanning. Therefore, determined displacements and strains represent only the arbitrary measurement error of the DIC method superposed by the systematic error due to scanner drift. CMP (Chemical Mechanical Polishing) treated silicon surfaces which exhibit ideal pattern for correlation technique have been chosen for this analysis. All measurements were carried out after several hours of idle scanning in order to minimize piezo drifts.

By analyzing the data of Figure 9.6 obtained from an AutoProbe M5 device several conclusions can be made:

1. For suitable SFM choice and installation some of the commercially available equipment exhibits a scanner stability, which does not significantly reduce the possible measurement accuracy as limited by the correlation algorithms (see Section 9.2.3). For the SFM micrographs 0.1 pixel (and better) for local displacement values and 4×10^{-3} for local strain values are feasible. These results relate to topography scans obtained one immediately after the other.
2. Obviously, higher line scan frequencies, i.e., smaller scan time, improve accuracy. Values along the cantilever line scan direction are more accurate, what should be expected from the same stability considerations. The standard deviation for displacements keeps at levels as known from correlation analysis with SEM [10,24].
3. For strain values slightly higher systematic measurement errors are found than under SEM imaging. It is assumed that improvements of strain measurements are possible, because no optimizations of data refinement procedures (smoothing, grid building algorithms) have been included into the referred to analysis.

Drifts introduced by loading stages and the objects under investigation are a separate issue. Already slight drifts of loading forces and applied temperatures as well as material creep can result in large pseudo strains. The cause is a possibly accumulated displacement over the whole sample/stage size, which appears locally at the scan position as a large amount of "rigid body motion." In some cases this "rigid body drift" over the scan time leads to not negligible values of additional pseudo strains. For example, the accumulation of thermally induced displacements over an aluminum specimen of 1 cm length can give rise to already measurable pseudo strains for temperature drift rates as small as 1×10^{-3} K/min.

Figure 9.7 illustrates the impact of drift induced pseudo strains on real nanoDAC measurements. The 3D plot shows the displacement fields nearby a crack trip. The crack was opened by external forces into the direction perpendicular to the crack boundary (Mode I crack opening, see also Section 9.4.3). The comparison between the measured and the theoretical crack opening displacement fields reveals slight deviations.

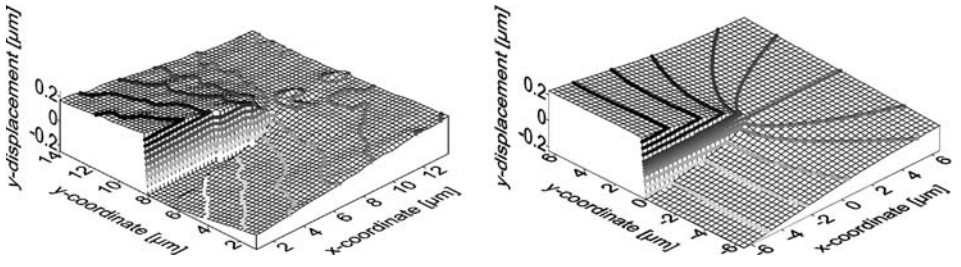


FIGURE 9.7. Crack opening displacement field near a crack tip (displacement component perpendicular to the crack boundary); (left): measured by SFM, (right): theoretical field, the incline of the overlaid displacement contourlines indicates a superposition of pseudo strains induced by drift.

In the measurement plots the displacement contourlines of equal y-displacement values are inclined on both sides of the crack boundary in different directions. This behavior is a result of a small, nearly constant pseudo strain in y-direction superposed to the real crack opening field. Nevertheless, the actual crack opening field dominates and is reproduced in the right way.

9.3.2. Technical Requirements for the Application of the Correlation Technique

There exist two main issues to be discussed if Digital Image Correlation (DIC) is supposed to be applied in SFM: suitable specimen loading must be realized within the equipment and high level scanning reproducibility must be provided.

The placement of thermal and/or mechanical loading stages in the SFM is mainly a question of compact design of loading stages, free access for SFM cantilevers to the specimen surface of interest and a scanning by cantilever-site actuation. In dependence on available equipment and loading stages it can be necessary to install additional spacers in between the cantilever head and the ground plate with $x-y$-stages for specimen adjustment.

I order to avoid scanning instabilities and different thermo-mechanical drifts (see Section 9.3.1) common tools of environmental isolation offered by equipment suppliers should be installed. This comprises active vibration compensation of equipment tables, acoustic enclosures against ambient sound and if possible also temperature stabilization. Advanced fast line scan equipment can be very helpful to meet this requirement. Because of the possible drift of object loading parameters it might be necessary to develop special tools for drift compensation. These measures can be taken actively by piezo driven displacement compensation or passively by numerical corrections of measured displacement and/or strain fields.

9.4. DEFORMATION ANALYSIS ON THERMALLY AND MECHANICALLY LOADED OBJECTS UNDER THE SFM

Most of the published work is aiming at the characterization of materials, either taking into consideration local material structures like grain size or property gradients (e.g., [7,25]) or focusing on the determination of material properties on microscopic or nanoscopic structures (e.g., [9,12,26]). The following two section present two examples, which give an impression about possible application fields of the DIC under SFM conditions.

9.4.1. Reliability Aspects of Sensors and Micro Electro-Mechanical Systems (MEMS)

Modern sensors and MEMS/NEMS devices consist of extremely fragile functional structures. Because of the desired device functionality, quite different materials in terms of material properties have to be combined with one another. Loading such structures thermally and/or mechanically means to implement severe material mismatch within submicron and nano-scale volumes. Therefore, functional or environmental loading causes local stresses and strains due to different material properties such as coefficient of thermal expansion (CTE), Young's modulus or time depended viscoelastic or creep properties. The smallest existing material imperfections or initial micro/nano-scale defects can grow under stress and strain and can finally lead to the failure of the device [10,27–30].

Thin layers used in sensor and MEMS technology undergo local stresses remote from elastic material behavior, where permanent device alterations are feared after each load cycle. Nowadays, responses of nanomaterials to applied external loads from temperature, vibrations, or chemical agents are not well understood. The same is true for actual failure mechanism and damage behavior.

The way to achieve this aim is the combination of displacement and strain measurements on the micro- and nano-scale with modeling techniques based on finite element analysis. Parameterized finite element models of MEMS are applied for faster prediction of life time and failure modes. The parameterization allows the variation of model geometries and materials in order to accelerate the MEMS design process [32].

9.4.2. Thermally Loaded Gas Sensor under SFM

Sensor applications with local temperature regulation such as the gas sensor shown in Figure 9.8 are usually thermally loaded with rapid and frequent change in temperature [29]. This thermal cycling and the temperature gradients over the structure imply thermal stresses and may cause failure of the component [33]. In the operation mode of the gas sensors thermal stresses are induced due to the activated micro-heater.

With in-situ SFM measurements on this micro system the capability of the nanoDAC approach is demonstrated measuring material deformation resulting from mismatch of material properties.

The gas sensor is designed to tolerate several hundreds of °C thermal loads. The thermal mismatch between the platinum electrodes (CTE = 9 ppm K^{-1}) and the SiO_2 substrate (CTE = 0.65 ppm K^{-1}) leads to high local stresses, if the entire device is heated up. Local displacements resulting from the thermal load have been measured by means of the nanoDAC technique.

9.4.2.1. In-plane Displacements In-situ non-contact SFM scans on top of the gas sensor membrane have been carried out at room temperature and at 100°C. The area which was observed is illustrated in Figure 9.9 as location 2. At this area an overlap of the SiO_2 membrane by the platinum electrodes should result in a thermally induced stress/strain field. The temperature was achieved by applying a defined voltage to the microheater of the gas sensor.

The determined thermally induced displacement field shows that the platinum layer with its higher CTE value reveals an inherent expansion toward the edge of the layer. In supplementary tests with heating cycles with maximum temperatures in the range of 450°C severe delaminations of the platinum layer at the edges to the SiO_2 substrate layer were observed (Figure 9.11). Details of this testing cycle are described in more detail in [29].

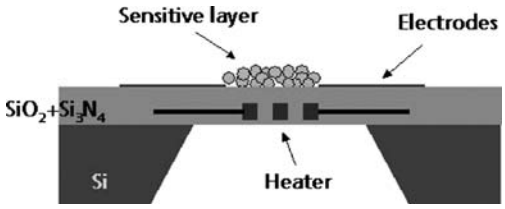
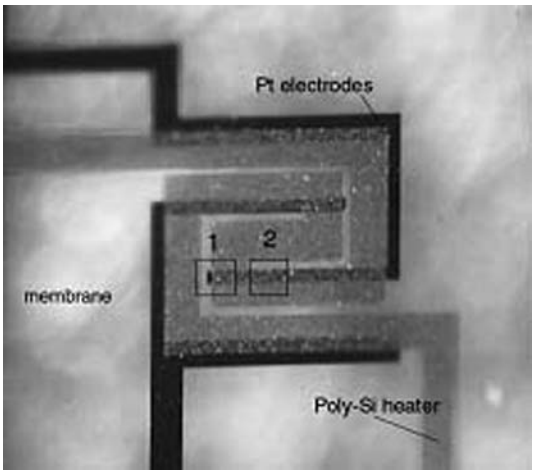
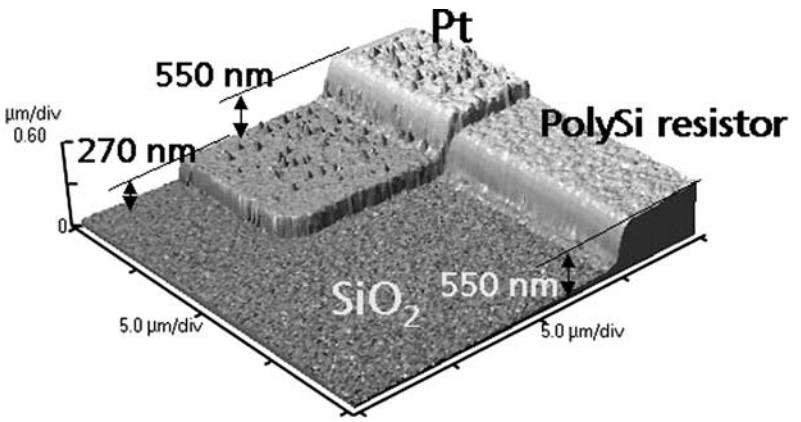


FIGURE 9.8. Layout of gas sensor.



(a)



(b)

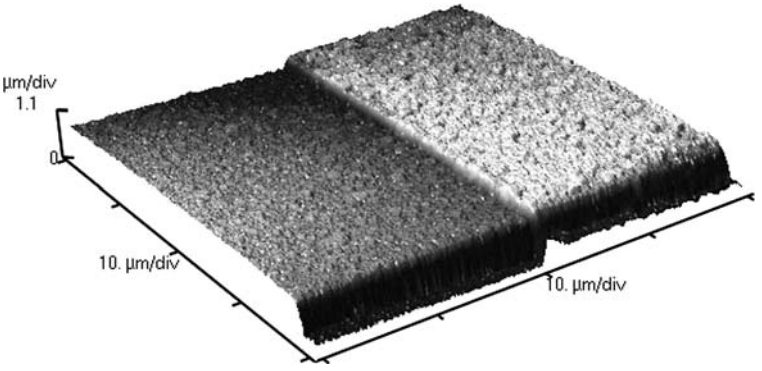
FIGURE 9.9. (a) Microscopic image of flow and gas sensor membrane, overall membrane thickness: approx. 2 µm, field of view: approx. 500 µm; (b) SFM topography scan of gas sensor depicting the Pt layer on top of the SiO_2 membrane and part of the Poly-Si heater embedded (detail 1 of Figure 9.9(a) source [29]).

9.4.2.2. Out-of-plane Displacements Besides the information on structural deformation in the $x-y$ plane the SFM measurement technique allows the determination of the out-of-plane displacement component. The height information of the SFM topography images before and after loading is analyzed for evaluation of movements or deformations in the z-direction.

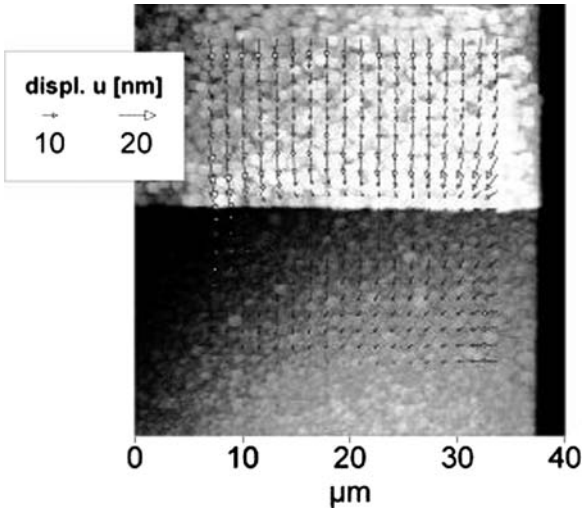
Applying this technique to in-situ measurements of thermal deformations by SFM on the top of the sensor membrane have revealed a high value of remaining deformations even after a single heat cycle (25 to 100°C). Inelastic strains remain after cooling down to room temperature (Figure 9.12).

9.4.3. Crack Detection and Evaluation by SFM

Tiny defects or cracks in microelectronics components can lead to severe crack propagation and complete failure if electronic devices are stressed. Because of intrinsic stress



(a)



(b)

FIGURE 9.10. (a) SFM topography image of platinum and SiO_2 layers; (b) vector plot of displacements u measured by nanoDAC.

sources, like e.g., thermal mismatch, the changes in environmental conditions (temperature, pressure, mechanical vibrations) can initiate crack propagation and cause fatal damage. Experimental crack detection can be a crucial issue bearing in mind original crack sizes of about some micrometers. These cracks will open only some tens of nanometers or even less under sub-critical load. Their detection, however, is possible by DIC displacement measurements.

In the following, crack detection and evaluation will be shown at a cyanate ester resin polymer material. A typical application of this thermoset is the area of microelectronic systems, where it may be used as underfiller between chip and substrate or as matrix material for printed circuit boards. The unmodified resin has a high modulus of elasticity but poor resistance to fracture [3,34].

9.4.3.1. In-situ Measurement Technique For the crack detection experiments, a simple specimen configuration is selected to demonstrate the fundamental approach. With a compact tension (CT) crack test specimen as shown in Figure 9.13 Mode I (opening) loading of

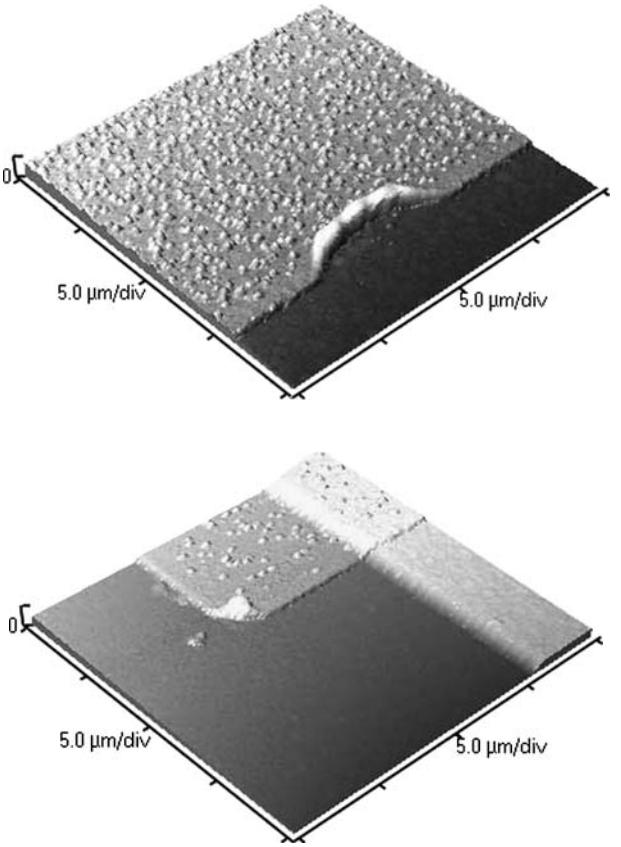


FIGURE 9.11. SFM topography scan of membrane layers after tempering at 450°C, Pt electrode destruction at edge and corners (compare to Figures 9.9 and 9.10).

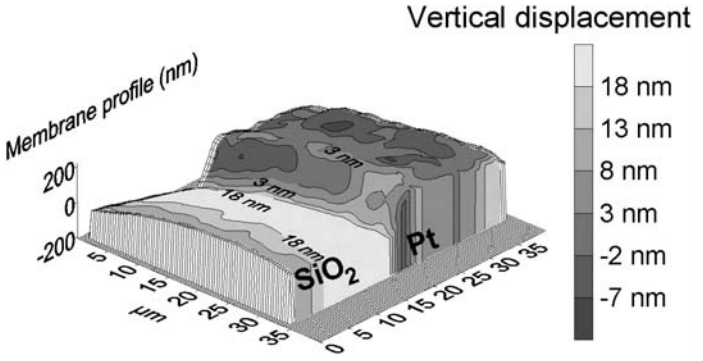


FIGURE 9.12. Residual sensor deformation after heat cycle (SFM based deformation measurement), 3D plot shows part of the membrane layer profile, the coloring (gray scale) indicates the remaining vertical deformation after a heat cycle.

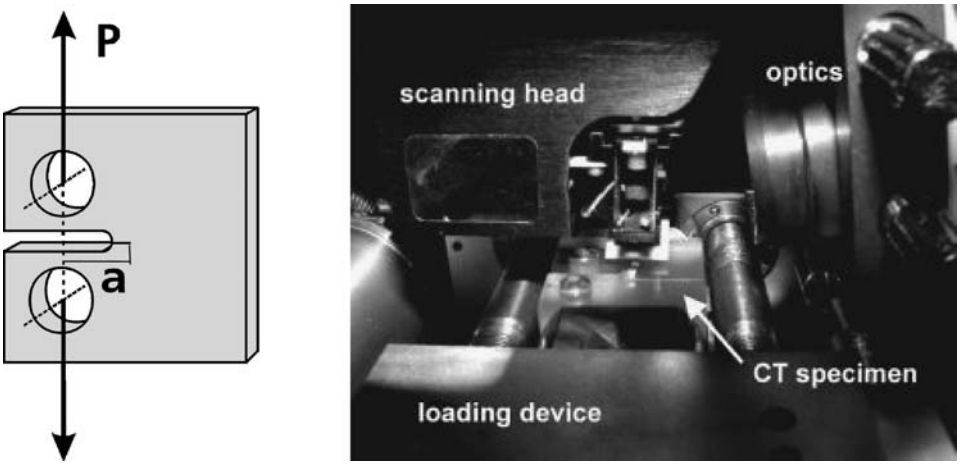


FIGURE 9.13. (left) Compact tension (CT) specimen; (right) In-situ loading of CT specimen under SPM.

the crack tip is enabled. Due to a high accuracy in the machining process of the specimen, in-plane and out-of-plane shear (Mode II and III) components are avoided to a considerable extent.

The CT specimen is loaded by a special tension/compression testing module, which can be utilized for in-situ SEM and SPM measurements. Figure 9.13 shows the CT specimen and parts of the loading device under the SPM.

SFM topography scans are taken at different locations of the crack face before and after loading. In the following presentation of measurement results the first location is approximately 50 μm away from the crack tip and the second is directly at the crack tip.

9.4.3.2. Crack Detection At the first location (crack face) of the CT specimen, the capability of the DIC methods for displacement measurements at nano-scale is demonstrated. SFM non-contact topography scans are taken at the crack face approximately 50 μm away from the crack tip. The CT specimen is loaded with a force far below the critical fracture load. The SFM images are taken before and after loading with a size of 33×33 μm and as 256×256 image arrays, i.e., the lateral resolution is approximately 130 nm/pixel.

Figure 9.14 shows the scans before and after loading with height profiles perpendicular to the crack.

A comparison of the two topography images and the height profiles of Figure 9.14 shows that the crack opening is not clearly recognizable due to the low load and the relatively coarse resolution of the scan. There are also scratches on the image which could be identified as cracks.

However, if the digital image correlation algorithm is applied to the images, the crack opening can easily be detected. Figure 9.15 shows the result obtained by nanoDAC analysis.

As illustrated in Figure 9.15 the crack opening due to loading of the CT specimen is about 200 nm. Obviously the crack cannot be identified from the SFM images themselves, because the crack opening is only in the order of 1 image pixel. This fact is illustrated by the height profiles from the topography plot (Figure 9.14), where scratches and the crack do not clearly differ from each other.

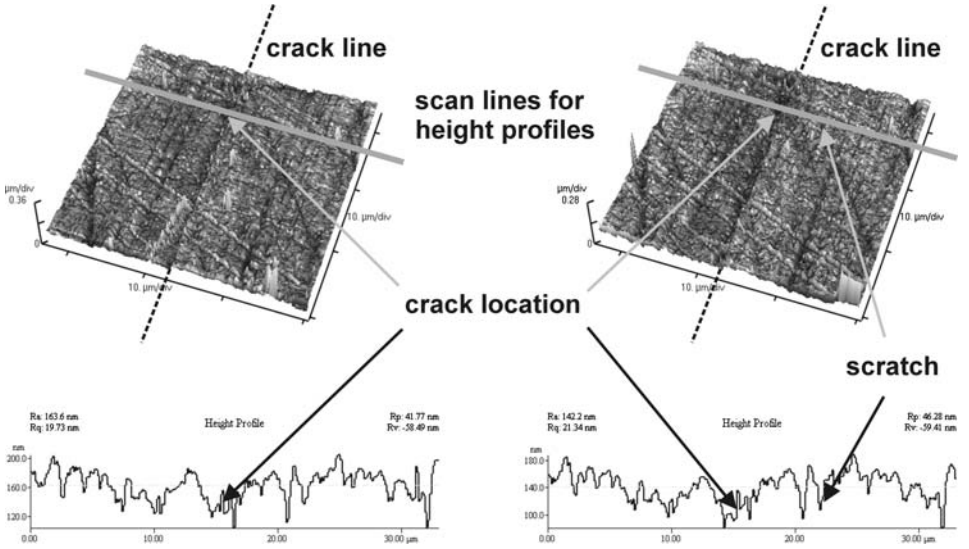


FIGURE 9.14. SFM topography images near a crack tip on cyanate ester resin specimen (33 μm × 33 μm image size); (left) SFM scan before crack opening; (right) SFM scan after crack opening.

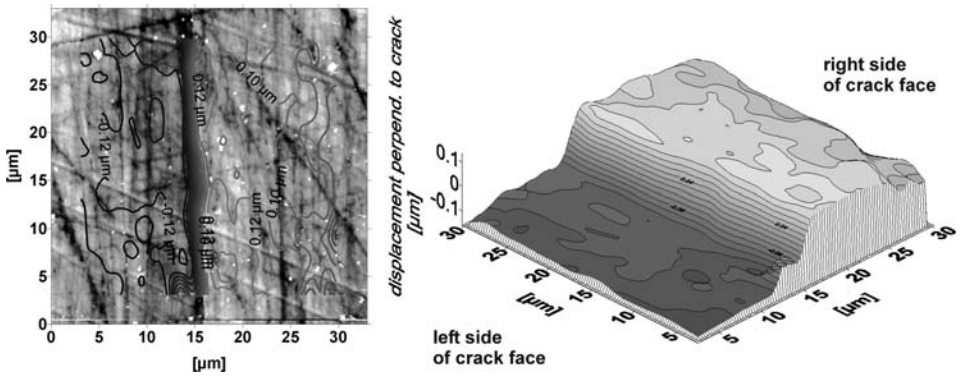
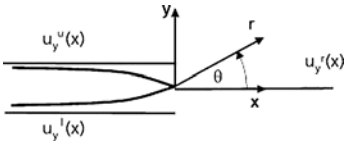


FIGURE 9.15. Displacement measurement from SFM images near a crack tip on a cyanate ester resin specimen; (left) SFM topography scan, with an overlaid contour plot showing the displacements in x-direction, u_x (component perpendicular to the crack boundaries); (right) Displacement field u_x as a 3D plot.

As a conclusion from these measurements, it can be stated that even for larger scan sizes (10 . . . 100 μm) displacements in the range of 10 nm can be measured. Cracks in the micron and submicron range can be detected and evaluated. Therefore, it is possible to apply the DIC technique for future reliability issues of MEMS and NEMS.

9.4.3.3. Crack Evaluation by COD Concept With successful detection of cracks at the micro and nano-scale the foundation is laid for a more detailed analysis of material faults and defects. The question if available fracture and damage criteria from macro or micro approaches can be transferred to a nanoscopic level is an important issue for reliability evaluation of MEMS and NEMS.

TABLE 9.1.
Crack opening displacement in LEFM for infinite bulk material and Mode I crack opening.

	$u_y^u = \frac{K_I}{2\mu}\sqrt{\frac{x}{2\pi}}(k+1),\ u_y^l = -\frac{K_I}{2\mu}\sqrt{\frac{x}{2\pi}}(k+1)$ for $x \le 0$ $u_y^u = u_y^l = 0 \quad$ for $x > 0$ (9.4)

The classical stress intensity factor K is conventionally determined by means of macroscopic fracture tests at standardized fracture specimen such as the CT specimen. This specimen type which is also used for the crack detection test described in the Section 9.4.3.1 is used for the verification of the nanoDAC technique for crack evaluation [11].

A straightforward approach for crack evaluation in the SFM is the technique of crack opening displacement (COD) determination. For the combination of the COD concept and the K concept, the following assumptions have to be made:

- Linear Elastic Fracture Mechanics (LEFM) apply within the measurement area and for the applied load,
- the specimen consists of homogeneous material.

In order to determine the Mode I stress intensity factor K_I crack opening displacements u_y^u and u_y^l have been measured along both the upper and lower crack boundaries. If determined by LEFM they must equal the values of Table 9.1.

In Equation (9.4) μ is the shear modulus and k is a function of Poisson's ratio. For the surface of the specimen where plane stress predominates k is given by $k = (3-\nu)/(1+\nu)$ [35].

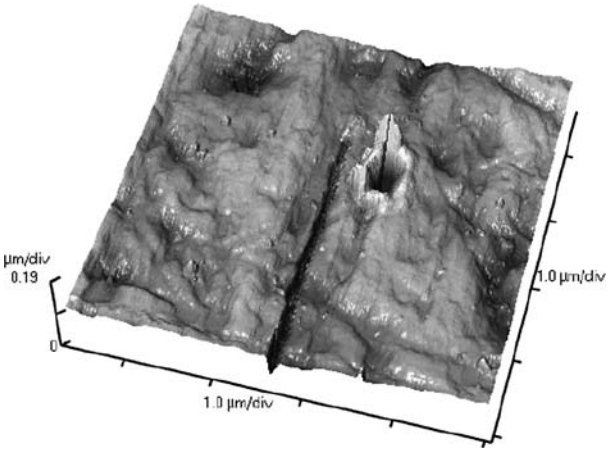
Taking the square of the difference of upper and lower displacements, we obtain a linear function of the x-coordinate or 0, in dependence, at which side of the crack tip we are:

$$\left(\frac{u_y^u - u_y^l}{2}\right)^2 = Cx, \quad x \le 0$$
$$= 0, \qquad x > 0. \tag{9.5}$$

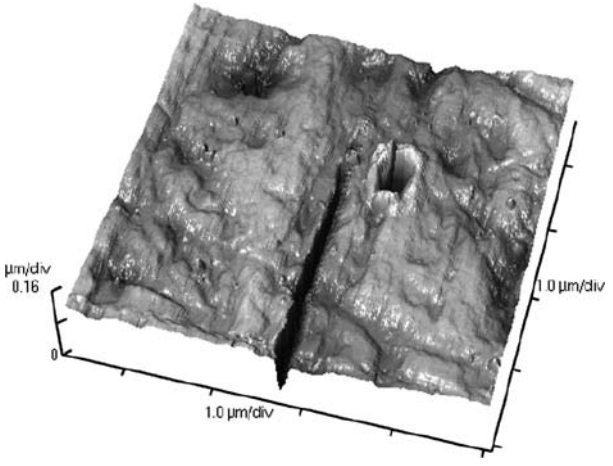
The expression of Equation (9.5) does not change if specimen rotation due to load is included into the considerations. In this case, equal rotational terms on both sides of the crack boundary are subtracted from each other. For the equation above, the crack tip is set at location $x = 0$. The crack tip location on the real specimen can be found at the interception of a linear fit of the curve Cx with the x-coordinate axis. The slope C allows to estimate the stress intensity factor K_I, which is a measure of the crack tip load and is given by:

$$K_I = \frac{E}{1+\nu}\frac{1}{k+1}\sqrt{2\pi C}. \tag{9.6}$$

In Equation (9.6) E is the Young's modulus.



(a)



(b)

FIGURE 9.16. SFM image of crack tip (size: 4.6 × 4.6 µm) for the evaluation of stress intensity factors K_I by measuring crack opening displacement around the crack tip; (a) load state 1; (b) load state 2; (the indentation near the crack tip is a indentation caused by a cantilever approach).

Continuing, an example for this procedure will be presented. The crack tip of a cyanate ester resin CT specimen is mapped by the SFM equipment at two different load states. Figure 9.16 shows the corresponding images.

To demonstrate the down-sizing capabilities of the DIC approach the scan size of these images have been chosen smaller than that of the examples described in the previous section. For a scan size of 4.6 × 4.6 µm the crack opening is already recognizable in the SFM images. Figure 9.17 illustrates the displacement results in y-direction, u_y calculated by digital image correlation along with two lines marking the upper and lower crack face. The displacement results at these lines are u_y^l and u_y^u used for the determination of the slope C (Figure 9.17).

The determined value for K_I equals 0.033 MPa m$^{1/2}$ which is a value of about $1/20$ of the critical fracture toughness K_{IC} for this type of cyanate ester thermoset.

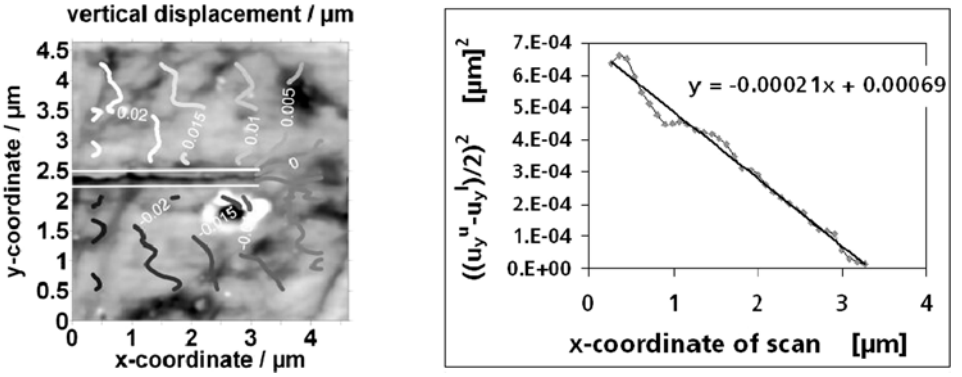


FIGURE 9.17. (left) SFM image of crack tip area (size: 4.6 × 4.6 µm) with overlaid displacement results in y-direction, u_y, lines for the upper and lower crack face are included (right) evaluation of slope C for the calculation of the stress intensity factor K_I.

9.5. CONCLUSION AND OUTLOOK

It has been shown that SFM images can be utilized to measure displacement and strain fields in very narrow regions of micro- and nanoobjects. In that way the versatile information stored in the micrographs can be converted into quantitative data. As a result characterization of materials behavior as a response to external macroscopic loading can be performed aiming at the consideration of the complex influence of the microscale and nanoscale structure. At present especially methods of Digital Image Correlation have been attracted to extract displacement and strain data from SFM images. On this occasion it must be mentioned that these tools have been developed within past very few years. One can expect, that capabilities of these tools will be extended within the next years. Stability and reproducibility issues restricting applications today may have less importance, as advanced SFM equipment and measurements approaches will come up. Also the determination of material properties by SFM based strain measurements is only at the very beginning. More advanced micro material failure analysis will be possible. Moreover, the better understanding of material behavior in the accessible tiny volumes should have an impact on the development of failure criteria.

REFERENCES

1. T. Kinoshita, Stress singularity near the crack-tip in silicon carbide: investigation by atomic force microscopy, Acta Materialia, 46(11), pp. 3963–3974 (1998).
2. K. Komai, K. Minoshima, and S. Inoue, Fracture and fatigue behavior of single crystal silicon microelements and nanoscopic AFM damage evaluation, Microsystem Technologies, 5(1), pp. 30–37 (1998).
3. C. Marieta, M. del Rio, I. Harismendy, and I. Mondragon, Effect of the cure temperature on the morphology of a cyanate ester resin modified with a thermoplastic: characterization by atomic force microscopy, European Polymer Journal, 36, pp. 1445–1454 (2000).
4. M.S. Bobji and B. Bushan, Atomic force microscopic study of the microcracking of magnetic thin films under tension, Scripta Materialia, 44, pp. 37–42 (2001).
5. C.J. Druffner and Sh. Sathish, Improving atomic force microscopy with the adaptation of ultrasonic force microscopy, Proc. of SPIE, 4703, pp. 105–113 (2002).
6. H. Xie, A. Asundi, C.G. Boay, L. Yungguang, J. Yu, Z. Zhaowei, and B.K.A. Ngoi, High resolution AFM scanning moiré method and its application to the micro-deformation in the BGA electronic package, Microelectronics Reliability, 42, pp. 1219–1227 (2002).

7. A. Asundi, X. Huimin, L. Chongxiang, C.G. Boay, and O.K. Eng, Micro-moiré methods—optical and scanning techniques, Proc. SPIE, 4416, pp. 54–57 (2001).
8. A. Asundi, H. Xie, J. Yu, and Zh. Zhaowei, Phase shifting AFM moiré method, Proc. of SPIE, 4448, pp. 102–110 (2001).
9. I. Chasiotis and W. Knauss, A new microtensile tester for the study of MEMS materials with the aid of atomic force microscopy, Experimental Mechanics, 42(1), pp. 51–57 (2002).
10. D. Vogel, J. Auersperg, and B. Michel, Characterization of electronic packaging materials and components by image correlation methods, Advanced Photonic Sensors and Applications II, Nov. 27–30, 2001, Singapore, Proc. of SPIE, 4596, pp. 237–247 (2001).
11. D. Vogel and B. Michel, Microcrack evaluation for electronics components by AFM nanoDAC deformation measurement, Proc. of IEEE-NANO 2001, Maui, Hawaii, Oct. 28–30, 2001, pp. 309–312.
12. D. Vogel, J. Keller, A. Gollhardt, and B. Michel, Displacement and strain field measurements for nanotechnology applications, Proc. of the 2002 2nd IEEE Conference on Nanotechnology, IEEE-NANO 2002, August 26–28, Washington D.C., 2002.
13. D. Post, B. Han, and P. Ilju, High Sensitivity Moiré, Springer-Verlag, Berlin, 1994.
14. R.S. Sirohi and F.S. Chau, Optical Methods of Measurement—Wholefield Techniques, Marcel Dekker Inc., N.Y., Basel, 1999.
15. B. Michel and R. Kuehnert, Mikro-Moiré-Methode und MikroDAC-Verfahren anwenden, Zeitschrift Materialprüfung, 38(6) (1996).
16. D. Vogel, J. Auersperg, A. Schubert, B. Michel, and H. Reichl, Deformation analysis on flip chip solder interconnects by microDAC, Proc. of Reliability of Solders and Solder Joints Symposium at 126th TMS Annual Meeting & Exhibition, Orlando, 1997, pp. 429–438.
17. D. Vogel, V. Grosser, A. Schubert, and B. Michel, MicroDAC strain measurement for electronics packaging structures, Optics and Lasers in Engineering, 36(2), pp. 195–211 (2001).
18. M.A. Sutton, W.J. Wolters, W.H. Peters, W.F. Ranson, and S.R. McNeill, Determination of displacements using an improved digital correlation method, Image and Vision Computing, 1(3), pp. 133–139 (1983).
19. Y.-J. Chao and M.A. Sutton, Accurate measurement of two- and three-dimensional surface deformations for fracture specimens by computer vision, in J.S. Epstein, Ed., Experimental Techniques in Fracture, VCH Publishers, N.Y., 1993, pp. 59–93.
20. M.A. Sutton, S.R. McNeil, J.D. Helm, and M.L. Boone, Measurement of crack tip opening displacement and full-field deformations during fracture of aerospace materials using 2D and 3D image correlation methods, IUTAM Symp. on Advanced Optical Methods and Applications in Solid Mechanics, 2000, pp. 571–580.
21. D.L. Davidson, Micromechanics measurement techniques for fracture, in J.S. Epstein, Ed., Experimental Techniques in Fracture, VCH Publishers, N.Y., 1993, pp. 41–57.
22. D. Vogel, A. Schubert, W. Faust, R. Dudek, and B. Michel, MicroDAC—a novel approach to measure in-situ deformation fields of microscopic scale, Proc. of ESREF'96, 1939-42, Enschede, 1996.
23. D. Vogel, J. Simon, A. Schubert, and B. Michel, High Resolution Deformation Measurement on CSP and Flip Chip, Technical Digest of the Fourth VLSI Packaging Workshop of Japan, 84–86, Kyoto, 1998.
24. D. Vogel, E. Kaulfersch, J. Simon, R. Kühnert, A. Schubert, and B. Michel, Measurement of thermally induced strains on flip chip and chip scale packages, Proc. of ITherm 2000, Las Vegas, USA, May 23–26, 2000, pp. 232–239.
25. E. Soppa, P. Doumalin, P. Binkele, T. Wiesendanger, B. Bornert, and S. Schmauder, Experimental and numerical characterization on in-plane deformation in two-phase materials, Computational Material Science, 21, pp. 261–275 (2001).
26. L. Cretegny and A. Saxena, AFM characterization of the evolution of surface deformation during fatigue in polycristalline copper, Acta Mater., 49, pp. 3755–3765 (2001).
27. K. Pinardi, Z. Lai, D. Vogel, Y.L. Kang, J. Liu, Sh. Liu, R. Haug, and M. Willander, Effect of bump height on the strain variation during the thermal cycling test of ACA flip-chip joints, IEEE Trans. on Comp. and Pack. Techn., 23(3), pp. 447–451 (2000).
28. A. Schubert, R. Dudek, H. Walter, E. Jung, A. Gollhardt, and B. Michel, Lead-free flip-chip solder interconnects—materials mechanics and reliability issues, Proc. APACK: Int. Conf. on Advances in Packaging, Singapore, Dec. 5–7, 2001, pp. 274–287.
29. J. Puigcorbé, D. Vogel, B. Michel, A. Vilà, N. Sabaté, I. Gràcia, C. Cané, and J.R. Morante, High temperature degradation of Pt/Ti electrodes in micro-hotplate gas sensors, J. Micromech. Microeng., (13), pp. 119–124 (2003).
30. C.B. O'Neal, A.P. Malshe, W.F. Schmidt, M.H. Gordon, R.R. Reynolds, W.D. Brown, W.P. Eaton, and W.M. Miller, A study of the effects of packaging induced stress on the reliability of the sandia MEMS microengine, Proc. of IPACK2001: The Pacific Rim/ASME International Electronic Packaging Technical Conference and Exhibition, Hawaii, USA, July 8–13, 2001.

31. D. Vogel, Chen Jian, and I. de Wolf, Experimental validation of finite element modeling, in G.Q. Zhang, Ed., Benefiting from Thermal and Mechanical Simulations in Micro-Electronics, Kluwer Academic Publishers, Boston, 2000, pp. 113–133.
32. J. Auersperg, R. Döring, and B. Michel, Gains and challenges of parameterized finite element modeling of microelectronics packages, in B. Michel, Ed., Micromaterials and Nanomaterials, No. 1, Fraunhofer IZM, Berlin, 2002, pp. 26–29.
33. J. Puigcorbé, A. Vilà, J. Cerdà, A. Cirera, I. Gràcia, C. Cané, and J.R. Morante, Thermo-mechanical analysis of micro-drop coated gas sensors, Sensors and Actuators A, 97, pp. 379–385 (2002).
34. I. Hamerton, Chemistry and Technology of Cyanate Ester Resins, Blackie Academic and Professional, Glasgow, 1994.
35. T.L. Anderson, Fracture Mechanics, CRC Press LLC, Boca Raton, 1995.

10

Interconnect Reliability Considerations in Portable Consumer Electronic Products

Sridhar Canumalla[a] and Puligandla Viswanadham[b]

[a]*Mobile Devices Unit, Enterprise Solutions, Nokia, 6000 Connection Drive, IRVING, TX 75039, USA*
[b]*Nokia Research Center, Nokia, 6000 Connection Drive, IRVING, TX 75039, USA*

10.1. INTRODUCTION

Revolutionary changes have taken place in digital information processing in recent years—the world has gone wireless and life has gone mobile. Handheld computers, personal digital assistants, mobile phones, computer controlled domestic appliances and other portable consumer electronic hardware have become pervasive in daily life. These mobile consumer electronic products are considerably different from other consumer electronic products from a variety of perspectives. In addition to being function-rich, lightweight, and portable, they often serve as fashion accessories. Hence, in the design and construction of mobile electronic hardware, visual appeal needs to be considered in addition to durability. Indeed, industrial design and functionality guidelines often take precedence over design for durability.

Most hand held/portable consumer electronic products can be characterized as high volume, low cost devices. In a competitive business environment, manufacturing and product development costs, time-to-market, functionality, yields, customer satisfaction and product quality all have an impact on the business. Original equipment manufacturers face pressure to develop new, more advanced technology products in record time, while at the same time improving productivity, product field reliability and overall quality. Although product quality encompasses several measures, the relevant link between quality and reliability can be described as follows. Reliability is defined as the probability that a product will perform its intended function under encountered operating conditions for a specified period, whereas quality, in narrow terms of reliability alone, can be defined as the reliability at time zero. While there are other definitions to quality, there is general agreement that an unreliable product is *not* perceived as a high quality product [1].

TABLE 10.1.
Comparison of typical application conditions of desktop, mobile, and automotive hardware.

Hardware type	Product life/yrs	Power-on cycles/day	Power-on hours	Relative humidity/%	Environment temperature range/°C	Operational temperature range/°C	Voltage/ V
Desktop	5	1−17	13,000	10−80	10−30	20−60	12
Mobile terminal	5	20	43,800	10−100	−40−40	32−70	1.8−3.3
Automotive under-the-hood	15	5	8200	0−100	−40−125	−40−125	12

*Source JEDEC.

The highly personal use profile and mobility for these products implies that consumers will take these products with them wherever they go, and expect the same dependable performance irrespective of the exposure of the product to the elements, for example, rain, snow and accidental drop. Under such conditions, meeting the reliability expectations for portable products can be a challenge, especially since reliability expectations need to be met without compromising profitability. Therefore, one primary driver for product reliability is perceived quality and customer satisfaction. Another reason for ensuring reliability is that product field-failure rate, which plays a key role in controlling warranty and repair costs, tends to be higher for an unreliable product. In other words, all other factors remaining the same, a more reliable product will be more profitable. However, in reality, there is a level of optimum reliability beyond which additional reliability improvements have a decreasing rate of return. Therefore, it is prudent to develop products that meet specific business or customer requirements driven reliability target rather than aiming to have the most reliable product possible at the expense of profitability or time-to-market. A third reason to strive for product reliability is that reliability (and quality) could be employed as product differentiators in product marketing (advertising), which will only increase the business value of product reliability.

The operating environment for mobile electronic equipment also differs considerably from that for desktop or business computers, and is often more varied in terms of thermal excursions, and exposure to humidity and corrosive environments. The products are more prone to high humidity exposures in both non-condensing and condensing atmospheres. Additionally, portability makes the product more likely to experience mechanical loads such as drop, bend, twist, etc. Mechanical drops from such heights as a meter and half on hard surfaces are not uncommon. The number of power ON cycles, the operational voltages, and other conditions are also different. Table 10.1 shows a comparison of the typical operating environments for portable hand-held telecommunication devices with conventional desktop and automotive under-the-hood electronics.

Irrespective of the operating environment, programs for producing reliable products require quantitative methods for predicting and assessing various aspects of product reliability. This involves the collection of reliability data from the following [2]:

1. Laboratory life tests to assess product reliability.
2. Degradation tests of materials, devices, and components.
3. Design of experiments for reliability improvement.
4. Tests on early prototype units to learn about possible failure modes and mechanisms.

5. Monitoring of early-production units in the field.
6. Analysis of warranty data and samples from warranty population.
7. Systematic longer-term tracking of product in the field.

The need for shorter design cycle time is a driver for reducing the time and resources spent in reliability testing. Since non-accelerated tests can take an excessively long time to yield valuable data for reliability improvement, different kinds of accelerated tests have been developed to estimate relatively quickly the failure-time distribution or long-term performance of the product in the field, based on a careful study of the operating environment. However, in and of itself, accelerated life tests to *assess* reliability do not yield actionable data to *improve* reliability. Analysis of the failures to uncover failure mechanisms and the root causes of failure are crucial for formulating corrective actions that can improve reliability. Sometimes, a second round of reliability tests may be required to assess the reliability of the improved products.

The focus of this chapter is on the physics behind accelerated laboratory life tests to assess reliability in thermal, mechanical and electrochemical environments. Since failure analysis is ideally an integral part of any reliability assessment and improvement exercise, some representative failure mechanisms commonly observed in each of these reliability tests will also be discussed.

The study of interconnection reliability, until a few years ago, was driven primarily by the computer industry. Therefore, the vast majority of literature on electronic packaging reliability is comprised primarily of thermal cycling reliability, and to a lesser extent, corrosion and electromigration phenomena. In fact, there was a tacit assumption that reliability always implied thermal cycling reliability. As such, the titles of some reliability publications did not even indicate that the investigation pertained only to thermal cycling. Until recently, this did not cause any serious consternation among the packaging community in the days where much of the information processing hardware was confined to environments with controlled temperature and humidity.

The material discussed in this chapter is intended to introduce interconnection reliability issues in thermal, mechanical and electrochemical environments for portable, consumer electronic products to readers who are primarily familiar with similar issues in business, office and telecommunication applications. The scope of the chapter is limited to interconnection reliability and excludes important topics such as electromechanics or liquid crystal display issues, which are complex enough to justify a separate chapter.

10.2. RELIABILITY—THERMAL, MECHANICAL AND ELECTROCHEMICAL

10.2.1. Accelerated Life Testing

Some of the pitfalls of accelerated life tests (ALTs) need to be considered to avoid seriously incorrect inferences about the product reliability in the field based solely on laboratory tests [1]. The following aspects need to be recognized when interpreting the results of ALTs:

1. Multiple or unrecognized failure mechanisms—high levels of accelerating variables can induce failure mechanisms that would not normally be observed at operating conditions. For example, instead of just accelerating corrosion or electrochemical migration, higher temperatures may cause melting or material deformation or degradation. Higher humidity may cause swelling and delamination. In less extreme

cases, high levels of accelerating variables will change the relationship between life and the variable. If different failure mechanisms are operative at high levels of the accelerating variables, and this is recognized, failure times for that mechanism can be censored out. Sometimes, such censoring can result in inadequate data. If the presence of undesirable failure mechanisms is not recognized, it is possible that seriously incorrect inferences are drawn.

2. Failure to properly quantify uncertainty—it is important to recognize that all statistical estimates have some uncertainty associated with them. Using point estimates alone can be misleading in many cases. Uncertainty can result either from the experiment or from the model, and in general, statistical confidence intervals do not account for model uncertainty. Extrapolations, fundamentally, are fraught with errors, especially when based on inadequate sample sizes and point data. Performing a sensitivity analysis to assess model uncertainty or testing adequate number of samples is one solution.
3. Multiple time scales and degradation affected by more than one accelerating variable—in ALT, particularly when there is more than one failure mechanism, it should be recognized that all mechanisms may not be accelerated in the same manner. For example, when performing ALT of solder interconnections under accelerated conditions, creep and fatigue are accelerated differently depending on ramp rates and hold times at the different temperatures.
4. Masked failure mechanism—if there is more than a single failure mechanism, it is possible that one mechanism is accelerated more than others. In such cases, the masked failure mechanism will not show up in laboratory testing but can dominate field failures. It is not only prudent, but also cost effective, to verify that the failure mechanisms seen in the field are the same as the failure mechanisms observed in accelerated testing.
5. Faulty comparison—a popular use of ALT is in the comparison of alternative designs or materials from vendors, in addition to its use in predicting field reliability. The rationale behind is that if material from one vendor or one design performs better in laboratory tests, relative field reliability would follow a similar relationship. However, in cases where the reliability in the field is governed by a different failure mechanism than that observed in the laboratory test, ALT results can mask the actual field performance and serve as the basis for inaccurate prediction of field reliability.
6. Accelerating variables can cause deceleration—the most common examples involve failure mechanisms that require specific combinations of humidity, stress, and temperature. For example, when the usage rate is accelerated for a connector undergoing wear, the accelerated test can inhibit a secondary corrosion failure mechanism by continuously removing corrosion products and not giving enough time for the reaction to occur. Another example is failure due to tin whisker formation, which has a high propensity at a certain temperature and humidity for certain substrate and coating compositions and thicknesses. Optimum temperatures for tin whisker growth have been reported to be between 50 and 70°C by several researchers, for example [3]. Unfortunately, since the working temperature of most electronic equipment is relatively close to the optimum temperature for tin whisker growth, an injudicious selection of temperature acceleration can yield incorrect results.
7. Differences between prototype and production samples—It is important to test units manufactured under actual production conditions, using materials and parts that

will be employed in actual production samples. Sometimes, test methods capable of handling functional products may need to be developed. For example, ball-shear tests are widely used to assess the quality of the ball attachment process for area array packages such as ball grid array packages (BGAs) or chip scale packages (CSPs) [4]. However, the ball-shear test method cannot be applied to assess the interconnection quality or strength in a functional product after surface mount assembly because an individual ball is no longer accessible for test. To accommodate interconnection strength data requirements on functional products, tests such as the package-to-board interconnection strength test method (PBISS) can be used [5,6].

10.2.2. Thermal Environment

Historically, for office and business machines, accelerated thermal cycling tests are carried out in the 0 to 100°C range with 10–15 minute dwell times at ramp rates in the 10–15°C/minute range. A life requirement of 1000 cycles translates into a product life of about 7–10 years. These machines hardly experience other mechanical stresses in the operational environment. In contrast, hand held electronic hardware can experience extreme ambient temperature fluctuations in the range of −30°C to 45°C depending on the geographic location. When the appliance is left in an automobile it can experience even more severe temperature conditions depending on the climate and diurnal variations. Thus, accelerated thermal cycling tests applicable to business machines will be not be severe enough to assess the performance of handheld electronic appliances. Another difference in regard to the portable hardware is the shorter product design life. The average product design life is in the range of 2 to 5 years instead of the 7 to 10 years in other consumer products such as desktop machines.

Owing to the aforementioned considerations, portable electronic PWB assemblies are generally subjected to accelerated thermal cycling of −40°C to 125°C for 200 to 800 cycles in order to assess the product performance.

10.2.3. Mechanical Environment

One way to classify the mechanical environments for a portable electronic product is based on the rate of deformation: (a) low deformation—as experienced in bending and twisting, (b) medium to high deformation rate—as experienced in vibration or (c) high rate of deformation—as in case of drop or shock. Another way to characterize the environment is based on the life expectancy in number of fatigue cycles as being high cycle fatigue (vibration) or low cycle fatigue (drop, bending and twisting). In comparison to thermomechanical reliability, relatively little has been published in the public domain on reliability under mechanical loading. Broadly, mechanical loading can be divided into the following categories

(1) Drop or impact loading—typically high strain rate loading that can also cause bending and twisting of the product due to impact forces. The number of cycles to failure is generally low.
(2) Bending and twisting—typically low strain rate events such as encountered during key presses. The life expectancy is generally a few hundred cycles.
(3) Vibration loading—typically high strain rate loading with low amplitude. In general, vibration failures are of relatively less concern in portable electronic products.

In addition, reliability evaluations of portable electronic products can also involve either shear or pull testing performed at the interconnection or package level for purposes of determining the strength distribution. It is pertinent to include them in the discussion because shear and pull tests serve to define the strength of the interconnection between the package and PWB, which is closely related to reliability in drop, bend, twist or vibration loading.

10.2.3.1. Drop or Impact Environment When portable electronic products are subjected to mechanical drop or impact, it is important to recognize that failure can occur (a) at the solder or other interconnects, (b) connector or spring contacts, (c) inside the components such as LCD, housing, lens, etc. or (d) at the system level. Usually, these failures are due to the following causes:

(a) High inertial forces (g-forces) due to rapid change in velocity upon impact,
(b) Large strains in the solder interconnects between the PWB and package due to excessive dynamic buckling, flexure, or twisting of the PWB, and/or
(c) Shock waves that travel through the product assembly upon impact.

It is reasonable to assume that all three effects can co-exist during any single event and that the interactions among them can be relatively complex.

The drop tests carried out can be at the product level, or at the board assembly level. Product level drop tests involve tests on the entire product including the housing, while the board level drop tests are performed on just the PWB assembly with components mounted on it, as described below.

10.2.3.1.1. Product Level Drop or Impact Testing. Product level drop testing can be classified as constrained or free. In constrained drop testing, which is by far the most common, the product is clamped rigidly to a heavy table that is guided along vertical rails to have a single impact against a target surface.

Clatter is probably best understood in terms of drop impact of an elongated or flat object onto a surface. Invariably, one corner touches down first, the object begins to rotate, and clattering occurs as the various corners encounter the impact surface before the object finally comes to rest. In that sense clatter refers to the condition where the second or third impact of the object probably occurs before the deformation from the first impact has returned to zero. On an oscilloscope time readout, strain or acceleration data due to clatter will resemble an extended but single impact sequence. In contrast, multiple impacts refer to the condition when the object bounces up and lands at a different location and orientation. In the case of multiple impacts, it is probable that the deformation in the assembly has had a chance to return to zero after the first impact, and the second impact occurs a short time later. On a time scale, they appear as two distinct events rather than as a single event. A third type of secondary impact, chatter, refers to the condition when subsystems or components impact each other within the product, for example, a battery impacting the case or a component. On an oscilloscope display monitoring the deformation-time response, chatter will appear as two events superimposed on each other and not distinctly separated in time. One of the main effects of the resultant secondary impacts in real life situations is that, depending on the moment and the coefficient of restitution, the ends of the object can strike at much higher velocities than during the first impact. The increased amplitude of velocity shocks, the possibility of exciting resonant conditions and repetitive shocks are reasons why the damage in a “real life” drop can be significantly higher [7].

On the other hand, free fall testing replicates the abuse a portable product will experience in actual usage. The main disadvantage is that it is difficult to control the orientation

of the product at impact and this affects the repeatability of the test results and ease of monitoring by instrumentation. There is little mention of experiments in literature where free fall drop testing has been automated. Goyal et al. [8] proposed that the object being tested be suspended onto the guided drop table in the precisely desired drop orientation, and that, just before impact, the object is released from its suspension. The intended result, theoretically, is that although the required orientation at first impact is maintained, the object is free to move unconstrained subsequent to the first impact. Another variation of the quest for greater repeatability in product level drop testing consists of using grippers to control the orientation of the product until just before impact [9].

Lim et al. [10] surveyed the response of several commercial portable products (Nokia 8250 and 8310, Sony Ericsson T68i, Compaq 3850, HP Palm m105 and m505) using strain gauges and accelerometers to monitor the response of the PWB during drop tests. Maximum strain values ranged from 500 to 2500 microstrain and varied considerably between the different drop orientations depending on the product. Although the horizontal drop orientations generally yielded the highest strains in the PWB and the highest accelerations, there was considerable variability, which indicates that the actual behavior is quite complex and eludes simple generalizations.

In a study of the role of the rigidity of the mobile housing in determining the impact tolerance [8], it was found that thin-walled clamshell case constructions, currently favored for its size and weight advantages, may not provide sufficient rigidity to impact induced loads. Housing modifications to increase the stiffness improved the drop reliability. In addition, it is believed that the drop tolerance of the mobile phone would improve if the battery pack were to remain firmly attached to the phone, minimizing velocity amplifications and possible chattering.

The Shock Response Spectrum (SRS) approach was applied in using compliant suspensions to reduce peak acceleration and increase drop impact performance [11]. Results from another study with a personal digital assistant (PDA) using accelerometers and strain gauges, located along both the longitudinal and transverse directions, suggests that although there is a reasonably good correlation between acceleration and strain, it is often very difficult to completely unravel the complex strain-time or acceleration-time data except in select orientations [12].

10.2.3.1.2. Board Level Drop Testing. Because of the complexities inherent in product level drop testing, alternative ways of estimating the product reliability from simpler tests have received much attention. One such technique is the board level drop test, where the PWB assembly is subjected to impact loads or high accelerations while measuring the acceleration, velocity and strain on the assembly. Such board level drop tests provide a common basis to evaluate the impact tolerance of electronic products if one assumes that the conditions during product level drop impact can be reproduced adequately by dropping a test PWB assembly. The advantages of board level testing are:

(a) the shock pulse amplitude can be fairly well controlled,
(b) the orientation of the PWB assembly is controlled closely, and
(c) the tests are relatively more repeatable.

The primary disadvantage is that the test is not a true reflection of reality because it does not include the effect of secondary effects such as clatter, chatter or multiple impacts, which can have a significant bearing on reliability.

Ong et al. [13] examined the relevance of a board level drop tester by comparing it with the data collected from an instrumented drop of a Nokia 3210 model phone. It was

found that, for the product level drop test, depending on the orientation of drop, the impact force can vary by up to a factor of five. Their results indicate that an axial impact exerted the highest forces. Further, because of the possibility of multiple impacts, the damage induced in a single drop in a product level test may actually be much higher than the damage induced due to a single drop in a board level test. In addition, Ong et al. [13] report that in board level drop tests, flexure of the PWB can last much longer than in product level drop tests.

Despite these differences, board level drop tests are attractive for investigating package reliability and process quality issues. Mishiro et al. [14] observed a correlation between solder joint stresses and PWB strains in a study where numerical analysis and strain measurements were employed to assess CSP reliability for 3 different package constructions. Even if the PWB strain is the same, the package structure played a significant role in controlling the solder joint stresses and hence drop impact reliability. In particular, the package structure with a 0.15 mm thick elastomer between the die and polyimide substrate performed better than the package where the interposer consisted of a multilayer laminate, which in turn was better than package with only a polyimide substrate. Further, nonsolder mask defined (NSMD) pad structure was shown to be significantly better than solder mask defined (SMD) pad structure for drop reliability. With regard to PWB build-up layer, aramid-epoxy PWBs with low adhesive strength performed poorly because of premature delamination in the build-up layer. Underfilling the package to board interspace was found to improve the reliability when the Young's modulus was sufficiently high. However, when the underfill modulus was low (5 MPa), however, drop test reliability was much worse. Similar results were also reported in another study, where underfilling the CSP improved the reliability significantly in drop loading and the degree of improvement depended on the underfill modulus [15]. However, if the underfill quality was not optimal, the presence of even a small void encompassing the corner solder joint can magnify the stresses in the solder joint, effectively negating any anticipated benefit of underfilling.

Recognizing the relative complexity of a product level drop test, the relatively simpler board level drop test has been used to quantify drop reliability in terms of the package structure, materials, and processing. For example, Hannan and Viswanadham [16] evaluated the drop reliability of CSPs with reworkable underfills for reliability enhancement. Kujala et al. [17] used a board level drop test to compare the relative performance of land grid array (LGA) package and CSPs under both thermal cycling and drop impact. The board level drop test was used as a means to study the reliability of a "corner-reinforced-only" CSPs for portable product applications [18], where the CSP was held down only at the corners with epoxy, without actually having any underfill surrounding the solder joint in the package to board interspace. The drop reliability of such corner reinforced CSPs was lower than in the case of complete, capillary underfill. However, the relatively modest 3–4$\times$ improvement in the performance may be sufficient for some portable product applications [19]. Board level drop tests have also been used to investigate the effect of PWB and component pad surface finish, and concomitant interfacial strength, on drop test performance, and this is discussed in a later section.

10.2.3.1.3. Simulation of Drop Test Behavior of PWB Assemblies and Products. Faced with the complexities of purely empirical product level drop testing, there have been several attempts to complement experimental studies with finite element simulation to better understand the drop phenomena. The key issues for a successful understanding of drop impact reliability are (a) sophisticated and consistent analysis tools, (b) test correlation for model validation and refinement, (c) specification to define reliability requirements, and (d) material property data, especially over a broad range of strain rates [20].

Simulation when combined with board level drop testing can enable accurate prediction of not only the failure location but also durability to within 10% [21]. It was found that drop orientation with the components oriented face down was a more stringent test condition than one with the components facing up. Results indicate that during the drop test, greater PWB bending induces larger stress to the solder joints. As anticipated, it was found that the outermost solder joints have larger stresses and that smaller PWBs enhance drop performance. More importantly, it was reported that the lead-free solder studied had better board level thermal cycling reliability but worse drop test reliability. It should be recognized that thermomechanical reliability alone does not assure product reliability under mechanical loads.

Relatively accurate correlation of model prediction and experimental data were reported using smeared property models [22]. Significant error may be introduced due to aliasing of the experimental and computational data and under-sampled experimental data acquisition may mask the recognition of peaks in strain or displacement. In addition, smeared property models may not capture structural degradation during successive drops produced due to progressive delamination between materials. A validated modeling technique can be used to accurately predict failures observed in portable electronic products, such as disengagement of snap-fit housings and CSP solder joint cracking [23].

While state-of-the-art simulation was shown by various people to accurately predict different aspects of the drop test, a combination of simulation and experiments can be expected to be the most effective approach for improving and predicting reliability under drop or impact loading.

10.2.3.1.4. Analytical Modeling of Drop Phenomena. In addition to numerical simulation, closed form analytical modeling has been employed to understand the physics behind drop related phenomena. Suhir [24] obtained formulae to calculate the maximum displacements, velocities and accelerations of surface mounted devices when a shock load is applied to a flexible PWB at its support contour. Consideration of the nonlinearity of the PWB vibrations was found to be important in the case of large shock-induced deflections. The dynamic response of a rectangular plate element assembly subjected to drop impact was simulated as a box within a box, with one gasket between the outer and inner boxes and another between the PWB and the inner box. Results suggest that lower g-forces can be ensured by having the lower natural frequency considerably different from the higher frequency [25]. For example, the inner cushioning gasket could be made substantially stiffer than the outer one. Probabilistic approaches could also be employed to ensure a low failure rate. The effect of the stiffness of a "spring" shock protector was also studied [26]. Because the possibility of a "rigid impact" needs to be avoided at all costs, if the maximum drop height is not known, the advantages afforded by a soft spring cannot be fully utilized. The effect of viscous damping on the maximum displacement and the acceleration of a 1-DOF linear system subjected to a shock load during drop impact was also studied [27]. Sometimes, the application of materials with high energy absorption can result in even higher acceleration levels, and this needs to be avoided by a careful consideration of the system's mass and spring constants.

Whether maximum acceleration is an adequate criterion of the dynamic strength of a structural element in an electronic product has been investigated using a simply supported beam and a cantilever with heavy end mass [28]. Surprisingly, it was found that even if the accelerations experienced are not severe, one can expect significantly high dynamic stresses. These results are supported by observations during product level drop tests, where the bending of the board plays a bigger role in controlling failure compared to purely inertial forces, especially for light components such as flip chips and CSP assemblies. Until

recently however, acceleration has been measured preferentially because it is easier to measure. Regarding alternatives to drop testing, it was found that the applicability of shock tests to replace product level drop tests depends on whether the dominant frequency of the shock impulse (which is inversely proportional to duration) is sufficiently high in comparison to the fundamental frequency of the vulnerable structural element [29].

10.2.3.2. Bend or Twist Environment Most portable electronic products experience more severe PWB bending related stresses than thermal stresses. PWB bending failure in the creep regime can be caused by localized bending near a screw location or in the high cycle fatigue regime due to key press action. A third bending failure mode occurs when portable products are dropped [30]. Recognizing the importance of understanding the reliability under bending loads, several studies in recent years have been aimed at characterizing the deformation and failure of solder joints.

Darveaux and Syed [30] have used both 3-point and 4-point bending tests to examine the failure mechanisms under a range of conditions for different CSPs along with finite element simulation of the damage processes. For displacement controlled fatigue tests, life decreased with (a) reduction of span length, (b) increase in test board thickness, (c) increase in die size, and (d) increase in molding compound thickness. In load-controlled tests, which are more closely related to actual product reliability, opposite trends were observed. Simulation results indicate that the optimum component/PWB pad size ratio in bending is different than under thermal loading.

The failure modes observed can be summarized as (a) fracture in the solder or in the intermetallic layer at the component pad, (b) fracture in the solder or in the intermetallic layer at the PWB pad, (c) trace peeling and eventual laminate cracking of the PWB or the component, or (d) build-up layer fracture leading to trace cracks on the PWB. Since these are very similar to the failure modes frequently seen under drop or impact conditions, bend testing is generally perceived to be a relatively simpler alternative to more complicated drop tests. Improving the strength of the weakest failure link can offer improvements in performance. For example, in 3-pt bend and drop impact tests on CSPs, anchoring the pads with via-holes improved the performance over having no-via-in-pads [31].

A few studies have been reported on the effect of strain rate during the bending test. For example, Geng et al. [32] reported that the solder joint interconnection fails at approximately 50% lower board deflection when the test speed increases by two orders of magnitude (0.25 to 2.54 cm/s). It is relatively well known that although solder strength increases with increasing strain rate, strain to failure decreases. In that context, as long as failure occurs in the solder, solder joints can be expected to fail at lower strain rates in high displacement rate bending tests. However, the data does not show a very distinct trend at higher strain rates (25.4 cm/s), possibly due to experimental artifacts. In a different study, with increasing ram displacement rate in a 4-pt bending test, strain gages mounted on the PWB showed increasing strain at solder joint failure sites [33]. It was shown that Kirkendall voids at the intermetallic interphases between the Ni and the Ni-Sn-P layers degraded the interfacial strength enough to cause failure preferentially at these locations. It should be noted that Kirkendall-like voids were also reported at Cu-Sn interfaces in lead free solder joints on OSP pads by Chiu et al. [34], with severe drop performance degradation in strength upon thermal aging in the 100 to 150°C range. Some reliability studies also focused on testing methods for flexible or low stiffness PWBs. Rooney et al. [35] reported an offset bend test configuration that is useful for testing assemblies with thin PWBs (0.5 mm) having stiff components.

A planar 3-pt bending fatigue test method to assess the reliability of the CSP solder joints was recently proposed [36]. The applicability of this method was demonstrated for standard plastic ball grid array (PBGA) components mounted on a PWB. The same method to establish that via-in-pad structure by itself does not pose a reliability risk in bend fatigue [37]. This is in accordance with the results reported previously by Juso et al. (1998). The applied load could induce dielectric (build-up layer) cracking, which in turn can lead to trace and via failures. Although lead free solders (Pb-free) have been found to be more durable than tin-lead solders in bend tests [38], it should be remembered that different lead free solders can be expected to behave differently, and some can perform worse than Sn-Pb solders depending on surface finish, test conditions, sample history, and several other variables. Moire interferometry coupled with 4-pt bend testing can reveal the localized influence of solder ball interconnections on chip carrier and PWB deformation [39]. Large shear strains were found in solder balls across the entire array. It was found that maximum strains occur in the outermost row of the solder balls, which agrees with the observations from a study on underfilled CSPs with corner defects [15]. In another study, the effect of cyclic bending on CSP assembly reliability was investigated in addition to monotonic bend bending [40]. The average overstress limit for a CSP studied was determined to be 2550 N mm. It was concluded that the CSPs showed worse durability when the PWB assembly subjected to negative curvature (CSP mounted surface of the PWB is convex). This is understandable since negative curvatures would subject the corner joints of the CSP to tension and lead to premature failure.

Portable electronic products were also evaluated for reliability under twist loads in addition to bend loads. For example, Perera [41] reported on the effect of twist loads of 9% and 12% and observed that solder joint failures occurred mostly by fatigue processes.

10.2.3.3. Shear Tests Interconnection failure is a common mode in portable electronic products, and it is widely accepted that interconnection strength and solder joint quality can play a central role in determining product reliability. Thus, measurements of the interconnection strength are useful in understanding reliability of the product. The term interconnection strength in this context denotes the effective strength of the package-to-board-interconnection, and includes the strength of (a) the package-solder interface, (b) solder, (c) the solder-pad interface and (d) the build-up layer on the PWB. This interconnection strength plays a role in determining product reliability.

Conventionally, the ball shear strength is used to denote the strength of attachment of a solder ball of a BGA or CSP to the component *prior to board assembly* [4]. This measure of ball strength, although useful in measuring the quality of the ball attachment process, cannot be easily translated into a product level estimate of durability. This is because of the following reasons: (a) the bare component is no longer accessible for ball shear tests, and (b) the interconnection quality is determined not only by the solder/component bond but also by the solder strength, solder/printed wiring board interfacial strength and build-up layer quality.

Product level tests such as mechanical drop, twist and bend tests yield valuable information on the reliability in the field. However, the primary drawbacks of these tests are the complexity and the time required to analyze the results in terms of targeted improvement actions. Thus, there is a need for a product level interconnection strength test that can yield relatively rapid results and simultaneously provide targeted quality improvement actions.

One candidate method is a recently developed product level test, the package to board interconnection shear strength (PBISS) technique [5,42]. It was shown that the shear test is an effective tool to quantify the shear strength of CSPs and examine the effect of pad

finish and build-up layer strength. Only low strain rate PBISS behavior was characterized because product level twist and bend tests are performed at a low strain rate.

However, the strain rates experienced by the solder joint during drop tests are significantly higher. Therefore, the shear strength behavior measured at slow deformation rates is not directly applicable as a proxy for drop reliability of portable electronic products. Solder behavior changes significantly with the rate of deformation, and the damage to the CSP interconnection can be expected to be significantly different also during high rate of strain [43,44,76]. In this context, it was demonstrated that high strain rate shear tests essentially mimic the failure mechanisms and relative performance observed in drop tests [6].

10.2.4. Electrochemical Environment

The failure mechanisms that are of importance in portable electronic products exposed to electrochemical environments can be described as:

1. Corrosion.
2. Electrochemical migration (ECM).
3. Conductive anodic filament (CAF).

The fundamental difference between the two is that corrosion involves the destructive attack of a metal by the environment as anodic oxidation without the necessity for electrical bias, whereas ECM involves the transport of metal ions from the anode to the cathode under the influence of an applied electric field. From a failure perspective, corrosion results in product failure primarily by causing electrical open or intermittent interconnections, while ECM results in failures primarily due to electrical shorts or intermittent connections. Some factors affecting these failure mechanisms are the environment (temperature, humidity, presence of corrosive elements), operating conditions (bias voltage, current density, temperature and conductor spacing), and materials (nature of metal or alloy, surface condition, ability to absorb humidity, coating composition and thicknesses).

10.2.4.1. Corrosion Corrosion, depending on the severity, results in the following failure pathways:

(a) Oxidative materials degradation resulting in loss of electrical continuity,
(b) Partial degradation of materials accompanied by the formation of conductive oxidation product, such as a salt, that could result in lower surface insulation resistance (SIR),
(c) Electrical shorts between adjacent conductive features, or
(d) Intermittent shorts or opens depending on the humidity levels and the ionic nature of the corrosion product.

Corrosion is often discussed in terms of half-cell reactions because all corrosion processes are essentially electrochemical reactions. The electrodes in question could be on the macro- or micro-scale. Macroscopic galvanic corrosion cells can occur when dissimilar metals are coupled electrically and exposed to a corrosive environment, while microscopic corrosion cells tend to occur on the scale of grains. In either case, oxidation occurs at the anode and reduction at the cathode. In other words, the metal dissolution occurs only at the anode. The medium or electrically conductive environment in which these chemical reactions proceed is usually referred to as the electrolyte even if the electrolyte may extend to a thickness of a few monolayers. Since all the cations produced by the anodic reaction are

consumed by the cathodic reaction, both anode and cathode reactions proceed at the same rate for corrosion to occur in a continuous manner.

The propensity of a metal to undergo corrosion is described in terms of the standard electrode potentials, where the hydrogen electrode potential is arbitrarily assigned a value of zero. When two dissimilar metals are coupled, the less noble metal will corrode in relation to the more noble one. However, it is possible to promote corrosion of the more noble metal in a galvanic couple by electrical biasing, which makes the more noble metal the anode. Some forms of corrosion [45] that are relevant to portable electronic products are:

1. Uniform corrosion—this form of corrosion is evenly distributed over the surface, and the rate of corrosion is the same over the entire surface. A measure of the severity is the thickness or the average penetration.
2. Pitting and crevice corrosion—this localized form of corrosion appears as pits or crevices in the metal. The bulk of the material remains passive but suffers localized and rapid surface degradation. In particular, chloride ions are notorious for inducing pitting corrosion, and once a pit is formed, the environmental attack is locally autocatalytic.
3. Environmentally induced cracking—this form of corrosion occurs under the combined influence of a corrosive environment and static or cyclic stress. A static loading driven cracking is called stress corrosion cracking and a cyclic loading driven cracking is called corrosion fatigue. Residual stresses in electronic leads from lead bending operations were observed to cause stress corrosion cracking failures in the presence of moisture [46]. Stress corrosion cracking of package leads was also reported in the presence of solder flux residues [47].
4. Galvanic corrosion—this type of corrosion is driven by the electrode potential differences between two dissimilar metals coupled electrically. The result is an accelerated corrosive attack of the less noble material. Galvanic corrosion tends to be particularly severe if the anodic surface is small compared to that of the nobler cathode or cases where a nobler metal is coated onto a less noble one. For instance, when a porous Au plating over a Ni substrate is exposed to a corrosive environment, the gold coating acts as a large cathode relative to the small area of exposed Ni. This sets up a galvanic cell at the exposed substrate which experiences intense anodic dissolution. It has been observed that pore corrosion can be enhanced by a galvanic corrosion process when the substrate metal is less noble than the coating, and vice versa [48].

10.2.4.2. Electrochemical Migration The distinguishing feature of ECM from corrosion is the formation of dendrites that cause a short between adjacent conductors. There are some similarities to corrosion as well, and the oxidation of the metal at the anode is common to both processes. ECM, which is also known as migrated metal shorts [49,50], is probably best described as due to transport of ions between two conductors in close proximity, under applied electrical bias and along an electrically conductive medium. In general, three conditions are necessary and sufficient for ECM failures to occur, and they are (1) presence of sufficient moisture (sometimes as little as a few monolayers), (2) presence of an ionic species to provide a conductive medium, and (3) presence of an electrical bias to drive the ions from the anode to the cathode. In the presence of sufficient moisture, the process is accelerated by temperature, and several mechanisms of ECM have been in vogue.

The first step in the classical model of ECM consists of metal ion formation by anodic oxidation (similar to corrosion), which may be either direct electrochemical dissolution or a

multi-step electrochemical process. At the anode, for example, where M represents a metal atom,

$$M \rightarrow M^{n+} + ne^{-}.$$

The second step is the transport of metal ions from the anode, through an electrolyte, toward a cathode. In the final step, at the cathode, the positively charged ions are reduced to a neutral metal atom. At the cathode

$$M^{n+} + ne^{-} \rightarrow M.$$

Successive cationic reductions facilitate the growth of dendrites toward the anode along energetically favorable crystallographic orientations. Therefore, the surface insulation resistance of the material progressively decreases as the migration advances toward the anode. Eventually, an electrical short results when the dendritic filament touches the anode. Silver [49], Cu, Pb, Sn [51,52,78], Mo and Zn [77] have all been observed to form dendrites by this process. The presence of flux containing ionic species is a known contributor to ECM and has been studied widely using surface insulation resistance measurements [53]. Following the migration ability of pure metalization systems, the propensity for ECM may be ranked as follows: Ag > Pb > Cu > Sn [54,55].

A second mechanism of ECM was proposed to explain the migrated metal short formation involving noble metals such as Au, Pd and Pt. Because of the relative chemical inertness of these metals, a halogen contaminant is needed to induce anodic dissolution [56,57]. In an acidic medium, a positively charged metal ion may form by the following route at the anode,

$$Au + 4Cl^{-} \rightarrow AuCl_4^{-} + 3e^{-},$$

$$AuCl_4^{-} + H^{+} \rightarrow H[AuCl_4^{-}] \rightarrow HCl + AuCl_3 \rightarrow H^{+} + 4Cl^{-} + Au^{3+}.$$

These positively charged Au ions can migrate toward the cathode and form dendrites in a similar fashion as the classical model.

A third mechanism to explain the ECM of Ni starting at the anode involves the presence of a strongly alkaline electrolyte. The first step is the formation of a cation ($HNiO_2^{-}$) by anodic corrosion followed by a chemical process resulting in secondary ionic species [58].

$$Ni \rightarrow Ni^{2+} + 2e^{-},$$

$$Ni^{2+} + 2OH^{-} \leftrightarrow Ni(OH)_2.$$

It is suggested that instead of migrating to the cathode, the Ni^{2+} ions thus formed undergo the following reaction to form an anionic complex

$$Ni^{2+} + 2H_2O \rightarrow HNiO_2^{-} + 3H^{+}.$$

This anion complex migrates through the electrolyte under the applied electrical field. Finally, the metal atoms are deposited at the anode in the form of metallic dendrites due to the electrochemical reaction of the cationic species with the H^{+}, Ni^{2+} or OH^{-} ions. Similar process could be operative for Co and Cu ECM as well in cases where anodic deposits of the metal are observed.

10.2.4.3. Conductive Anodic Filament (CAF) Formation CAF is the type of electromigration failure mechanism where the loss of insulation resistance between neighboring conductors is caused by the growth of a subsurface anodic filament along delaminated fiber/epoxy interfaces [59]. The first step in the formation of the CAF is the physical degradation of the fiber/epoxy bond. This is followed by an electrochemical reaction requiring both the presence of moisture and a potential gradient across the cathode and anode. The metal undergoes oxidation at the anode to yield a positively charged ion that migrates toward the cathode. As the metal species migrate toward the cathode, they precipitate at locations where the pH is thermodynamically conducive, and in time, the filament extends from the anode to the cathode causing a short. The CAF formation may occur along the surface of a PWB or between conductors in different layers separated by a dielectric or along the glass fibers in the weave [60].

10.2.5. Tin Whiskers

Single crystal whiskers, of several metals including Sn, Cd, Zn, Sb, In, Pb, Fe, Ag, Au, Ni and Pd have been reported (for example, [61–63]). While the mechanism for the growth of whiskers of different metals may possess similarities, the mechanism of Sn whisker growth has been studied extensively. However, due to recent emphasis on the implementation of Pb-free solders and the consideration of Sn as a component terminations and PWB finish, there has been an increased effort to study the reliability implications of Sn whiskers. Several reported field failures have been collected from medical, military, and space applications by Siplon et al. [62]. It is generally agreed that whisker growth occurs at the base of the whisker in response to imposed stresses or residual stresses below the surface. The formation of Cu_6Sn_5 or other intermetallic compounds at the interface between the tin and the substrate layer has been shown to result in a compressive stress in the Sn film [64,63]. Once the oxide layer covering the tin has ruptured, tin whiskers can be extruded as a means of releasing compressive stress. It has been demonstrated that the use of certain substrate-coating combinations, such as Ni over Cu, significantly reduces whisker growth [65]. It was also demonstrated that avoiding brighteners, annealing of any residual stresses, using thicker tin layers, and addition of Pb are beneficial in reducing the propensity for whisker growth. On the other hand, the use of brighteners, lack of annealing, tin layers thinner than 2 μm, copper-based substrates and addition of Zn were shown to promote tin whisker formation [3]. The study of Sn whisker related reliability issues in portable consumer electronic products is in its infancy insofar as published reports of whisker related failures. Owing to considerable variations in Sn plating formulations and test methodologies, estimation of product failure risk has not been easy. However, decreasing pitch and increasing circuit density coupled with the drive toward Sn-rich solder compositions can be expected to elevate the risk of failure due to Sn whisker related issues in the near future.

10.3. RELIABILITY COMPARISONS IN LITERATURE

Reliability testing and accompanying failure analysis that are needed to fully understand the magnitude and nature of reliability concerns can be expensive in terms of time and resources. As discussed earlier in this chapter, there is a constant business driven need to minimize or accelerate reliability tests. Therefore, it is only natural that every effort is made to utilize any available historical data to assess current reliability risks and minimize

the reliability testing that needs to be performed. While the value of reliability comparisons is clear in terms of reducing the need for testing and saving time and money, comparison and utilization of reliability data from different sources is a difficult exercise at best, and one has to be cognizant of the multitude of factors that influence the final reliability projections. In this section, some of the relevant aspects in comparing reliability results from different sources are discussed.

10.3.1. Thermomechanical Reliability

Effects of thermal fatigue are generally evaluated through accelerated thermal cycling tests. Test units, in statistically significant numbers, are subjected to a predetermined thermal profile over a number of cycles until all or 50% of the samples fail, and failure distributions are determined. In evaluating technologies, comparisons of failure data from a variety of sources are attempted to verify, substantiate or discern significant variations in reliability and understand the mechanisms. There are several pit falls in this approach. The first one is the definition of failure. Some regard a percent change in the resistance of total risk net consisting of a number of solder joints. Others may consider resistance spikes of a given magnitude and lasting over a specified duration, and still others may consider only an open joint as constituting a failure. The number of joints in a risk net may be different from study to study as well as in the same study depending on the I/O s of the packages being studied. The actual value of the resistance change can be significantly different in each case, if only percent change in resistance is considered. In great many instances, the failure criterion is not even included. A comparison of the probability plots can lead to misleading conclusions if the failure criteria are not identical in all of them.

Test parameters are also crucial and need to be considered explicitly for meaningful reliability comparisons. For example, in a thermal cycling test, the important parameters are the ramp rates and dwell times at the temperature extremities. A ramp rate of 15°C/minute and a dwell time of 10 minutes at each extremity are generally considered appropriate in many instances. However, literature contains data with 6 cycles per hour all the way up to 2 cycles per hour. Differences in the dwell time at extremities can have significant influence on the thermal fatigue and creep behavior of interconnection alloys. The temperature that the package and board experience in a given profile can be different from settings of the temperature chamber. Many studies only indicate the temperature values involved and do not provide the actual temperature the product under test sees. It is only prudent to compare temperature profile of the chamber versus the actual temperature experienced by the product under test as a function of time.

Other important factors that influence the discrepancies between the two are: the number of layers, copper and epoxy content, thickness of the board, its heat capacity, nature and size of the components, presence and absence of heat sinks. For example, a high I/O large ceramic component may take a longer time to attain steady state in comparison to a thin small package such as a chip scale package. If the cycle profile is not set correctly, it can alter the dwell time on some packages. Thus, a package of high heat capacity is more likely to experience a shorter dwell than a smaller package. The net result is that the solder joints in the bigger package may not experience the anticipated creep relaxation, and hence the failure may be altered by an unpredictable amount. In addition, during the ramp-up portion of the cycle, temperature can overshoot the preset values and it takes some time for the temperature to reach the set value. If a number of boards are being tested in the chamber the location of the boards in the chamber, and their disposition can

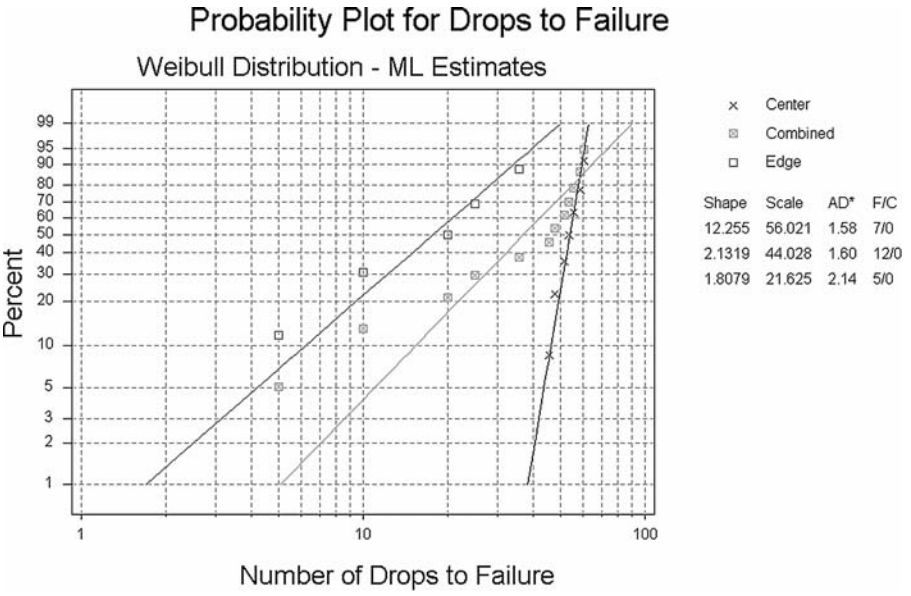


FIGURE 10.1. The presence of two different failure modes can be discerned in the data for the samples with no underfill in the drop test.

influence the temperature each board or package experiences. Boards stacked together and aligned perpendicular to the direction of air flow in the chamber will result in the boards immediately facing the air flow experiencing a different profile than other boards in the stack. In addition, the likelihood of blind spots in the chamber cannot be ruled out. Thus, a complete characterization of the thermal chamber to ensure that packages and the board attain the equilibrium temperature is very important.

Comparison of failure distributions can be complicated if the statistical distributions are not properly chosen and failure mechanisms are not well understood. The most popular solder joint failure distributions are the two-parameter Weibull distributions and occasionally three-parameter Weibull distributions. Even while using the two-parameter Weibull distributions, a single average line is often drawn through two apparently distinct distributions, as shown in Figure 10.1. This often leads to erroneous N_{50} values. In addition, a tacit assumption is made that there is only a single failure mechanism.

Sometimes, reliability results are reported without a failure analysis. Even when the failure mechanism is reported, the mechanism that is reported is based on the analysis done at the end of the test and not immediately following the detection of failure by electrical test. Thus, the understanding of the failure mechanism is corrupted or distorted by crack propagation, and micro structural changes occurred subsequent to the failure detection. When the distribution plots exhibit failures that indicate differing slopes, it is important to delineate them and conduct failure analysis to determine the exact failure mechanisms.

Thus, comparison of thermal cycling reliability tests has to be carried out with extreme care and caution taking into account all the factors that affect the inferences and conclusions. The current literature on reliability does not appear to readily lend itself to definitive correlations and comparisons.

10.3.2. Mechanical Reliability

Mechanical reliability comparisons for portable consumer electronics are more complicated and difficult than thermomechanical reliability comparisons because of dynamic and structural complexities. There are many more variables to be taken into account in the assessment of board-level mechanical reliability. These include package size, solder ball size, board structure and dimensions, drop height, orientation, impact duration, strike surface, etc. At the product-level, reliability comparisons are even more complicated due to additional dependence on the product form factor, weight distribution, impact orientation, occurrence of secondary impacts, and other test related variables. Therefore, the ability for comparison of mechanical drop test reliability is at its infancy.

Consistent test procedures with consistent acceleration and impact and failure criteria are critical in ensuring that results from one reliability test can be compared with results from another. For example, peak acceleration and the impact energy attained by the product depend on the frictional forces induced by the guide mechanism in the test equipment. Therefore, actual impact velocity can be different from the theoretically computed value.

The number of mounting screws and their location also has significant effect on the drop reliability. Boards mounted with only four screws can have lower impact life compared to those mounted with six screws under the same loading conditions due to greater bending. The type of screws and the torque applied to them can have a pronounced effect on the drop performance. The likelihood of screws loosening after subsequent drops cannot be ruled out. The dislodged screws can dramatically alter the board response during the drop. In addition, it needs to be verified that the failure locations and mechanisms are identical before attempting to compare reliability values. For example, failures that occur during drop can be due to interfacial brittle fracture at the package pad/solder interface, printed wiring board pad/solders interface, or the copper trace break at entry to the pad.

Location of the package on the PWB also plays an important role in determining reliability. Board bending and warpage can be very dependent on the board dimensions, and are usually greater along the longer dimension of the product. Typically, but not always, packages positioned at the center of the board are more susceptible to failure than the ones away from the center when the product is dropped on its face or back.

Package construction plays a significant role as well. Many portable electronics use low profile packages to accommodate the rather slim product form factors. These packages, such as ball grid array packages like Very-thin-profile Fine-pitch BGA (VFBGA), Thin-profile Fine-pitch BGA (TFBGA), and Quad Flat pack No-lead (QFN) packages, have low solder joint stand off, thinner die, and thinner molding compound. For example, VFBGAs have been shown to have slightly better performance than TFBGAs having the same I/Os [66]. In a different study, the 208 QFP package solder joints were observed to fail in a relatively small number of drops due to their mass and FLGA 300 (0.8 mm pitch) packages were relatively more durable [79].

Materials' aspects such as surface finish on the package and PWB pads can be expected to have a significant effect on the drop test reliability. Compatibility between PWB and component termination finishes, sometimes even inside the component module, can play a significant role in determining drop reliability. For example, incompatibility between Cu finish on resonators and ENIG finish on interposer PWB was found to severely degrade drop test performance [67]. In this case, the copper from solder/component interface migrated to the solder/interposer interface during the reflow and impeded the growth of Ni−Sn intermetallics, and instead, promoted the formation of a ternary Ni−Cu−Sn

intermetallic phase. In the absence of a strong metallurgical bond between the Ni on the interposer PWB and the solder, premature failures occurred in drop testing.

Although, in general, Sn−Cu interfacial bond has been found to be superior to the Sn−Ni interfacial bond, recent evidence seems to suggest that Cu−Sn intermetallic bond can have risks as well. For example, the Kirkendall type of voiding found at the Cu/Cu_3Sn interface, especially after thermal aging, has been shown to impair board level drop performance [34]. Modification of the IMC bond strength by addition of trace amounts of some elements also needs to considered when comparing reliability results from different studies. For example, addition of 0.3%In and 0.04%Ni to Sn-Ag-Cu solder was shown to improve drop test reliability by as much as 20% even after 150°C thermal aging in comparison to the Sn-Ag-Cu solder [68].

10.4. INFLUENCE OF MATERIAL PROPERTIES ON RELIABILITY

10.4.1. Printed Wiring Board

The proliferation of portable electronic appliances in the form of mobile phones, personal digital assistants, pagers, etc., has brought about a "density revolution" in the printed wiring board technologies. Ever-smaller board features have necessitated new approaches to design, materials, fabrication, assembly, and testing. The consumer demand is for faster, cheaper, lighter, and more reliable electronic hardware. Conventional multilayer boards with 150 μm lines and 150 μm spaces with 325 μm drilled through hole vias cannot always accommodate the wiring densities for fine pitch high I/O area array devices such as ball grid array and chip scale packages. Therefore, weight reduction and high density requirements have resulted in the need for high density interconnect (HDI) boards. For portable electronic hardware with high density, thinner boards with finer lines and spaces with very small vias were needed. Thus evolved a completely new printed wiring board industry of HDI micro-via board technology featuring extremely thin laminates, and multilayer microvias. Several techniques such as Surface Laminar Circuitry (SLC), laser drilled micro-via techniques, Any Layer Inner Via Hole (ALIVH) technology have evolved. Buried, blind, and through-hole vias were needed to accommodate the product functionalities. These features are significantly different from the conventional printed wiring board technologies, and are approaching those used in the semiconductor industry. A semiconductor technology attitude is being cultivated by the printed wiring board industry to meet the new challenges. At the same time, the reliability requirements for portable electronic hardware are often more stringent than the conventional hardware. The only relaxation in the reliability requirement is one of product life; they are shorter than those required for desktop and business products. However, the mechanical and environmental requirements are more severe.

The complexity of the product varies considerably and may contain PWB assemblies that are either single sided or double sided. A double-sided assembly will be more rigid and display a different shock response. In some cases, depending on the product complexity, both buried and blind vias may be used simultaneously. The buried vias may be plated or filled with conductive paste and cured. The reliability of thin populated boards with blind and buried vias is inadequately understood under various mechanical loading conditions. Issues, such as mis-registration of the buried vias in the individual layers, can pose a reliability exposure.

Another important aspect of micro-via technology is the shape of the vias namely, square-well or bathtub, and the copper plating thickness and uniformity due to the variations in via shape. In addition, the registration of the micro-via on the capture pad is very important and crucial for product reliability. In case of poor registration the laser drilling may be partially off the pad and penetrate the adjacent laminate. This can result in voiding during reflow process due to the egress of the occluded moisture in the laminate, impacting the package to board interconnection integrity.

In a high density printed wiring board, different materials are used for the micro-via layer including non-reinforced epoxies, woven-fiber reinforced resins, chopped fiber reinforced resins, such as aramid-reinforced materials, and resin coated copper foils. The adhesion of the reinforcing material to the base resin can have a significant impact on reliability. Additionally, several Cu to laminate adhesion enhancement treatments, including mechanical abrasion have been in vogue. Each of these aspects can impact the reliability, especially under mechanical loading.

10.4.2. Package

In portable electronic products, package size and style can influence product concepts, and vice-versa. Packages have to fit the form factor of the product, which is usually very thin. Double sided surface mount assemblies with low standoff low profile packages are the order of the day. This limits the feasible options to chip scale packages, VSSOP, TSOP, lead less packages, LGAs, Quad Flat No-leaded package types, to name a few. With increasingly effective utilization of PWB real estate, an emerging trend is to explore the out-of-plane dimension to increase the packaging density within the constraints of the form factor and package height limitations. Device stacking and package stacking are becoming increasingly popular. An understanding of the failure modes and mechanisms of these packages on a variety of laminate materials, and their construction under thermal and mechanical loading is still in its infancy. Package size, materials and construction, die size and thickness, the order of the stacking, and the bonding methods used can all have significant impact on the failure nature and mechanisms. Failures can range from package damage such as popcorning, to silicon die damage, interconnection failures, delamination, laminate cracking, etc. Industry trends indicate that with thinner die, such as 50–70 μm thin die, packages with as many as six to seven die stacked together could be anticipated in the near future.

10.4.3. Surface Finish

Surface finish of printed wiring boards and the package termination play a significant role in the integrity and reliability of an interconnection. Hot-air-solder-leveling which has been the main PWB surface finish for well over half a century has outlived its usefulness since the advent of high I/O fine pitch surface mount and area array packaging technology. Several surface finishes have since come into use. Organic solderability preservatives (OSP) and electroless nickel-immersion gold (ENIG) have almost replaced solder leveling. ENIG has been used extensively owing to its long shelf life and for excellent solderability wherever co planarity requirements are stringent. However, as hardware integration and miniaturization continued, resulting in smaller feature sizes, problems related to defects in ENIG surfaced. The hypercorrosivity of immersion gold plating composition and attendant high phosphorous content can cause sporadic and unpredictable solderability problems

(also referred to as black-pad). In addition, as portable electronic hardware is more subject to mechanical loading, intermetallic brittle fracture at the solder-pad interface is some times encountered. Also, as has been mentioned earlier, it is generally recognized that nickel tin intermetallics are more brittle than the copper-tin intermetallics. Often dual surface finishes are employed, with OSP to preserve solderability, and ENIG for electrical contact surfaces.

With ever increasing emphasis on the implementation of lead-free solders as the interconnection material, surface finishes of PWB, package leads, and terminations are being reexamined to arrive at acceptable alternatives. Immersion silver, immersion tin, palladium, nickel-palladium-gold etc., are being looked at. There does not appear to be a consensus on surface finishes. While each surface finish has its merits, the industry has to weigh the alternatives in terms of cost, performance and reliability for a given product group.

In the ensuing sections, several failure mechanisms pertaining to printed wiring boards, packages, and interconnections under a variety of loading conditions are described.

10.5. FAILURE MECHANISMS

As mentioned earlier, the failure mechanisms in handheld electronic products are different from those commonly encountered in desktop or mainframe business machine environments. Broadly, they may be categorized as those caused by (a) thermal loading, (b) mechanical loading (including mechanical drop, vibration, bending and twisting loads), and (c) electrochemical environments that induce corrosion and electromigration.

10.5.1. Thermal Environment

Failures induced due to thermal stresses in portable electronic hardware are in general similar to those in other electronic products. In portable electronic hardware, where use of HDI with multiple micro-via layers is prevalent, the shape of the micro-via, copper thickness, and the voids in the microvia influence the nature of the interconnect failure. In general, interconnect failures tend to occur on the package side of the solder joint, and are influenced by the coefficient of thermal expansion (CTE) of the package and sometimes aggravated by the solder mask defined pad geometry on the package side. In conventional Sn-Pb solders, the fracture generally occurs in the solder adjacent to the intermetallic layer, where the region is Pb rich in composition.

For Pb-free solder alloys, the interconnect failure mechanisms may display different kinds of deviations from the previously observed mechanisms for Pb-Sn alloys. Depending on the surface finish and the pad metallurgy, the interconnection can have multiple types of intermetallic phases dispersed in the bulk joint. In the case of tin-silver-copper system with OSP and Electroless Ni Immersion Au (ENIG) surface finish, Cu-Sn, Ag-Sn, Au-Sn intermetallics were found to be dispersed in the bulk of the joint or near the pads [69]. Solder joint failures due to thermal cycling are influenced by shear forces induced by to CTE mismatch between the component and PWB, with both fatigue and creep damage mechanisms operative at the same time. A damage accumulation map for Pb-free solders is discussed next.

In Pb-free solders, there are several significant differences in the microstructure compared to the Sn-Pb eutectic or near-eutectic solders, and these microstructural differences result in a very different damage evolution process. The primary microstructural differences are as follows:

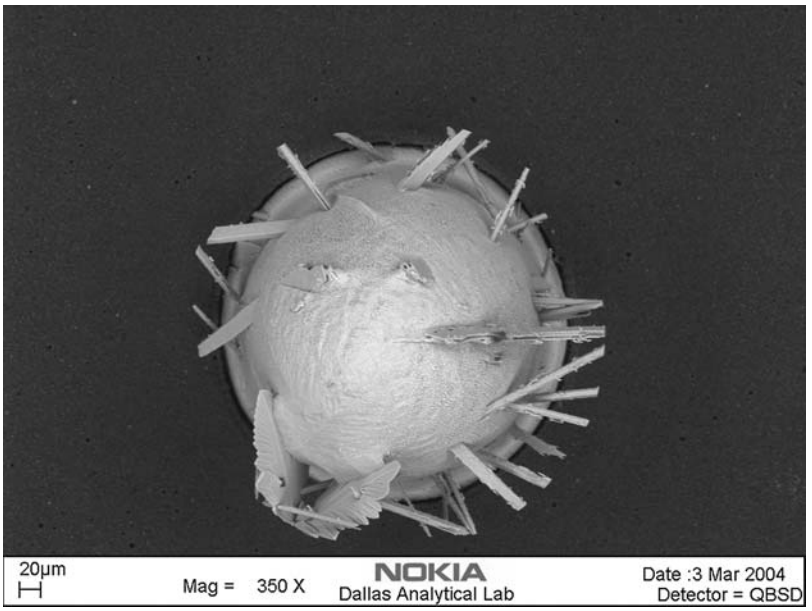


FIGURE 10.2. Partially etched solder ball of an unmounted revealing the IMC phase that can affect crack initiation and crack propagation thereby affecting thermomechanical reliability.

1. The intermetallic morphology is more complex and a multitude of small spheroidal Ag_3Sn IMCs are observed at the Sn dendrite boundaries. These could serve as initiation sites for voids and microcracks.
2. In addition to the increased presence of small particles in the interdendritic spacing, several large Cu_6Sn_5 IMCs and Ag_3Sn plates are distributed throughout the solder ball, which can effectively constrain the solder joint during shear deformation. A partially etched Sn3.5Ag0.7Cu solder ball on an unassembled CSP is shown in Figure 10.2 to illustrate how the IMCs in this solder system are distributed throughout the bulk of the solder joint to much larger extent than previously observed in Pb-Sn solders. A completely etched solder ball microstructure (in Figure 10.3) reveals the presence of Cu_6Sn_5 scallop shaped intermetallic phases adjacent to the Cu pad in addition to the IMCs distributed in the bulk of the solder.

These microstructural features can bring out damage mechanisms in Pb-free solders that were not a significant contributor to final failure in Pb-Sn solders under thermal fatigue/creep environments.

It should be noted that failure in thermal cycling in solders involves both fatigue and creep failure mechanisms. The relevant mechanisms of creep deformation are:

- Dislocation creep—involves the movement of dislocations which overcome barriers by thermally assisted mechanisms involving the diffusion of vacancies or interstitials ($10^{-4} < \sigma/G < 10^{-2}$).
- Diffusion creep—involves the flow of vacancies and interstitials under the influence of applied stress ($\sigma/G < 10^{-4}$).
- Grain boundary sliding—involves the sliding of grains past each other.
- Dislocation glide, which normally requires very high stresses, is probably not a major contributor to creep during thermal fatigue. Diffusion creep causes vacancies

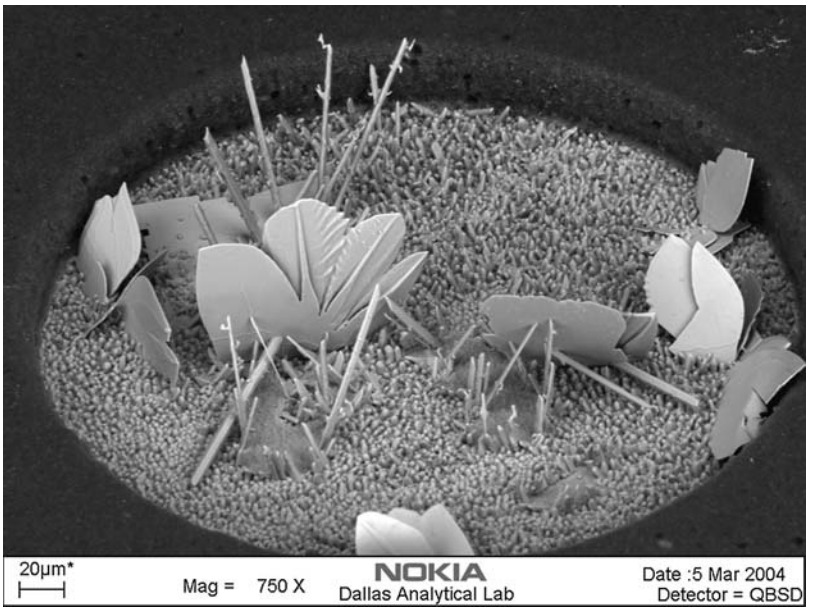


FIGURE 10.3. Completely etched solder ball of an unmounted revealing the scallop shaped Cu_6Sn_5 IMC phases on the Cu pad along with rod shaped Cu_6Sn_5 IMCs. In addition, planar dendrites of Ag_3Sn can also be seen interspersed throughout the surface.

from grain boundaries experiencing tensile stresses to flow toward those that are experiencing compressive stresses.

In a solder joint with the microstructure and IMC morphology described in the earlier section, the driving force for failure is the imposed cyclic shear stress due to CTE mismatch and creep under this stress. Because of the low homologous temperature of solder, fatigue damage mechanisms are accompanied by creep damage mechanisms.

Consistent with previously reported damage mechanisms for Pb-Sn solder, inhomogeneous shear stress fields can result in recrystallization at pads, corners, and voids. Additional damage mechanisms not widely reported for Sn-Pb solder but observed for Pb-free solders by Dunford et al. [69] are described next.

Zones of recrystallized material were observed at locations with high strain gradients and strain incompatibilities, such as grain boundaries. These recrystallized zones grow with imposed cycling, and a multitude of smaller recrystallized grains form to relieve the strain. In parallel, creep driven damage mechanisms were observed to a degree not reported in previous studies. Another creep driven damage mechanism is the initiation of voids and cracks at locations of high strain incompatibility. For example, triple-point grain junctions and IMC-grain boundary junctions in the interior of solder balls and grain boundary (GB) junctions at the surface of the solder ball appeared to be the favored sites for crack initiation.

Further damage evolution is governed by the interaction of the localized damage (in the form of recrystallized zone) with the distributed damage (in the form of microcracks and voids). The severity of damage of all three types, namely (a) recrystallization zones, (b) microcracks and voids at recrystallized grain boundaries (RGB), and (c) cracks and voids at GB, grows with increased cycling.

Final failure, however, is dominated by the weakening of the material due to recrystallization and distributed microcracking in the damage zone. A macrocrack forms by

the coalescence of the microcracks, primarily in the recrystallized zones. The propagation path of these macrocracks is very different from that observed for Pb-Sn solders. The IMC plates and rods sometimes serve to deflect the propagating macrocrack so that several macrocracks may exist in a solder joint without significantly impacting electrical continuity. These macrocracks coalesce with each other through the distributed damage, changing direction depending on local damage geometries and microcracks at the RGB or the cracks at the GB. Final failure occurs by propagation of the most dominant macrocrack traversing the solder ball, primarily near the pads on the board or the component. For example, in the solder joint of the CSP shown in Figure 10.4, one can see the tortuous path taken by the propagating macrocrack and the distribution of the microcracks near the fracture plane. Near the bottom of the solder joint, away from the component pad, an elongated void formed due to of creep related damage enlarging an initially small crack or void, is also seen. In the right half of the picture, the grain morphology with Ag_3Sn and Cu_6Sn_5 IMC particles interspersed in the interdendritic spaces is seen. A higher magnification picture of an elongated void caused by creep damage at grain boundaries in a different solder joint is shown in Figure 10.5. The damage evolution map for thermomechanical loading that brings together the different operative mechanisms just described is shown in Figure 10.6 [69].

10.5.2. Mechanical Environment

It is instructive to review the construction of a generic package mounted on a PWB before discussing the failure mechanisms. The PWB in portable electronic products serves not only as a carrier for the different electrical subsystems but also provides mechanical rigidity to the assembly. A typical PWB can have 4 to 12 electrical planes laminated between woven glass fiber reinforced epoxy layers that serve both a dielectric and mechanical support function. Electrical connection between these layers is often achieved through plated-through-hole vias, blind vias or buried vias. The outermost layer of the PWB, sometimes called the build-up layer, is the first interconnection layer between the solder joint and the PWB. Interconnection failures can be found at different levels as shown schematically in Figure 10.7, and can be classified as follows, based on the location of the crack:

- Die fracture within the package.
- Interposer level failure within the package.
- Solder joint fracture.
- Crack initiation inside the component and subsequent damage to the solder joint.
- Solder joint fracture.
- Interfacial failure—at the solder/PWB pad interface.
- PWB related failure—trace fracture.
- PWB related failure—micro-via fracture.

10.5.2.1. Die Fracture within the Package Sometimes, when the die is not supported optimally inside the package, almost all the flexure of the PWB can be transmitted to the relatively brittle semiconductor die inside the package. Cleavage fracture of the die can occur causing electrical failure. An example of this kind of failure is shown in the optical micrograph in Figure 10.8. The wire bonds on the die can also be seen along with the vertical crack in the die. The cracks at the inactive side of the die (bottom) are attributed to polishing damage during the grinding stage. Such artifacts have previously been observed in samples where excessive normal force was exerted on the sample, and should not be confused with cleavage type of cracking on the active side of the die.

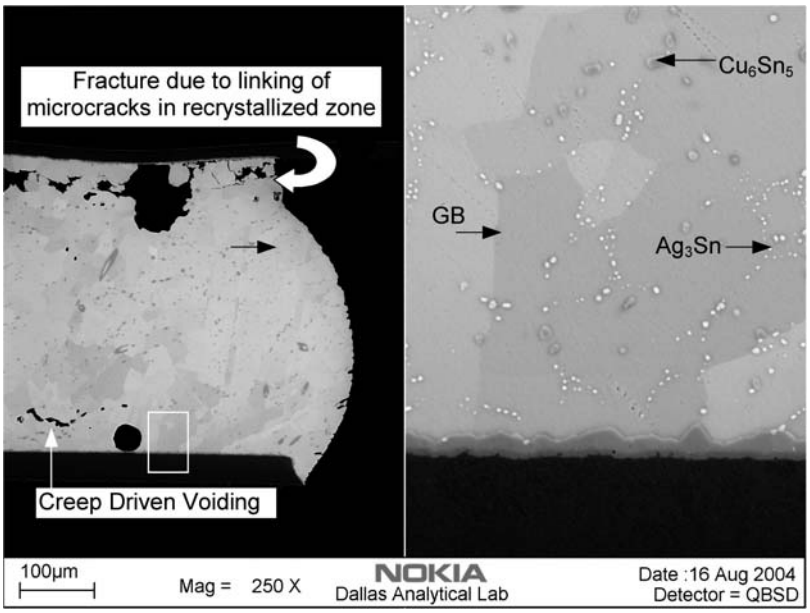


FIGURE 10.4. Interconnection fracture due to thermomechanical fatigue loading in Sn3.5Ag0.7Cu solder joint of CSP. The backscattered electron micrograph reveals the fatal crack near the component pad in addition to voiding and other damage near the PWB pad (Sample is a courtesy Michael Wellborn).

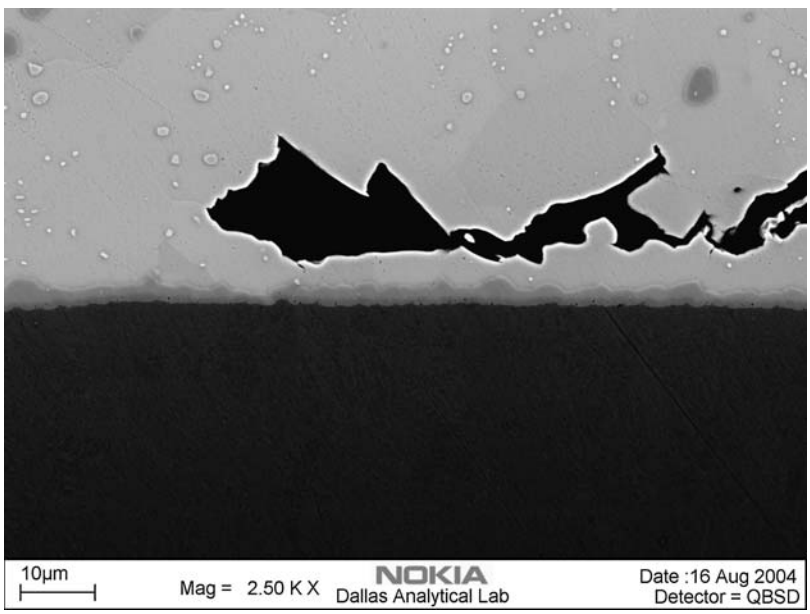


FIGURE 10.5. Creep driven damage resulting in elongated voids at grain boundaries.

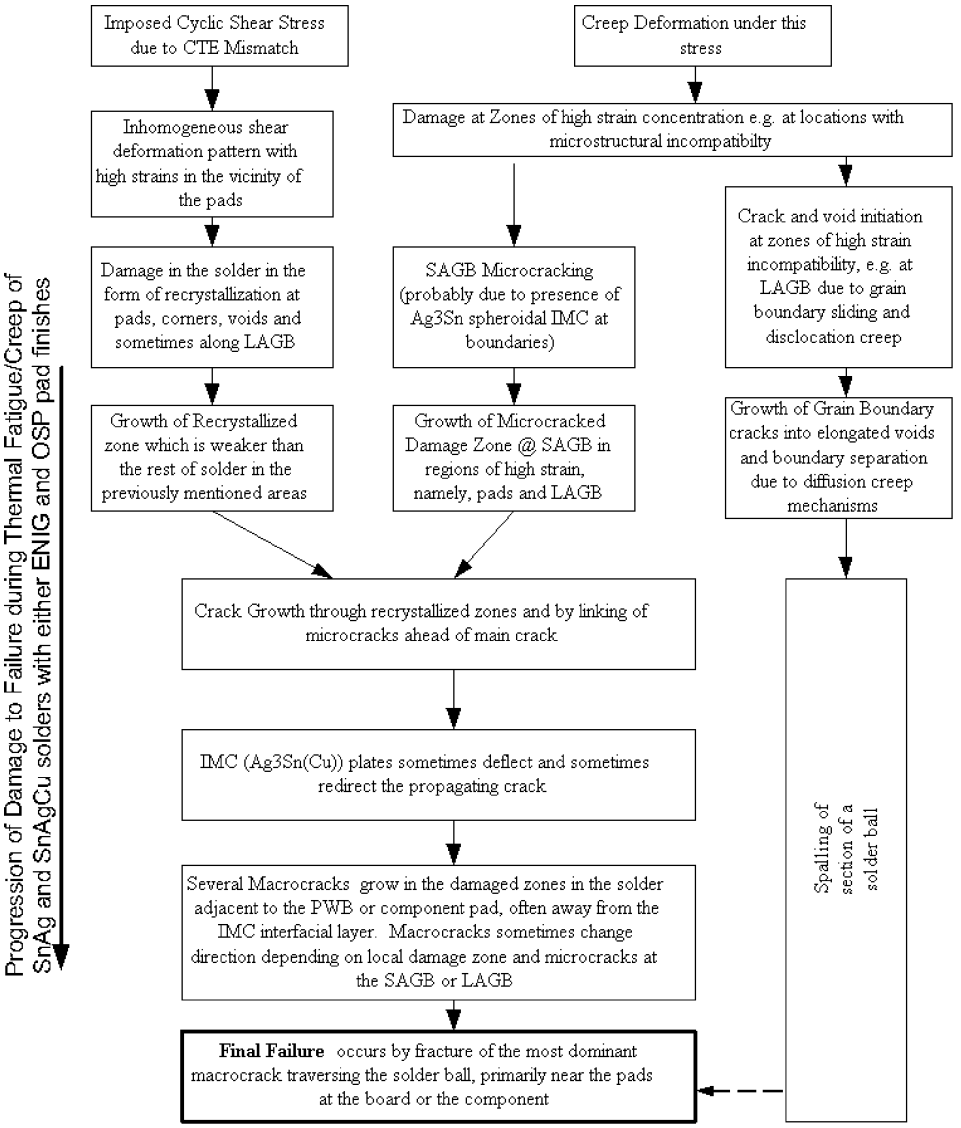


FIGURE 10.6. Damage mechanism evolution map for Pb3.5Sn0.7Cu solder under thermomechanical loading.

10.5.2.2. Interposer Level Package Failure The Cu circuitry inside the interposer can sometimes fail if the process conditions in the fabrication of the interposer are not optimal. The example shown in Figure 10.9 illustrates the particular case where sub-optimal adhesion between the via-barrel and the via-cap failed upon exposure to mechanical loading at the PWB level.

10.5.2.3. Crack Initiation Inside Component Leading to Solder Joint Damage Ceramic components, due to their weight and lower fracture toughness, are particularly susceptible to failure when the product is dropped. Local stress concentrations on the ceramic component, such as those created by machining can serve as crack initiation sites and cause

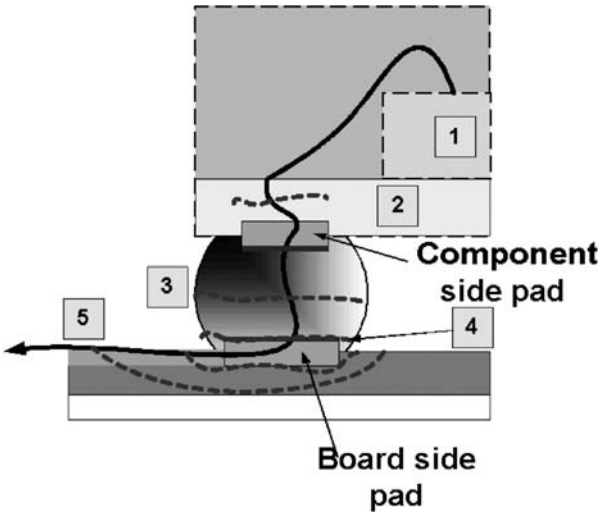


FIGURE 10.7. Simplified schematic of electrical interconnection from the Si die to multilayer PWB through different levels of packaging. Dashed line represents possible crack or open.



FIGURE 10.8. Die cracking due to mechanical loading.

premature failure as shown in Figure 10.10. The crack that originated at the machining groove caused an electrical open upon propagation. Apart from the machining on the ceramic component, a second factor contributing to the crack originating in the component is the relatively high strain rate of deformation during drop loading. Since solder deformation characteristics are highly strain rate dependent at room temperature, the solder joint is stiffer and stronger under higher deformation rates, thereby subjecting the ceramic compo-

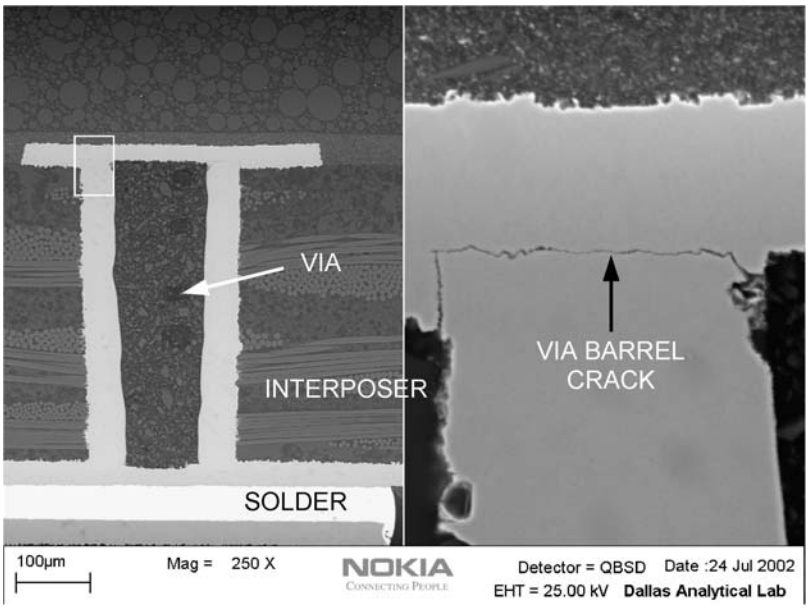


FIGURE 10.9. Via barrel cracking due to PWB level mechanical loading causing electrical failure.

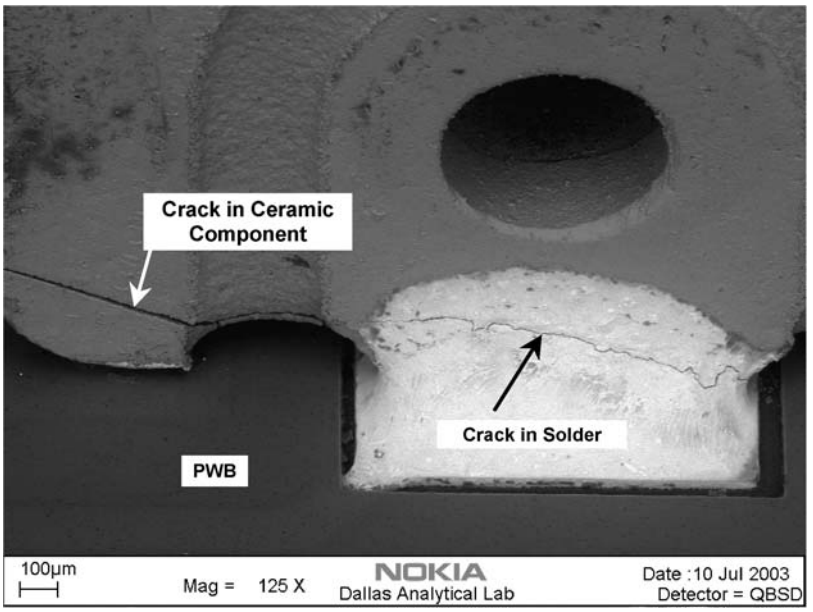


FIGURE 10.10. Crack in solder joint and ceramic component after mechanical shock (drop) reliability testing.

nent to proportionately higher stresses. For components operating at radio frequencies, a relatively minor partial crack, as shown in Figure 10.10, can sometimes cause parametric shift induced failures rather than a hard open.

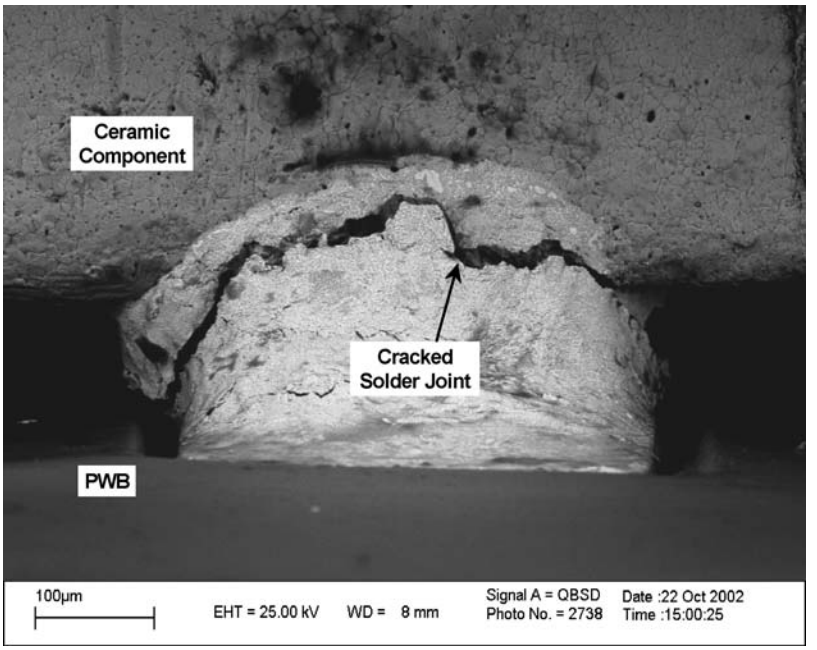


FIGURE 10.11. Crack in solder joint after twist testing.

10.5.2.4. Solder Joint Fracture due to PWB Level Twisting Bending and twisting are commonly encountered end-use environmental hazards for hand-held products. The deformation rates are much lower than those observed in mechanical drop. In such cases, the solder joint strength and stiffness are proportionately lower and promote fracture at the solder joint in contrast to locations within the ceramic component. An illustrative example is shown in Figure 10.11, where the solder joint is completely fractured without the damage extending into the ceramic. The lack of machining damage near the solder joint was probably a secondary factor in limiting damage to the solder joint without cracking the component.

10.5.2.5. Solder Joint Failure Related to Underfill Process It is a relatively common practice to provide additional reinforcement to a solder joint to improve its reliability under thermal and mechanical loads. For ball grid array (BGA) and chip scale packages (CSP) soldered onto PWBs, this reinforcement can be achieved by the use of a suitable underfill material in the package-to-board interspaces. This constrains the assembly against bending and thermal strains. One of the more commonly used procedures for underfilling a CSP soldered onto a board consists of dispensing liquid underfill along one or more edges of the CSP perimeter such that capillary action forces the underfill to fill the entire space between the CSP interposer and the PWB. Upon curing, the liquid underfill hardens and encapsulates the solder joints completely, thereby providing additional reliability by mitigating the deleterious effect of either thermal or mechanical strains.

The quality of the underfilling process is dependent on several variables such as temperature of the PWB or liquid underfill, cleanliness of the surfaces, speed of dispensing, etc. It has been shown that when the quality of the underfill is non-optimal and voids are present at the CSP corners, the benefit of the underfill is not realized even if the size of the void exposes only the corner solder joint [15]. An example of a partially underfilled CSP

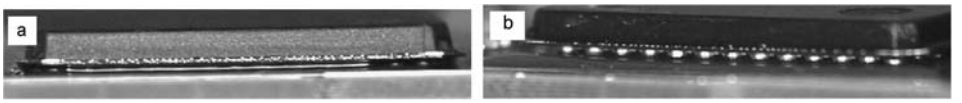


FIGURE 10.12. (a) A partially underfilled CSP with a corner underfill void, and (b) A more severe underfill defect exposing a whole row of solder joints.

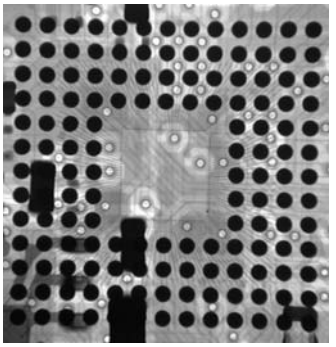


FIGURE 10.13. X-ray microscope image of a poorly underfilled CSP incorrectly indicating the lack of underfill defects. X-ray techniques can yield misleading results for certain kinds of defects.

is shown in Figure 10.12(a) and an optical micrograph of a more severe underfill defect is shown in Figure 10.12(b).

The true extent of an underfill defect cannot be ascertained by either visual or X-ray inspection. For example, Figure 10.13 shows a representative X-ray microscope picture of a CSP that does not reveal any underfill defect although visual inspection showed a substantial underfill defect at the perimeter. The scanning acoustic microscope, on the other hand, is very sensitive to voids and underfill defects. Difficulties encountered in acoustic inspection of CSP or BGA underfill include the signal to noise ratio due to material attenuation and uncertainty about the specific depth that the data includes. Both these problems are particularly severe for CSP and BGA underfill, unlike in flip chip underfill inspection. A judicious selection of transducer frequency, F# (ratio of focal length to diameter of the transducer), depth of focus and gating are essential for successful inspection. The acoustic image of the CSP in Figure 10.12(b) is shown in Figure 10.14, and the areas of incomplete underfill can be clearly identified in the top half of the acoustic image. A virtual cross-section along the dashed line is shown in the bottom half of the acoustic image, and the relative depths of the die, the interposer and the void can be seen. In addition, the bond wires extending from the die to the interposer are also visible. It is also useful to present the acoustic waveform along with the image to clarify the nature and location of defects.

An acoustic image of a different, improperly underfilled CSP is shown in Figure 10.15. The waveforms from three locations are presented alongside the acoustic image for ease of interpretation. The waveform from location 1 and 2 shows how the die and the interposer lie above the depth of the defect shown in location 3. The positive (upward) reflection from the top of the die and the Cu pads on the interposer are in contrast to the negative (downward) pulse from the underfill void. Thus, there is no ambiguity in concluding that the void lies below the interposer, where underfill would normally be expected in an underfilled sample. The lack of support for the solder ball can lead to failure of the interconnect that are now exposed to higher levels of loading. When exposed to ad-

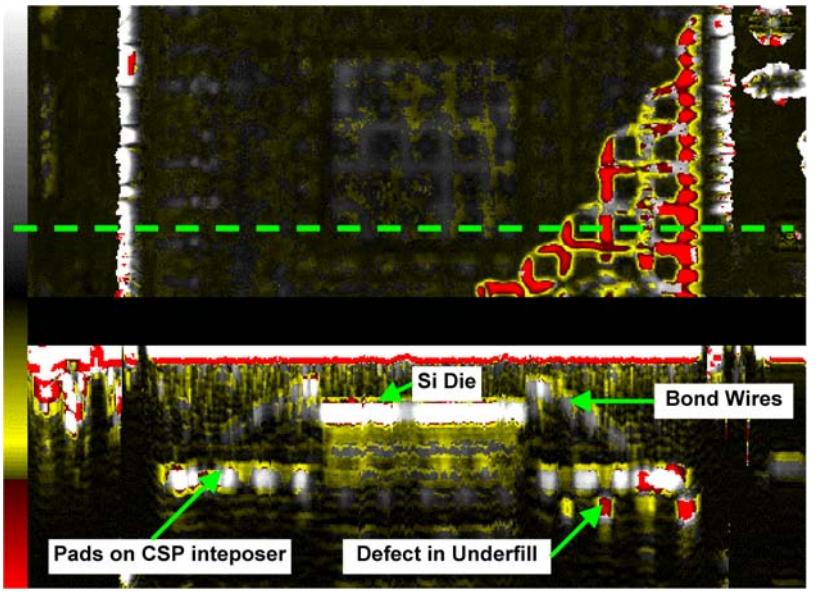


FIGURE 10.14. Acoustic image of the same CSP as in previous figure showing voiding in the underfill below the interposer of the CSP. A virtual cross-section (QBAM along the dashed line in the image) in the lower half of the image reveals that the underfill defect is below the interposer.

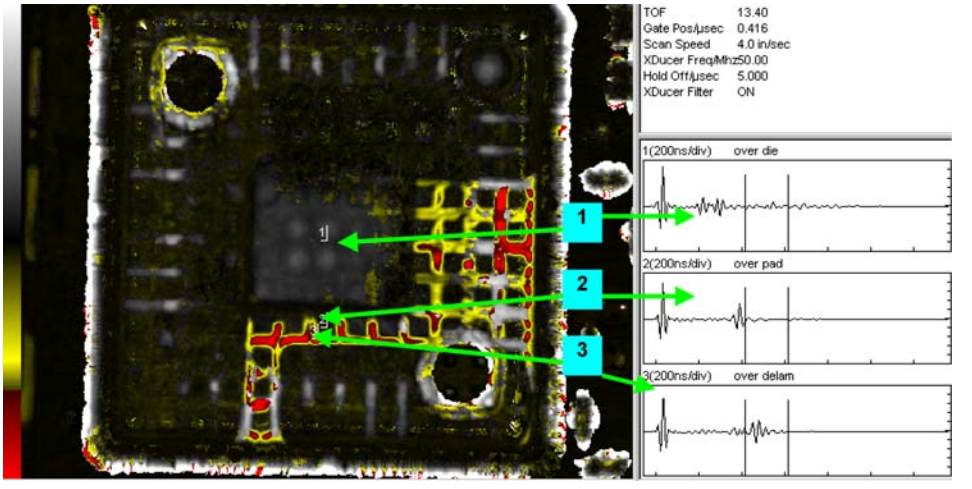


FIGURE 10.15. A more detailed acoustic image of a CSP with underfill defect showing the acoustic waveform traces over three locations: (1) the die, (2) the Cu pad on the interposer and (3) over the delamination.

verse environment such as mechanical loading, the solder joints or the build-up dielectric layer below the Cu pad on the PWB can develop cracks as shown in the scanning electron micrograph of the polished cross-section in Figure 10.16.

10.5.2.6. PWB Quality Related Fracture at Solder/PWB Pad Interface Electroless nickel/immersion gold (ENIG) plating of the Cu pads on the PWB gained considerable popularity as a pad surface finish in recent years. This is because it provides a cost ef-

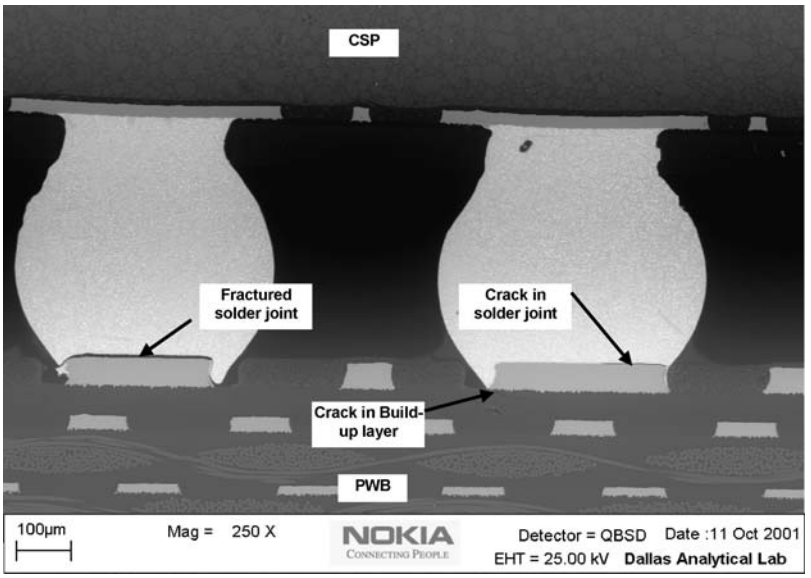


FIGURE 10.16. Scanning electron micrograph showing the fractured solder joint and concurrent damage at a neighboring solder joint.

fective means of ensuring coplanarity, which is a crucial requirement in high density, fine pitch assembly. The Ni layer was intended to provide a diffusion barrier between the gold and the Cu pad. The very thin gold layer (<0.5 µm) was intended to protect the Ni surface from oxidation and preserve its solderability until reflow. During reflow, when the solder melts and wets the gold surface, the gold layer dissolves instantly into the solder leaving a clean, solderable Ni surface for Ni-Sn metallurgical bond formation. Not so long ago, the interconnection pad sizes were relatively large because fine pitch packages were not widely used. When pad sizes were relatively large, quality variations in the ENIG plating did not immediately or always result in interconnection failures because the pad size-defect size ratio was substantial. Now, the pad size-defect ratio is smaller due to higher density of packaging. In addition to this, the increased use of less-aggressive organic solvent or water-soluble fluxes or no-clean fluxes can result in a pad surface that may not be as solderable. These trends in the industry have increased the risks due to ENIG surface finish related failures that are characterized by a brittle fracture at the solder/pad interface along with a dull, dark Ni surface exhibiting "mud crack" type of surface morphology [70,71].

One example of the brittle interfacial crack at the solder-PWB pad interface is shown in Figure 10.17. The fracture occurs below the intermetallic layer at the solder/ENIG pad, which indicates that the interfacial bond between the Ni-Sn intermetallic layer and the underlying Ni layer is weak. Indeed, the fracture surface on the PWB side pad is often devoid of any adhering solder as seen in the backscattered micrograph in Figure 10.18. The only evidence of solder adhering to the surface is seen in locations where a void in the solder offered easier crack propagation than fracture at the solder/Ni pad interface. The characteristic "mud cracking" types of features are also visible on the Ni fracture surface. A high magnification micrograph of the polished cross-section (Figure 10.19) reveals the high-P Ni layer and a transverse view of the hypercorrosion trenches in the Ni layer into which the solder has ingressed.

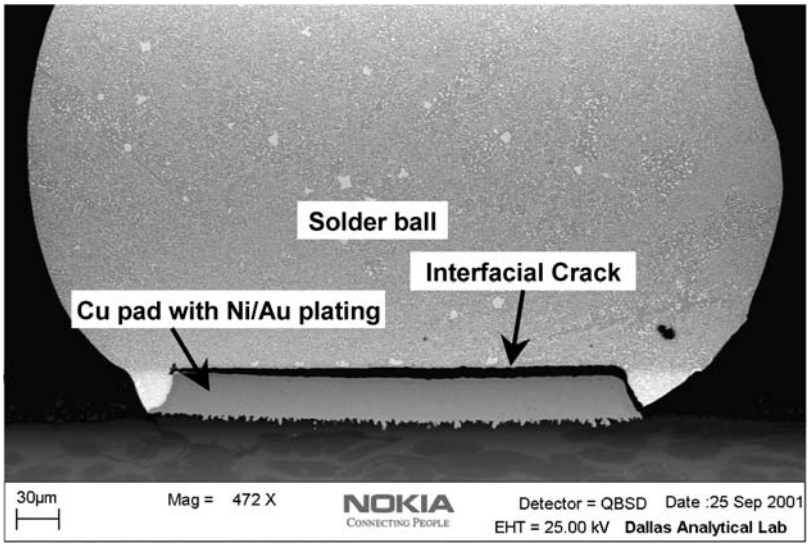


FIGURE 10.17. Interfacial fracture resembling brittle cleavage between solder ball and pad.

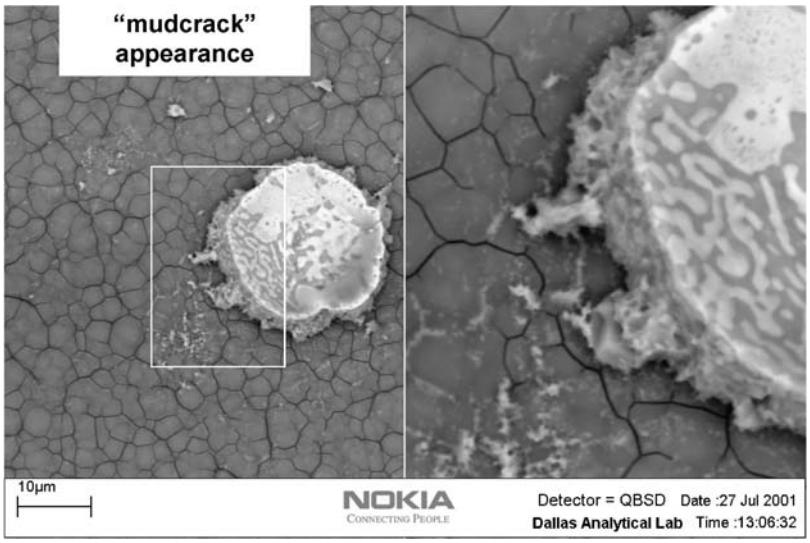


FIGURE 10.18. "Mud crack" appearance of Ni fracture surface showing the poor bond quality of solder to Ni/Cu.

10.5.2.7. PWB Build-up Layer Cracking Leading to Trace Fracture The PWB is usually a multi-layer laminate made up of layers of continuous, woven glass reinforced epoxy and Cu circuitry. The outermost layers sometimes referred to as the build-up or redistribution layers, serve both as the dielectric material and mechanical support for the Cu traces during PWB flexure or extension. Upon subjecting the assembled board to mechanical loading, such as encountered in mechanical drop, damage accumulates in the build-up layer in the form of cracking. Subsequent damage accumulation and electrical failure will depend on the redistribution method employed.

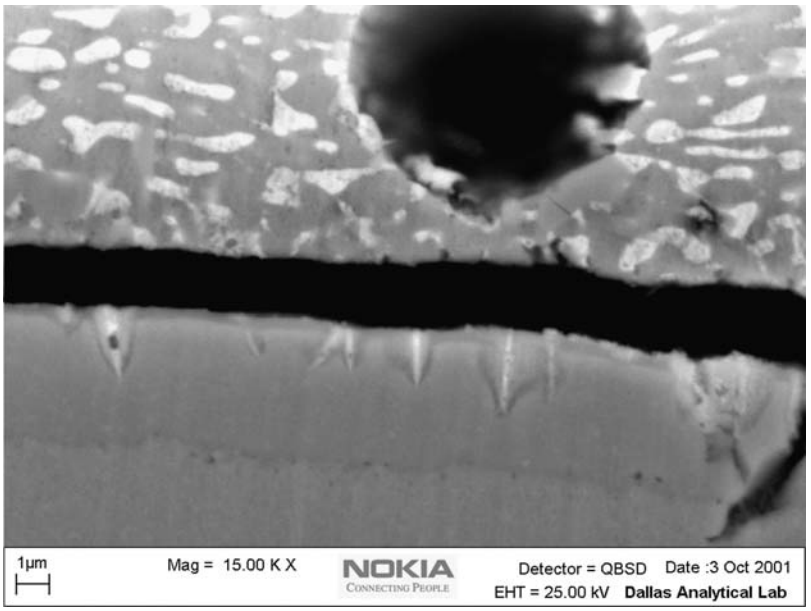


FIGURE 10.19. Hypercorrosion of Ni layer observed on a microsectioned sample with black pad defect.

The progression of damage for the case when redistribution is achieved through traces is described below. Initially, the damage in the build-up layer accumulates until the trace is no longer supported because of extensive cracking of the laminate under the Cu pad. The damage in the build-up layer is exemplified by the back-scattered electron micrograph in Figure 10.20(a). This weakening of the build-up layer forces the trace to shoulder an ever-increasing share of the mechanical loads imposed on the PWB, which eventually causes trace fracture by fatigue processes. An example of the fractured trace is shown in the top-view optical micrograph in Figure 10.20(b). The fracture process can be seen more clearly after a second microsectioning operation along the dotted line in Figure 10.20(b). The backscattered electron micrograph of the double-polished sample is shown in Figure 10.21(a) and a schematic explaining the crack under the pad causing the trace fracture is depicted in Figure 10.21(b).

The damage progression is similar in cases where a via-in-pad redistribution method is employed. The damage in the build-up layer, again, accumulates in the form of cracks. Once the support afforded by the build-up layer is diminished by the cracking, further mechanical flexure of the PWB subjects the via to increasingly higher stresses. Eventually, the via fractures due to fatigue, leading to an electrical open (Figure 10.22).

10.5.3. Electrochemical Environment

For portable and handheld electronic devices, two failure mechanisms related to electrochemical environments are of particular relevance—corrosion and electrochemical migration.

10.5.3.1. Corrosion Gold plating of connectors is a common practice designed to protect the underlying Cu and Ni layers from corrosive attack and promote good electrical contact. However, under the action of friction, the relatively thin and inert Au coating can be

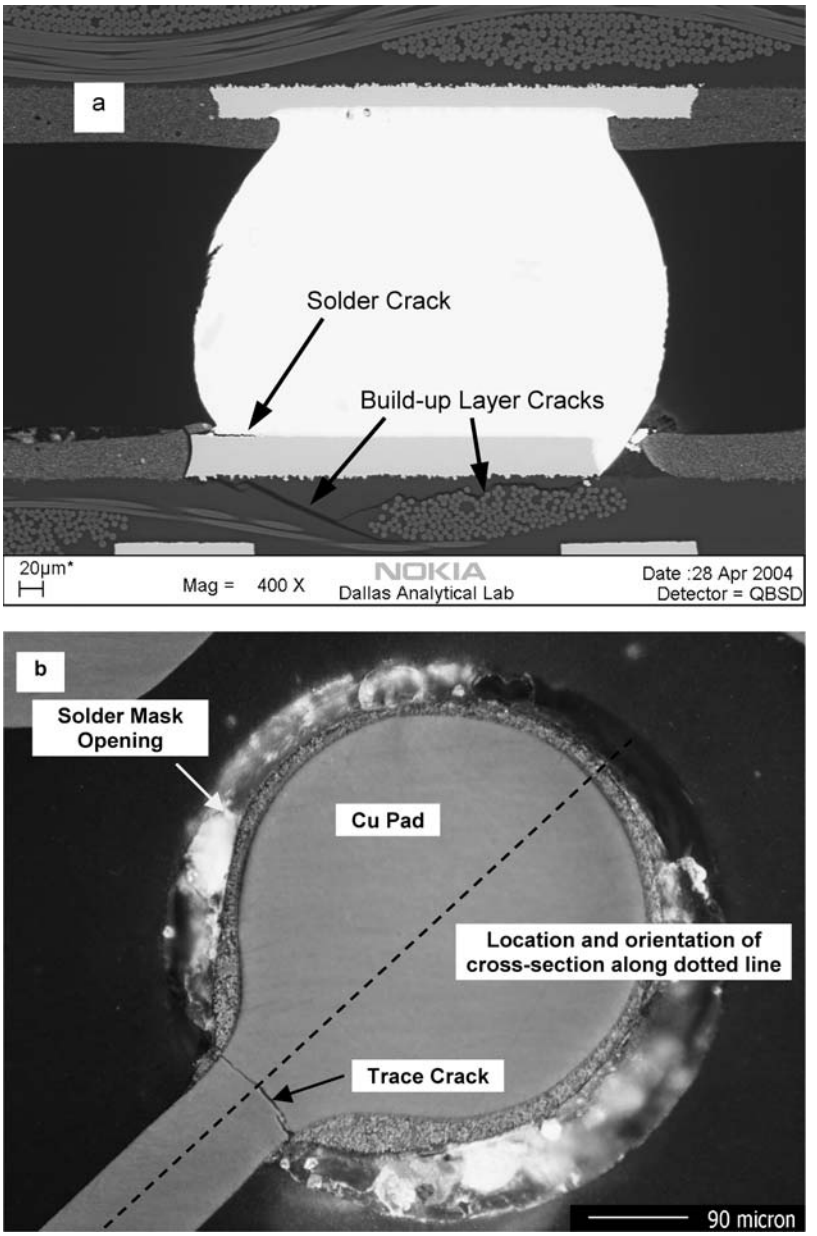


FIGURE 10.20. (a) Build-up layer cracking in a solder joint with trace and (b) Optical micrograph (top view) of a sample suspected to have a broken trace after the solder ball was removed by mechanical polishing. The dotted line represents the location and orientation of a second vertical microsectioning needed to show damage under the pad.

removed locally thereby exposing the Cu and Ni layers underneath. In such cases, fretting corrosion, pitting corrosion and localized galvanic corrosion can occur simultaneously, especially in the presence of an ionic species such as chlorides. This corrosion product, which is usually nonconductive, can cause electrical failure due to opens or intermittent. An ex-

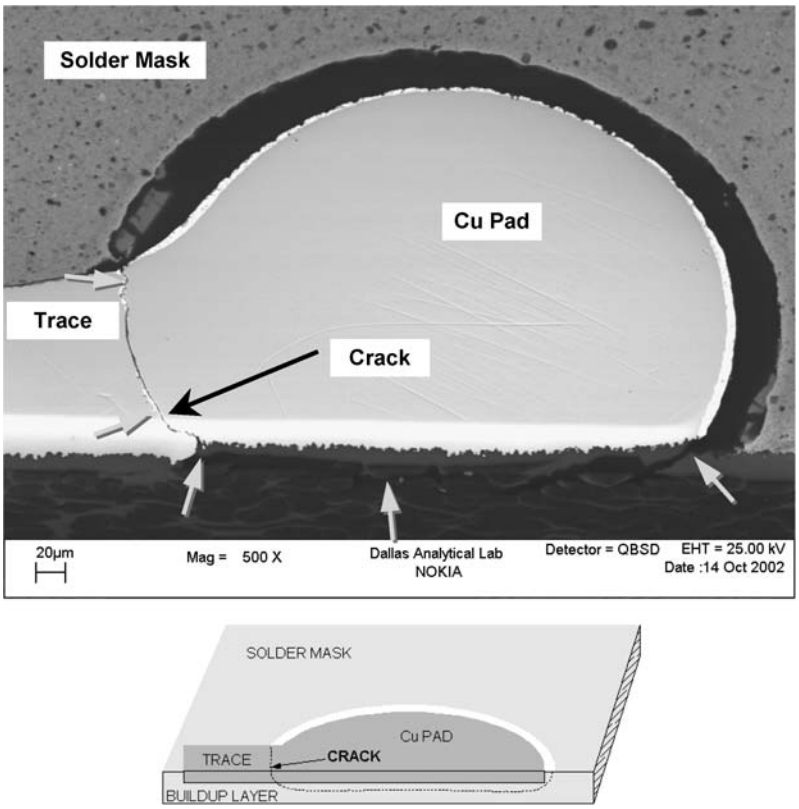


FIGURE 10.21. (a) Trace fracture accompanied by build-up layer cracking revealed in a double-cross sectioned sample (sample shown different than that depicted in Figure 10.18), (b) schematic showing the location of the crack in the build-up layer.

ample of gold plated connector corrosion is shown in Figure 10.23. The EDX elemental map for Au indicates that the coating is intact over the major portion of the area of interest. However, in the central portion of the image, the Au coating appears to have been removed completely, and the underlying Cu is exposed. This Cu surface, identified as a bright area in the Cu elemental map, also shows significant presence of O and Cl. The absence of any areas with high concentrations of Ni indicates that the mating surface of the connector has probably worn through the Ni layer in the area of contact.

10.5.3.2. Electrochemical Migration In several studies comparing the propensity for ECM of different metalization systems can be ranked as follows: Ag > Pb > Cu > Sn [54]. Although electrochemical migration phenomena have been observed with many metals, only Ag [49,51], and Cu to a limited extent [72], and perhaps Sn [73], have been found to exhibit this behavior in the presence of humid but non-condensing conditions. Indeed, Dumoulin et al. [73] concluded that silver migration presents the greatest risk because dendritic growth can occur whether Ag is outside the package or only partly exposed to humid air, on ceramic as well as on plastic substrates. Although Dumoulin et al. [73] suggested that Cu migration and Sn migration did not pose as big a risk based on their experimental data, in mobile electronic products which see a wide range of corrosive species during their lifetime, ECM of Cu and Sn can be as prevalent as Ag migration. In addition, residues

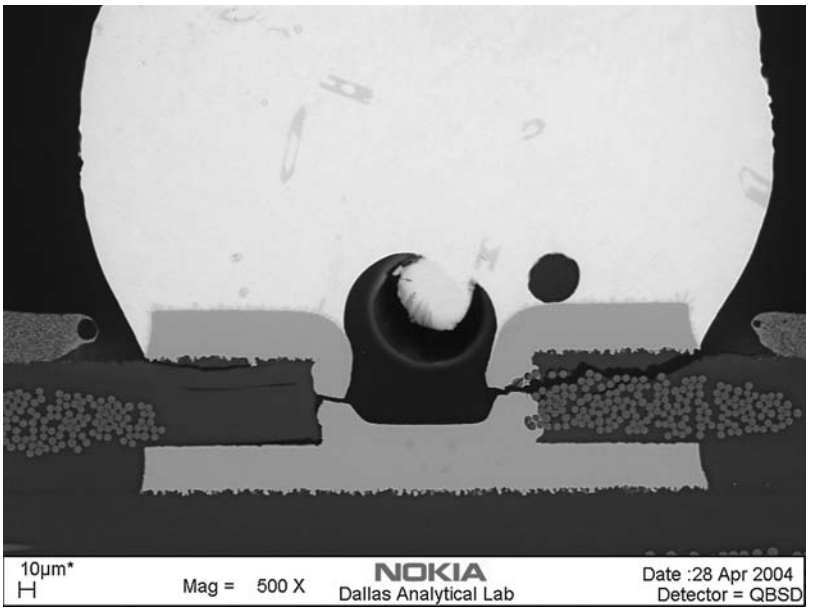


FIGURE 10.22. Build-up layer cracking in a solder joint with via-in-pad leading to via cracking upon further exposure to mechanical drop related stresses.

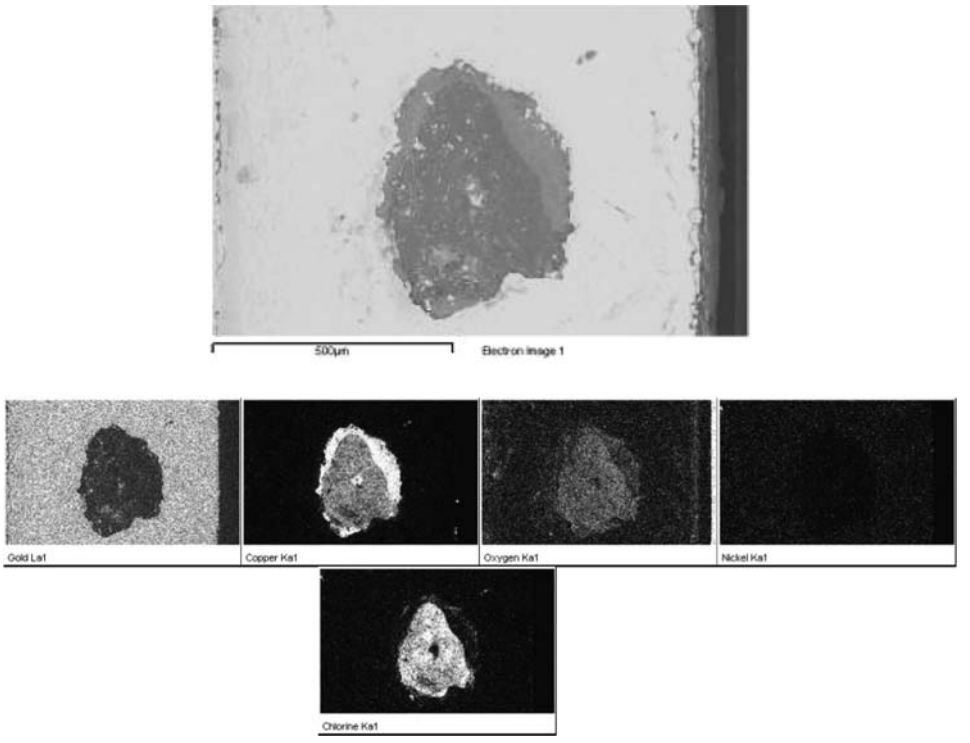


FIGURE 10.23. Corrosion of Au plated connector along with EDX elemental maps of Au, Cu, O, Ni, and Cl.

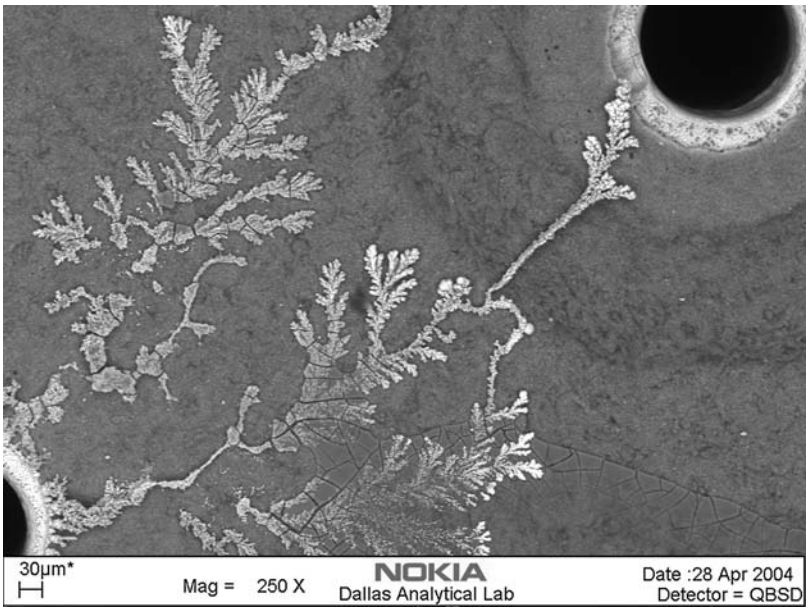


FIGURE 10.24. Cu electrochemical migration on a PWB between two through hole vias after damp heat exposure.

on the substrates that originate from the process, play an important role through water adsorption/conductivity behavior modification. One example each is provided next for ECM phenomena involving Cu, Sn and Ag.

Krumbein [74] noted that in practice, ECM can manifest itself as two separate, though not always distinct, effects that lead to impairment of the circuit's electrical integrity. Dendritic or filamentary bridging between the anode and cathode, which is one kind, has been discussed at length before. Colloidal staining is the second manifestation of ECM, which can also cause a short. Deposits of colloidal Ag, Cu or Sn have been observed to originate at the anode without necessarily remaining in contact with it. An example of this effect is also provided below.

Copper forms complex species such as $CuCl_4^{2-}$, $CuCl_2(H_2O)$, $Cu(H_2O)^{2+}$, etc. in the presence of halide containing species and moisture. An example of Cu electrochemical migration resulting in Cu dendrite formation is shown in Figure 10.24. If plated through hole vias or conductor pads are too close, Cu ECM can occur when the product is exposed to humid environments in the presence of an ionic contaminant.

Tin electrochemical migration mechanism is similar to that of Cu, but is much more prevalent because Sn constitutes a major portion of several commercial solder compositions such as 62SnPb2Ag, 10SnPb, Sn3.5Ag0.7Cu, etc. In addition, exposed Sn is more widespread on an assembled PWB as compared to Cu. The particular example shown in Figure 10.25 is from a test vehicle that failed upon exposure to damp heat testing. In this case, the potential difference between the terminals of a capacitor with Sn termination resulted in the migration of Sn from the anode toward the cathode. The right half of the picture shows a higher magnification view of the Sn dendrites at the cathode end of the termination. Elemental analysis mapping data of the surface of the capacitor is shown in Figure 10.26, where the Ba, Ti and O from the capacitor dielectric material can be seen clearly. In addition, the Sn map shows the presence of Sn between the terminations, where

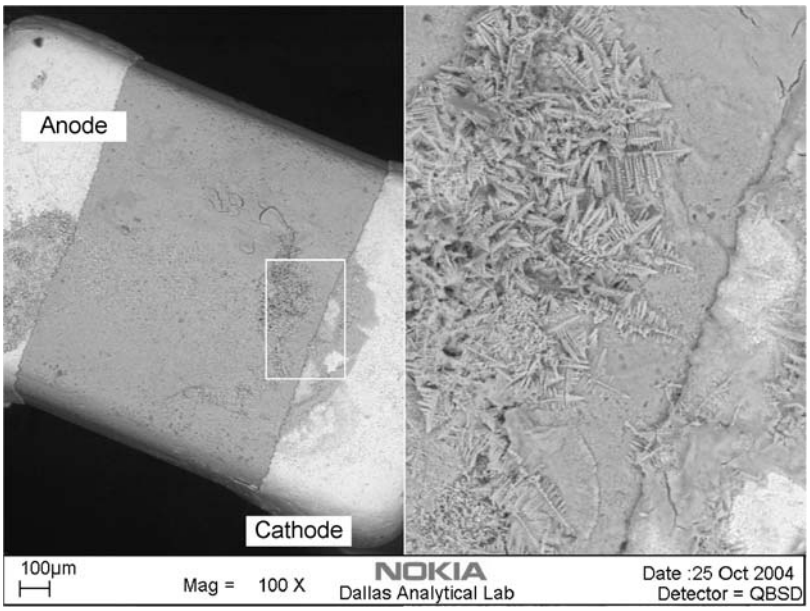


FIGURE 10.25. Tin electrochemical migration on a capacitor with tin termination. The right half of the picture shows a higher magnification view of the Sn dendrites at the cathode end of the termination.

there should be none. In several passives, Ni is used as a barrier layer between the silver adjacent to the dielectric and the tin termination. In this particular case, the Ni barrier layer at the anode is visible in areas where the Sn from the surface has been consumed by the ECM process. Another example of Sn ECM is shown in Figure 10.27, where colloidal form of ECM can be observed in addition to dendrite formation.

Silver ECM can occur on the PWB if there is exposed metal in the termination or pad finish, or it can occur on the surface of passive devices separate from the surface of the PWB. The occurrence of ECM on the surface of passive devices can potentially be a more serious reliability risk because of the current trend toward smaller size passives, which provides a ready site for ECM. A coating of Ag is commonly employed at the ends of the passive device to ensure that there is a good contact between the electrodes in a capacitor. However, since Ag is prone to ECM, it is advisable to isolate this Ag from the environment. Therefore, Ni is used as a barrier layer between the Ag base and the Sn outer layers. To be effective, this Ni layer should be continuous and free of cracks or gaps. In the event that the Ni layer is discontinuous, Ag can be exposed to the environment leading to dendrite formation as illustrated in Figure 10.28. Here, dendrites of Ag can be seen growing on the surface of the passive component after damp-heat reliability tests.

10.6. RELIABILITY TEST PRACTICES

Accelerated thermal cycling test practices are influenced not only by the design life and the operating environment but also by the nature of the PWB assembly. In a majority of cases, portable electronic hardware by its very nature has to be small, lightweight, and possess high I/O density. This implies the use of surface laminar circuitry or other HDI PWB technologies. Also, inherent is the use of small low profile packages.

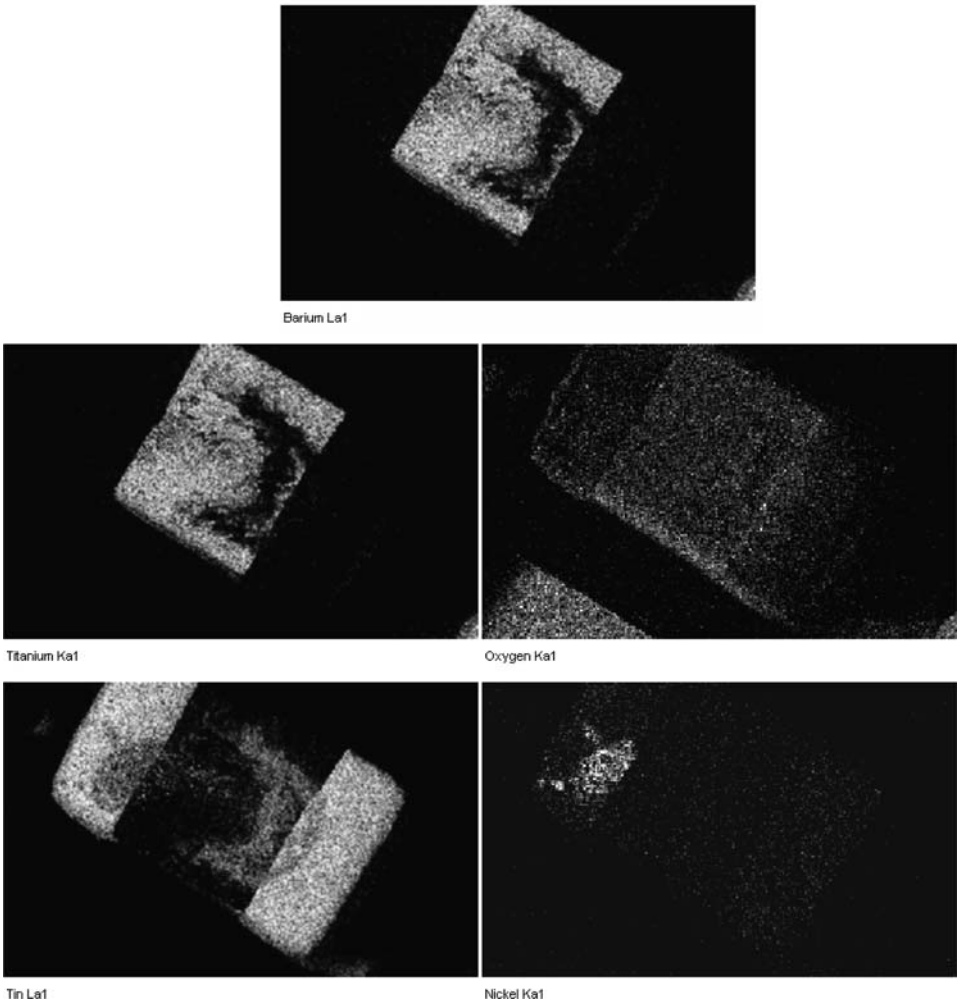


FIGURE 10.26. Elemental maps for the capacitor shown Figure 10.25 for (a) Ba, (b) Ti, (c) O, (d) Sn, and (e) Ni showing the presence of Sn ECM between the terminations and exposure of the Ni barrier layer under the consumed Sn surface at the anode.

It has been reported that the industry standard temperature cycle profile, where the upper and lower temperature dwells are invariant, leads to an underestimation of fatigue life [75]. It is well known that inelastic strain accumulation is generally proportional to fatigue life. It has been suggested that temperature fluctuations during upper dwell times can reduce elastic strain accumulation, and as such, using minicycles during dwell times will reduce the maximum inelastic strain. The magnitude of such inelastic strain reduction depends on the number of minicycles and their temperature ranges. Thus, selective superimposition of a judicious number of minicycles during the high temperature dwell may enable a more realistic fatigue life prediction.

In addition, portable electronic hardware involving radio communication features can have components such as power amplifiers and radio frequency devices that may run

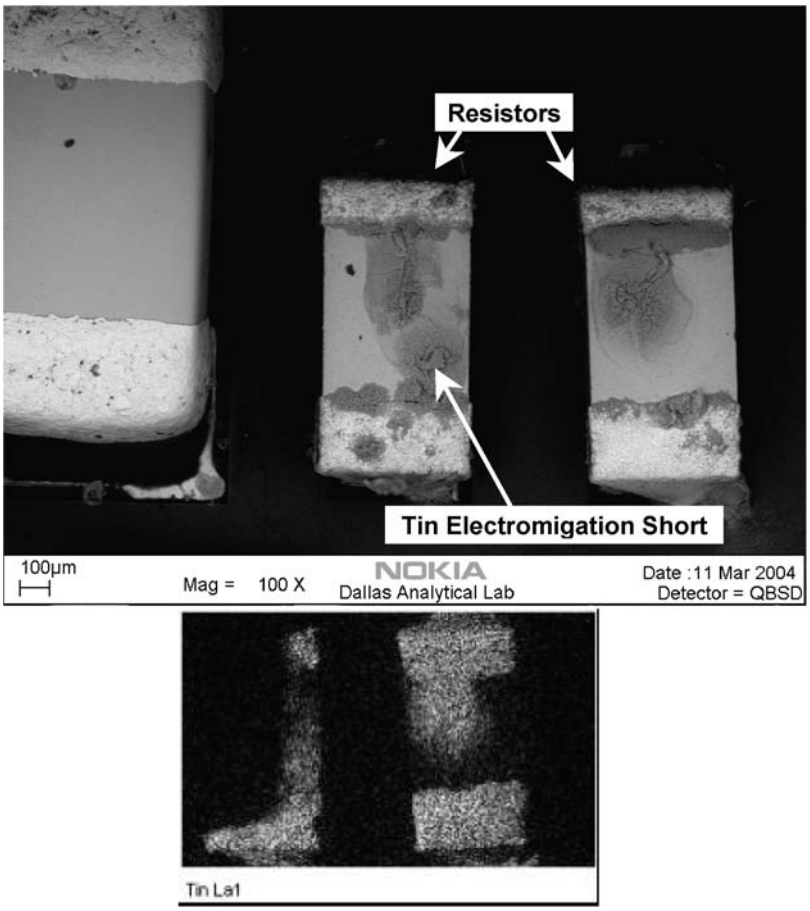


FIGURE 10.27. Tin electrochemical migration involving both formation of dendrites and on a resistor with pure tin termination.

hotter during operation in addition to the thermal exposure imposed by the environment. The thermal effects in such cases can cause excessive growth of interfacial intermetallics, which may be deleterious to the interconnection integrity. Power cycling tests may be much more appropriate in such cases.

Thermal and mechanical stress exposures in portable electronic hardware are rather frequent and some times concurrent in contrast to desktop machines, and the effect on product performance can be significant. For example the interfacial intermetallic growth, which by itself may not affect the solder joint integrity due to the compliance of the alloy, can progressively degrade the mechanical reliability. Thus, separate thermal and mechanical reliability assessments may not reflect the true product performance, as the synergistic effect is not taken into account. The effect of thermal aging on the mechanical reliability can be significant and should be considered in all reliability assessments.

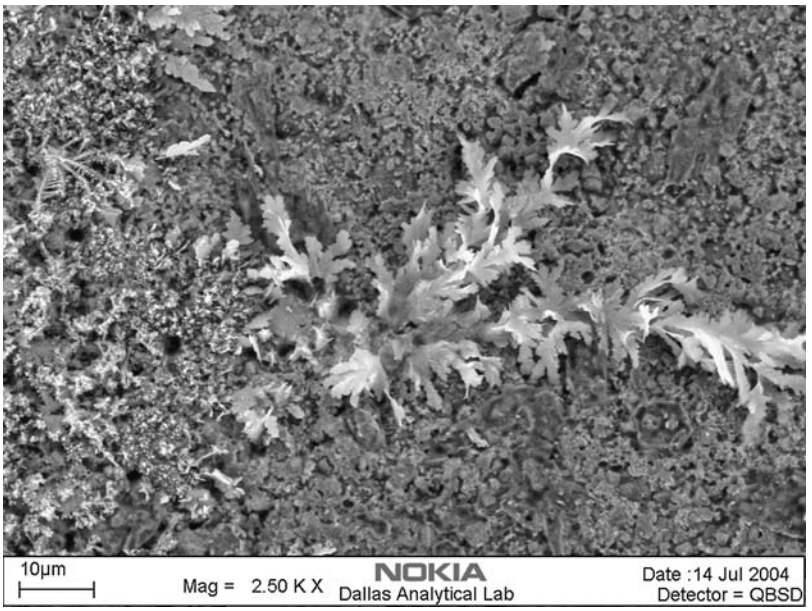


FIGURE 10.28. Silver electrochemical migration on a resistor with tin termination. Inadequate protection due to poor quality Ni barrier layers enabled the Ag to exhibit electrochemical migration.

10.7. SUMMARY

As portable consumer electronic hardware becomes more complex with multitudes of functions and increased data handling capacity, further miniaturization and higher levels of integration at all levels of packaging will be a natural trend. The reliability demands will be higher to ensure customer satisfaction and product acceptance. The implications for reliability, failure, and root cause analysis will be significant. More functions will be integrated into the device. The silicon device thickness will be in the range of 40–50 microns. Stacked devices, folded and stacked packages will be more prevalent with a combination of multiple levels of wire bonding and/or flip chip interconnection. Another emerging trend in packaging is the three dimensional integration at the wafer level. New materials that will have better mechanical properties and moisture resistance will be developed. More functions will be embedded into the printed wiring board and these may include active, passive and optical devices, attendant with new embedded interconnection schemes. The printed wiring board technology itself will witness revolutionary changes with thinner and improved materials capable of 10–25 µm vias, 10–20 µm lines and spaces, and structures involving several layers of stacked vias. Consequently, hitherto unknown failure mechanisms are likely to be encountered. As the feature sizes diminish the distinction between first and second level packaging becomes nebulous. Failure analysis, even at the printed wiring board PWB assemblies, will be a formidable challenge. With shorter product development cycles and faster to market business environment the need for more automated analytical tools with minimal operator intervention for rapid and repeatable root cause analysis will increase. Innovative reliability test practices will be needed to shorten the test durations to accommodate faster development schedules.

ACKNOWLEDGMENTS

The authors acknowledge the contributions of Sesil Mathew, Sambit K. Saha, Murali Hanabe, Steven Dunford, Nael Hannan and Laura Foss for technical discussions and permission to use their data, and appreciate the help provided by Leslie Landon, Tuula Stenberg, Elina Leivo, and Susanna Olli for technical discussions, and Michael Wellborn for providing a test sample. The management support of Timothy Fitzgerald was indispensable and appreciated.

REFERENCES

1. W.Q. Meeker and L.A. Escobar, Pitfalls of accelerated testing, IEEE Trans. on Reliability, 47(2), pp. 114–118 (1998).
2. W.Q. Meeker and M. Hamada, Statistical tools for the rapid development and evaluation of high-reliability products, IEEE Trans. on Reliability, 44(2), pp. 187–198 (1995).
3. R.J.K. Wassink, Soldering in Electronics, Electrochemical Publications Ltd., Great Britain, 1989.
4. R. Erich, R.J. Coyle, G.M. Wenger, and A. Primavera, Shear testing and failure mode analysis for evaluation of BGA ball attachment, IEEE/CPMT International Electronic Manufacturing Technology Symposium, 1999, pp. 16–22.
5. S. Canumalla, Test fixture and method, U.S. Patent #6,681,640, 2004.
6. M. Hanabe and S. Canumalla, Package to board interconnection shear strength (PBISS) behavior at high strain rates approaching mechanical drop, IEEE Electronic Components and Technology Conference, 2004, pp. 1263–1270.
7. S. Goyal and E.B. Buratynski, Methods for realistic drop testing, International Journal of Microcircuits and Electronic Packaging, 23(1) pp. 45–52 (2000).
8. S. Goyal, S. Upasani, and D.M. Patel, The role of case-rigidity in drop-tolerance of portable products, International Journal of Microcircuits and Electronic Packaging, 22(2), pp. 175–184 (1999).
9. C.T. Lim and Y.J. Low, Investigating the drop impact of portable electronic products, Proceedings of the IEEE Electronic Components and Technologies Conference, 2002, pp. 1270–1274.
10. C.T. Lim, C.W. Ang, L.B. Tan, S.K. Seah, W., and E.H. Wong, Drop impact survey of portable electronic products, Proceedings of the IEEE Electronic Components and Technologies Conference, 2003, pp. 113–120.
11. S. Goyal, E.B. Buratynski, and G.W. Elko, Role of shock response spectrum in electronic product suspension design, International Journal of Microcircuits and Electronic Packaging, 23(2), pp. 182–190 (2000).
12. S.K.W. Seah, C.T. Lim, E.H. Wong, V.B.C. Tan, and V.P.W. Shim, Mechanical response of PCBs in portable electronic products during drop impact, Proceedings of the IEEE Electronic Components and Technologies Conference, 2002, pp. 120–125.
13. Y.C. Ong, V.P.W. Shim, T.C. Chai, and C.T. Lim, Comparison of mechanical response of PCBs subjected to product-level and board-level drop impact tests, Proceedings of the IEEE Electronic Components and Technologies Conference, 2003, pp. 223–227.
14. K. Mishiro, S. Ishigawa, M. Abe, T. Kumai, Y. Higashiguchi, and K. Tsubone, Effect of the drop impact on BGA/CSP package reliability, Microelectronics Reliability, 42, pp. 77–82 (2002).
15. S. Canumalla, S. Shetty, and N. Hannan, Effect of corner-underfill voids on the chip scale package (CSP) performance under mechanical loading, 28th International Sympoium for Society for Testing and Failure Analysis, 3–7 November, Phoenix, AZ, 2002, pp. 361–370.
16. N. Hannan and P. Viswanadham, Critical aspects of reworkable underfills for portable consumer products, Proceedings of the IEEE Electronic Components and Technologies Conference, 2001, pp. 181–187.
17. A. Kujala, T. Reinikainen, and W. Ren, Transition to Pb-free manufacturing using land grid array packaging technology, Proceedings of the IEEE Electronic Components and Technologies Conference, 2002, pp. 359–364.
18. B.J. Toleno and J. Schneider, Processing and reliability of corner bonded CSPs, International Electronics Manufacturing Technology Symposium, 2003, pp. 299–304.
19. G. Tian, Y. Liu, P. Lall, R.W. Johnson, S. Abderrahman, M. Palmer, N. Islam, and J. Suhling, Drop reliability of corner bonded CSP in portable products, International Electronic Packaging Technical Conference, July 6–11, Hawaii, USA, 2003.

20. J. Wu, G. Song, C.-P. Yeh, and K. Wyatt, Drop impact simulation and test validation of telecommunication products, Intersociety Conference on Thermal Phenomena, 1998, pp. 330–336.
21. T.Y. Tee, H.S. Ng, C.T. Lim, E. Pek, and Z. Zhong, Impact life prediction modeling of TFBGA packages under board level drop test, Microelectronics Reliability, 43, pp. 1131–1142 (2004).
22. P. Lall, D. Panchagade, Y. Liu, W. Johnson, and J. Suhling, Models for reliability prediction of fine-pitch BGAs and CSPs in shock and drop impact, Proceedings of the IEEE Electronic Components and Technologies Conference, 2004, pp. 1296–1303.
23. L. Zhu, Modeling technique for reliability assessment of portable electronic product subjected to drop impact loads, Proceedings of the IEEE Electronic Components and Technologies Conference, 2003, pp. 100–104.
24. E. Suhir, Nonlinear dynamic response of a flexible printed circuit board to a shock load applied to its support contour, IEEE Electronic Components and Technology Conference, 1991, pp. 388–399.
25. E. Suhir and R. Burke, Dynamic response of a rectangular plate to a shock load, with application to portable electronic products, IEEE Trans. on Components, Packaging, and Manufacturing Technology, Part B: Advanced Packaging, 17(3), pp. 449–460 (1994).
26. E. Suhir, Shock protection with a nonlinear spring, IEEE Trans. on Components, Packaging, and Manufacturing Technology, Part A, 18(2), pp. 430–437 (1995).
27. E. Suhir, Dynamic response of a one-degree-of-freedom linear system to a shock load during drop tests: Effect of viscous damping, IEEE Trans. on Components, Packaging, and Manufacturing Technology, Part A, 19(3), pp. 435–440 (1996).
28. E. Suhir, Is the maximum acceleration an adequate criterion of the dynamic strength of a structural element in an electronic product? IEEE Trans. on Components, Packaging, and Manufacturing Technology, Part A, 20(4), pp. 513–517 (1997).
29. E. Suhir, Could shock tests adequately mimic drop test conditions? IEEE Electronic Components and Technology Conference, 2002, pp. 563–573.
30. R. Darveaux and A. Syed, Reliability of area array solder joints in bending, SMTA International Symposium, 2000.
31. H. Juso, Y. Yamaji, T. Kimura, K. Fujita, and M. Kada, Board level reliability of CSP, IEEE Electronic Components and Technologies Conference, 1998, pp. 525–531.
32. P. Geng, P. Chen, and Y. Ling, Effect of strain rate on solder joint failure under mechanical load, IEEE Electronic Components and Technologies Conference, 2002, pp. 97–978.
33. K. Harada, S. Baba, Q. Wu, H. Matsushima, T. Matsunaga, Y. Uegai, and M. Kimura, Analysis of solder joint fracture under mechanical bending test, IEEE Electronic Components and Technologies Conference, 2003, pp. 1731–1737.
34. T.C. Chiu, K. Zeng, R. Stierman, D. Edwards, and K. Ano, Effect of thermal aging on board level drop reliability for Pb-free BGA packages, Proceedings of the IEEE Electronic Components and Technologies Conference, 2004, pp. 1256–1262.
35. D. Rooney, Castello, N.T. , M. Cibulsky, D. Abbott, and D.J. Xie, Materials characterization of the effect of mechanical bending on area array package interconnects, Microelectronics Reliability, 39, pp. 463–477 (2003) (in press).
36. L. Leicht and A. Skipor, Mechanical cycling fatigue of PBGA package interconnects, Proceedings of 30th International Symposium on Microelectronics, 14–16 Oct., 1998, pp. 802–807.
37. K. Jonnalagadda, Reliability of via-in-pad structures in mechanical cycling fatigue, Microelectronics Reliability, 42, pp. 253–258 (2002).
38. K. Jonnalagadda, M. Patel, and A. Skipor, Mechanical bend fatigue reliability of lead-free PBGA assemblies, The Eighth Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems, (ITHERM 2002), 2002, pp. 915–918.
39. E.A. Stout, N.R. Sottos, and A.F. Skipor, Mechanical characterization of plastic ball grid array package flexure using moire interferometry, IEEE Trans. on Advanced Packaging, 23(4), pp. 637–645 (2000).
40. S. Shetty, A. Dasgupta, V. Halkola, V. Lehtinen, and T. Reinikainen, Bending fatigue of chip scale package interconnects, ASME International Mechanical Engineering Congress and Exposition, Orlando, Florida, Nov. 5–10, 2000.
41. U.D. Perrera, Evaluation of reliability of μBGA solder joints through twisting and bending, Microelectronics Reliability, 39, pp. 391–399 (1999).
42. S. Canumalla, H.-D. Yang, P. Viswanadham, and T. Reinikainen, Package to board interconnection shear strength (PBISS): Effect of surface finish, PWB build-up layer and chip scale package structure, IEEE Trans. on Components and Packaging Technologies, 27(1), pp. 182–190 (2004).
43. N.F. Enke, T.J. Kilinski, S.A. Schroeder, and J.R. Lesniak, Mechanical Behaviors of 60/40 Tin-lead solder lap joints, IEEE Trans. on Components, Hybrids, and Manufacturing Technology, 12(4), pp. 459–468 (1989).

44. T. Shohji, T. Yoshida, Takahashi, and S. Hioki, Tensile properties of Sn-Ag based lead-free solders and strain rate sensitivity, Materials Science and Engineering A, 366, pp. 50–55 (2004).
45. M. Tullmin and P.R. Roberge, Corrosion of metallic materials, IEEE Trans. on Reliability, 44(2), pp. 271–278 (1995).
46. J.D. Guttenplan, Corrosion in the electronics industry, ASM Metals Handbook, 9th edn, Vol. 13, ASM International, Metals Park, OH, USA, 1987.
47. A.J. Raffalovich, Corrosive effects of solder flux on printed circuit boards, IEEE Trans. on Components, Hybrids, and Manufacturing Technology, 7(4), pp. 155–162 (1971).
48. K. Yasuda, S. Umemura, and T. Aoki, Degradation mechanisms in tin- and gold-plated connector contacts, IEEE Trans. on Components, Hybrids, and Manufacturing Technology, 10(3), pp. 456–462 (1987).
49. G.T. Kohman, H.W. Hermance, and G.H. Downes, Silver migration in electrical insulation, Bell Systems Technolgy Journal, 34, p. 1115 (1955).
50. A. Shumka, and R.R. Piety, Migrated gold resistive shorts in microcircuits, Proceedings of the International Reliability Physics Symposium, 1975, pp. 93–98.
51. A. Dermarderosian, The electrochemical migration of metals, Proceedings of the International Society of Hybrid Microelectronics, 1978, p. 134.
52. G. Ripka and G. Harsanyi, Electrochemical migration in thick-film conductors and chip attachment resins, Electrocomponents Science and Technology, 11, p. 281 (1985).
53. L.J. Turbini, J.A. Jachim, G.A. Freeman, and J.F. Lane, Characterizing water soluble fluxes: Surface insulation resistance vs. electrochemical migration, IEEE/CHMT International Electronics Manufacturing Technology Symposium, 1992, pp. 80–84.
54. G. Harsanyi and G. Inzelt, Comparing migratory resistive short formation abilities of conductor systems applied in advanced interconnection systems, Microelectronics Reliability, 41, pp. 229–237 (2001).
55. T. Takemoto, R.M. Latanison, T.W. Eagar, and Matsunawa, A, Electrochemical migration tests of solder alloys in pure water, Corrosion Science, 39(9), pp. 1415–1430 (1997).
56. F.G. Grunhaner, T.W. Griswold, and P.J. Clendening, Migratory gold resistive shorts: Chemical aspects of failure mechanism, Proceedings of the International Reliability Physics Symposium, 1975, p. 99.
57. N.L. Sbar, Bias humidity performance of encapsulated and unencapsulated Ti-Pd-Au thin film conductors in an environment contaminated with Cl_2, IEEE Transactions on Parts, Hybrids, and Packaging, 12, p. 176 (1976).
58. G. Harsanyi, Electrochemical processes resulting in migrated short failures in microcircuits, IEEE Trans. on Components, Packaging, and Manufacturing Technology A, 18(3), pp. 602–610 (1995).
59. D.J. Lando, J.P. Mitchell, and T.L. Welsher, Conductive anodic filaments in reinforced polymeric dielectrics: Formation and prevention, International Reliability Physics Symposium, 1979, pp. 51–63.
60. P. Viswanadham and P. Singh, Failure Modes and Mechanisms in Electronics Packages, Chapman and Hall, New York, 1997.
61. M.E. McDowell, Tin whiskers: A case study, Aerospace Applications Conference, 1993, pp. 207–215.
62. J.P. Siplon, G.J. Ewell, E. Frasco, J.A. Brusse, and T. Gibson, Tin whiskers on discrete components: The problem, 28th International Sympoium for Society for Testing and Failure Analysis, 3–7 November, Phoenix, AZ, 2002, pp. 421–434.
63. K. Zeng and K.N. Tu, Six cases of reliability study of lead-free solder joints in electronic packaging technology, Materials Science and Engineering Reviews, 38, pp. 55–105 (2002).
64. B.-Z. Lee and D.N. Lee, Spontaneous growth mechanism of tin whiskers, Acta Materialia, 46(10), pp. 3701–3714 (1998).
65. R. Schetty, Minimization of tin whisker formation for lead free electronics finishing, IPC Works Conference, Miami, USA, 2000.
66. T.Y. Tee, H.S. Ng, D. Yap, X. Baraton, and Z. Zhong, Board level solder joint reliability modeling and testing of TFBGA packages for telecommunication applications, Microelectronics Reliability, 43, pp. 1117–1123 (2003).
67. S.K. Saha, S. Mathew, and S. Canumalla, Effect of Intermetallic Phases in Mechanical Drop Environment: 96.5Sn3.5Ag Solder on Cu and Ni/Au Pad Finishes, IEEE Electronic Components and Technologies Conference, 2004, pp. 1288–1295.
68. M. Amagai, Y. Toyoda, T. Ohinishi, and S. Akita, High drop test reliability of lead free solders, IEEE Electronic Components and Technologies Conference, 2004, pp. 1304–1309.
69. S. Dunford, S. Canumalla, and P. Viswanadham, Intermetallic morphology and damage evolution under thermomechanical fatigue of lead-free solder interconnections, IEEE Electronic Components and Technology Conference, 2004, pp. 726–736.

70. N. Biunno, A root cause failure mechanism for solder joint integrity of nickel/immersion gold surface finishes, IPC Printed Circuits Expo, Long Beach, CA, 1999, pp. 1–9.
71. E. Bradley and K. Banerji, Effect of PCB finish on the reliability and wettability of ball grid array packages, IEEE Transactions on Components, Packaging, and Manufacturing Technology, Part B, 19(2), pp. 320–330 (1996).
72. J.N. Lahti, R.H. Delaney, and J.N. Hines, The characteristic wearout process in epoxy-glass printed circuits in high density electronic packaging, Proceedings of the 17th Annual Reliability Physics Symposium, 1979, p. 39.
73. P. Dumoulin, J.-P. Seurin, and P. Marce, Metal migration outside the package during accelerated life tests, IEEE Trans. on Components, Packaging, and Manufacturing Technology, 5(4), pp. 479–486 (1982).
74. S.J. Krumbein, Metallic electromigration phenomena, IEEE Trans. on Components, Hybrids, and Manufacturing Technology, 11(1), pp. 5–15 (1988).
75. T. Dishong, C. Basaran, N. Cartwright, Y. Zhao, and H. Liu, Impact of temperature cycle profile on fatigue life of solder joints, IEEE Trans. on Components, Packaging, and Manufacturing Technology, Part B: Advanced Packaging, 25(3), pp. 433–438 (2002).
76. R. Darveaux, Constitutive relations for tin-based solder joints, IEEE Trans. on Components, Hybrids, and Manufacturing Technology, 15(6) (1992).
77. T. Kawanabe and K. Otsuka, Metal migration in electronic components, Proceedings of Electronic Components Conference, 1982, p. 99.
78. R.C. Benson, B.M. Romanesko, J.E. Weiner, B.N. Nall, and H.K. Charles, Metal electromigration induced by solder flux residue in hybrid microcircuits, IEEE Trans. on Components, Hybrids, and Manufacturing Technology, 11(4), pp. 363–370 (1988).
79. D. Xie, M. Arra, S. Yi, and D. Rooney, Solder joint behavior of area array packages in board level drop for hand held devices, IEEE Electronic Components and Technology Conference, 2003, pp. 130–135.

11

MEMS Packaging and Reliability

Y.C. Lee

Department of Mechanical Engineering, Campus Box UCB 427, University of Colorado, Boulder, CO80309-0427, USA

11.1. INTRODUCTION

Microelectromechanical systems (MEMS) technology enables us to create different sensing and actuating devices integrated with other microelectronic, optoelectronic, microwave, thermal, and mechanical devices for advanced microsystems. Semiconductor fabrication processes allow for cost effective production of these micro-sensing or actuation devices in the 1–100 μm size scale. Figure 11.1 illustrates a typical design and manufacturing process for a MEMS device. This illustration highlights some of the differences between MEMS and microelectronics fabrication and packaging. During the design, solid modeling is required since electro-thermal-mechanical coupling is essential to the functions of most of MEMS devices. The fabrication often involves deposition and etching of micron-thick layers with controlled mechanical and electrical properties [1,2]. In many devices, after the completion of the fabrication process, the sacrificial materials are removed by etching in order to release the device for mechanical movements. This release process is usually the first step in the MEMS packaging. The released device shown in the figure represents a configuration for pressure sensors or accelerometers or an element of an array for optical micro-mirrors and RF switches. After release, the devices can be tested on the wafer-level, followed by dicing. The released, diced device is assembled and sealed in a package. These testing, dicing, assembly and sealing steps are very challenging. Without proper protection, the micro-scale, movable features could be damaged easily during these steps [3]. As a result, it is always desirable to replace the process illustrated here by wafer-level packaging [4].

Hundreds of MEMS-based sensors and actuators and systems have been demonstrated and the number of their applications is growing. A few examples of their diverse applications are listed below [5]:

1. Pressure sensors: for sensing manifold air pressure and fuel pressure to decrease emission and fuel consumption; for measuring blood pressure.
2. Inertial sensors: accelerometers for measuring acceleration for launching air bags; gyros for measuring angular velocity to stabilize ride and to detect rollover.
3. Chemical micro sensors: for fast, disposable blood chemistry analysis; gas sensors.

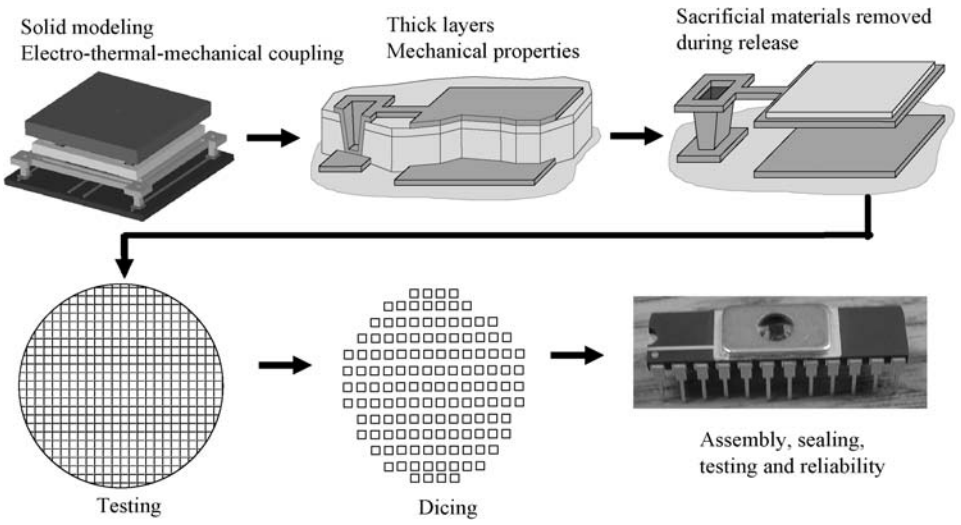


FIGURE 11.1. MEMS design, fabrication and packaging.

4. Optical MEMS: micromirrors for projection displays; optical switches for wavelength division multiplex switches; attenuators or micro-devices for active alignments for optical microsystems; micro-displays or paper-thin, direct-view displays.
5. Radio frequency (RF) MEMS: micro-resonators for integrated RF transceiver chips; RF switches for microwave systems.
6. Microfluidic MEMS: DNA hybridization arrays or similar lab-on-a-chips for biomedical and biochemical development, bio-analysis and diagnostic; printerheads for ink jet printing.
7. Power MEMS: on-chip power generation and energy storage for portable systems.
8. MEMS-based data storage: micro-positioning and tracking devices for magnetic, optical, thermal, or atomic force data tracks; micro-mirrors for optical beam steering.
9. Microsurgical instruments: for non-invasive techniques, intra-vascular devices, and laparoscopic procedures.

As new applications are developed, the MEMS market is experiencing a period of dramatic growth that is shown in Figure 11.2. Using Tire Pressure Monitoring System (TPMS) as an example, there is a need of 68 millions TPMS with a market value of $102M by year 2007. Similar high volume applications are for (a) mobile phones with microphones and acceleration sensors for human–machine interaction, gyroscopes for image stabilization and RF MEMS switches for transceivers; (b) hard disk drives with acceleration sensors for free fall detection; and (c) camcorders and cameras with gyroscope for image stabilization. In all these applications demanding low cost and small size, MEMS packaging is usually a major consideration.

MEMS packaging can be defined as all the integrations after the microfabrication of the device is complete. They include post-processing release, package/substrate fabrication, assembly, testing, and reliability assurance. Reliability is one of the performance measures that are strongly affected by the package as well as the device. Assurance of the reliability is considered as a packaging activity since packaging engineers rather than fabrication

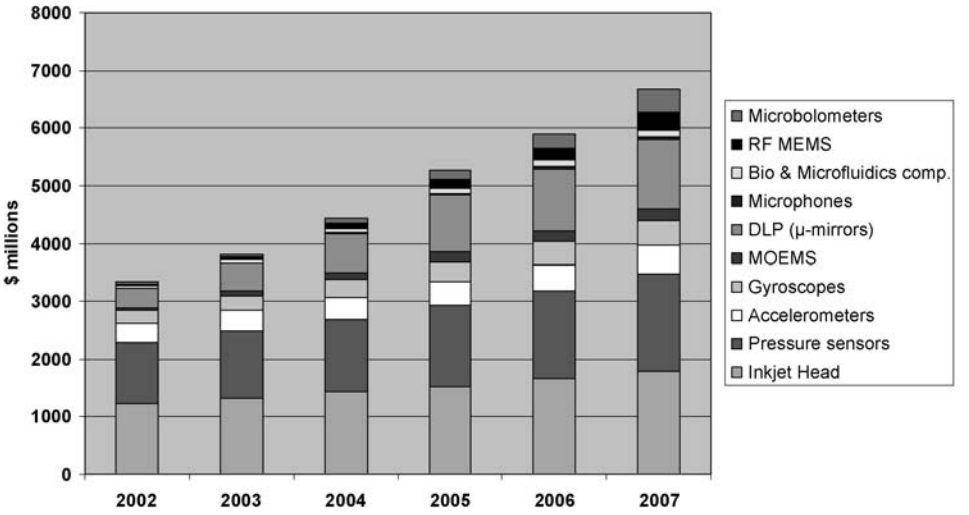


FIGURE 11.2. World wide market for MEMS devices.

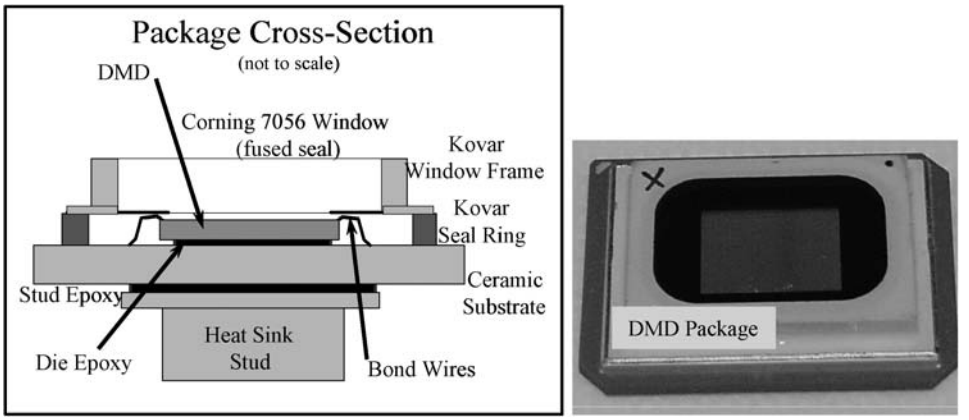


FIGURE 11.3. Package for digital mirror device (courtesy of John P. O'Connor, Texas Instruments).

engineers usually conduct environmental protection processes, burn-ins, and accelerated tests to ensure the production of a reliable MEMS device.

Figure 11.3 shows a package developed for Texas Instruments' digital mirror device (DMD). DMD has millions of micro-mirrors and is used for projection displays. This device has proven an important fact: mechanical devices can be switched over trillions of cycles while achieving the same reliability level as their electronic counterparts [6,7]. After release, a self-assembled monolayer (SAM) can be used to coat the device to avoid a moisture-induced striction problem. If needed, getters can be used to remove particles or moisture inside the package [8]. The DMD package is hermetically sealed with a Kovar ring. Particles can cause reliability failures, so the device has to be packaged in a Class-10 cleanroom. Outgassing of all the package materials should also be controlled. The large glass window is the critical optical interface between the DMD and other optical components for a projection system. Therefore, the window's alignment with the DMD is important [8]. Another concern is the hysteresis behavior of the mirror's aluminum material. The

mirror may be difficult to move when it stays at one titling angle for too long. This creep-related problem is temperature dependent; as a result, thermal management has to control the device temperature to avoid the hysteresis effect [7].

For inertial sensors, the above mentioned packaging approach is too expensive. It is replaced by another approach compatible with microelectronic packaging. The compatibility is achieved by the establishment of wafer-level capping [9–13]. As shown in Figure 11.4, silicon or glass caps are bonded onto a MEMS wafer for hermetic and/or vacuum seal. For a batching process, these caps are fabricated in another wafer as shown in Figure 11.5. The hermetic sealing is accomplished by wafer-to-wafer anodic bonding, soldering, or glass sealing. The capped MEMS devices are diced and packaged through injection plastic molding. As shown in Figure 11.6, this plastic packaging process is compatible with that used in microelectronic packaging. Therefore, packaging cost and size are reduced substantially. In addition to this example, there is an alternative with fully integrated device-to-cap fabrication. With a good design and custom-fabrication, a MEMS device and its encapsulation cover are fabricated in the same process run [14]. Another alternative is to replace the inorganic capping by liquid crystal polymer (LCP) [15,16]. Wafer-level capping and its improved version for wafer-level packaging are enabling technologies for cost and size reduction demanded by today's microsystems integrating MEMS and electronics.

MEMS packaging has been and continues to be a major challenge. The packaging cost is about 50% to 90% of the total cost of a MEMS product. Packaging should allow some moving parts to interact with other components through optical, electrical, thermal, mechanical, or chemical interfaces. As a result, many MEMS packaging problems are new

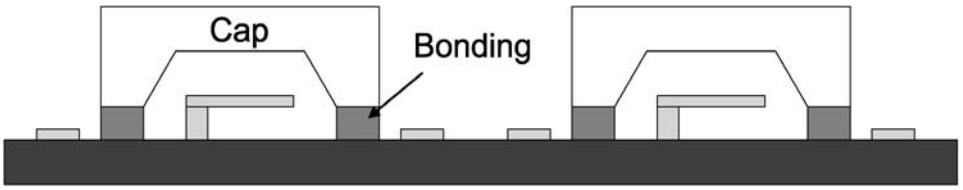


FIGURE 11.4. Wafer-level capping for hermetic sealing of MEMS.

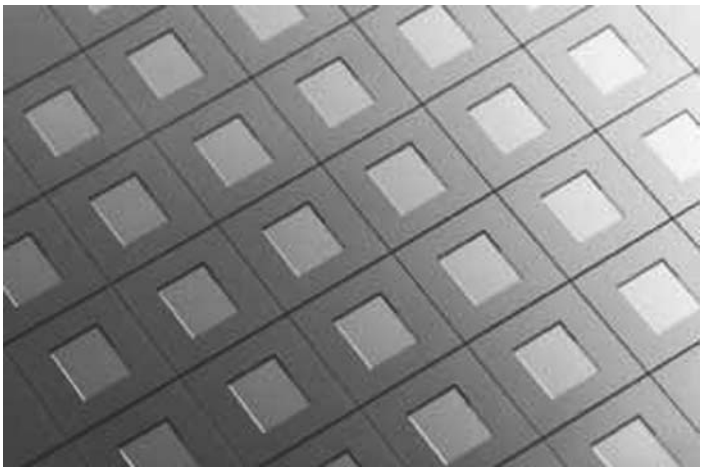


FIGURE 11.5. A wafer with silicon caps.

to most of the electronic packaging engineers. In a National Science Foundation (NSF) workshop, several major MEMS packaging and reliability challenges have been identified [17]. Here are a few examples:

- Vacuum packaging may be needed when viscous damping is important.
- Die-attachment may create severe thermal stresses that affect the accuracy of pressure measurement.
- Thermal strains may affect the performance of membrane devices.
- Moisture can cause striction problems.

These new problems are usually dependent on specific MEMS functions. MEMS package provides functional interfaces between the MEMS device and the environment. These interfaces are directly related to the applications. Unfortunately (or fortunately), MEMS has a large number of diverse applications as listed above. As a result, a variety of functional interfaces are needed such as: optical, RF, thermal (radiation, conduction, or convection), fluids (liquids or gases), mechanical (body or surface loadings), and others (e.g., radiation, magnetic, etc.). Clearly indicated by this long list of interfaces, there will be no "standard" packages to meet the requirements of all the MEMS applications. The above mentioned wafer-level capping for inertia sensors is one of the best solutions to insert MEMS packaging into existing microelectronic packaging infrastructure.

In addition to functional interfaces, reliability is another major packaging consideration. Striction, fracture and fatigue, mechanical wear with respect to frequency and humidity, and shock and vibration effects are the major causes of MEMS failures. During the last 20 years, MEMS products have proven to be reliable [7,18,19]. The most reliable MEMS devices are hermetically packaged single-point contact or no-contact devices. Recently, novel MEMS devices with surface contacts have reached impressive reliability levels with billions or hundreds of billions of surface impacts. It is a significant improvement from the early studies on RF MEMS [20].

With impressive technology advancement, MEMS sensors and actuators are no longer niche applications. In the near future, every automobile will use 50 to 100 MEMS components and every cell phone will have at least 3 MEMS components. Every MEMS component has to be integrated with other microelectronic, optoelectronic or RF compo-

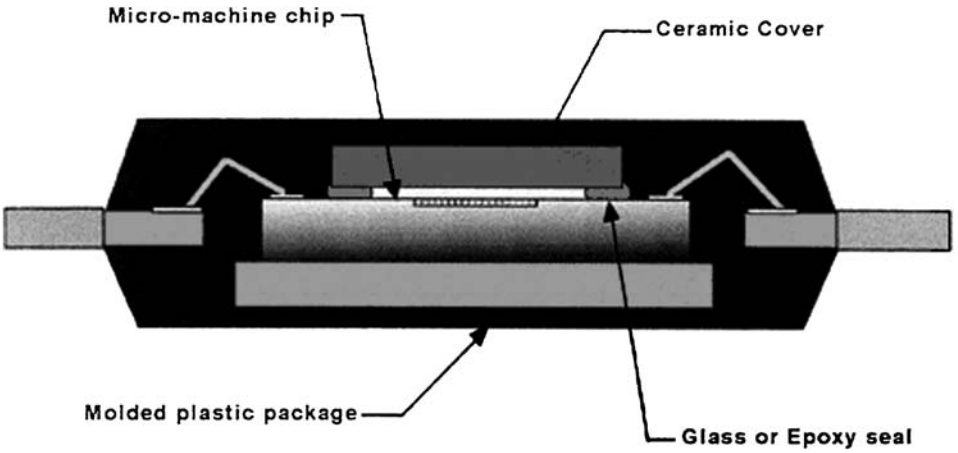


FIGURE 11.6. Plastic molding of MEMS devices capped.

nents. With such a large scale impact, we expect to see more advanced packaging technologies to be developed for the MEMS-based microsystems. In the following sections, we will describe (1) flip-chip assembly for hybrid integration, (2) soldered assembly for three-dimensional MEMS, (3) flexible circuit boards for MEMS, and (4) atomic layer deposition for reliable MEMS. They are different from the aforementioned MEMS packaging technologies being used to manufacture current products. The understanding of these new approaches will provide an insight into future MEMS packaging and reliability activities.

11.2. FLIP-CHIP ASSEMBLY FOR HYBRID INTEGRATION

MEMS devices have to be integrated with other electronic devices. Monolithic integration is always desirable; however, hybrid integration may be more practical due to its ability to integrate mixed-technology devices. For hybrid integration, flip-chip assembly could result in the smallest size while achieving superior performance. Such an assembly technology will be described in this section using PolyMUMPs-based MEMS as an example.

Figure 11.7 shows the cross-sections of PolyMUMPs (Polysilicon-Based Multi-User MEMS Processes), with its polysilicon and silicon oxide layers [1,21]. The oxide layers are sacrificial layers and are removed with HF after fabrication. An example of a typical design with over 50 different device layouts is shown in Figure 11.8. PolyMUMPs is only one of the MEMS foundry processes; there are quite a few other services using surface or bulk micro-machining or LIGA processes.

For hybrid integration, the MEMS devices manufactured in a foundry should be integrated with other devices on a new, common substrate. A flip-chip assembly process with silicon removal technology has been developed for such transfer and integration [22–24]. A 1D variable capacitor illustrated in Figures 11.9(a) and (b) is a good example. The MEMS fabricated in a silicon substrate was transferred to a ceramic RF substrate [25]. By using flexures having varying stiffness levels, the plates of the array would snap down indi-

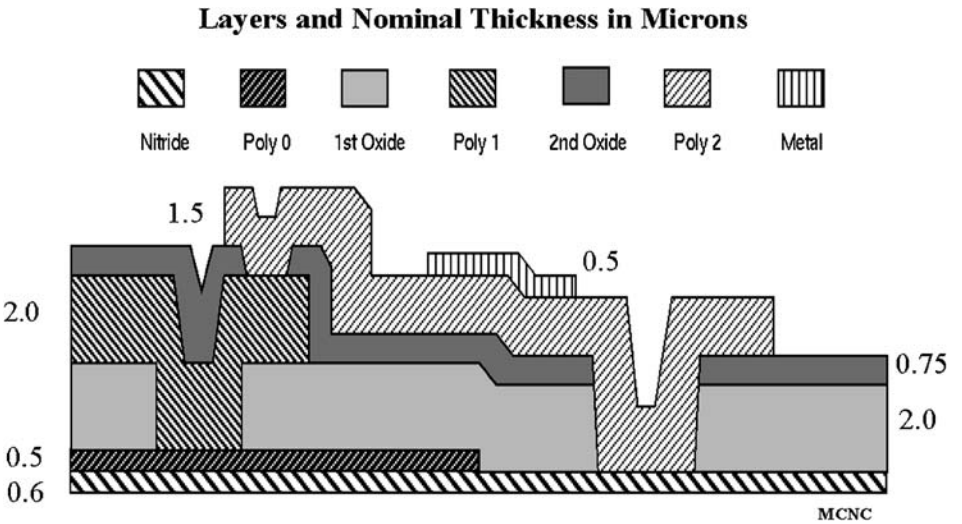


FIGURE 11.7. Cross-sections of PolyMUMPS (Multi-User MEMS Processes) foundry process.

vidually, and in sequence to change the capacitance. However, such an ideal operation was not achieved due to the following problems resulting from poor thermo-mechanical behaviors. As shown in Figure 11.10(a) for the capacitor with cantilever beams, bond height variations would result in non-repeatable capacitive performance. The warped beams would change their configurations away from two parallel plates defined by the bond and the dimple at the tip of each beam. With curved plates, the desirable digital snap-down sequence could not happen. The alternative fixed-fixed beams are shown in Figure 11.10(b), but this design still suffered large capacitance variations due to the warpage resulting from a thermal mismatch with different coefficients of thermal expansion (CTEs) between the MEMS device and the substrate. The digital increments in capacitance could be lost due to uncontrollable pull-in voltages associated with the varied bond heights and warpage.

In addition, this 1D variable capacitor flip-chip assembled required immersion of the chip in Hydrofluoric Acid after the assembly. Such immersion was slow and could damage some materials in the assembly [22–25]. Therefore, when the 1D variable capacitor was improved to a new two-dimensional (2D) device, we decided to improve the design and

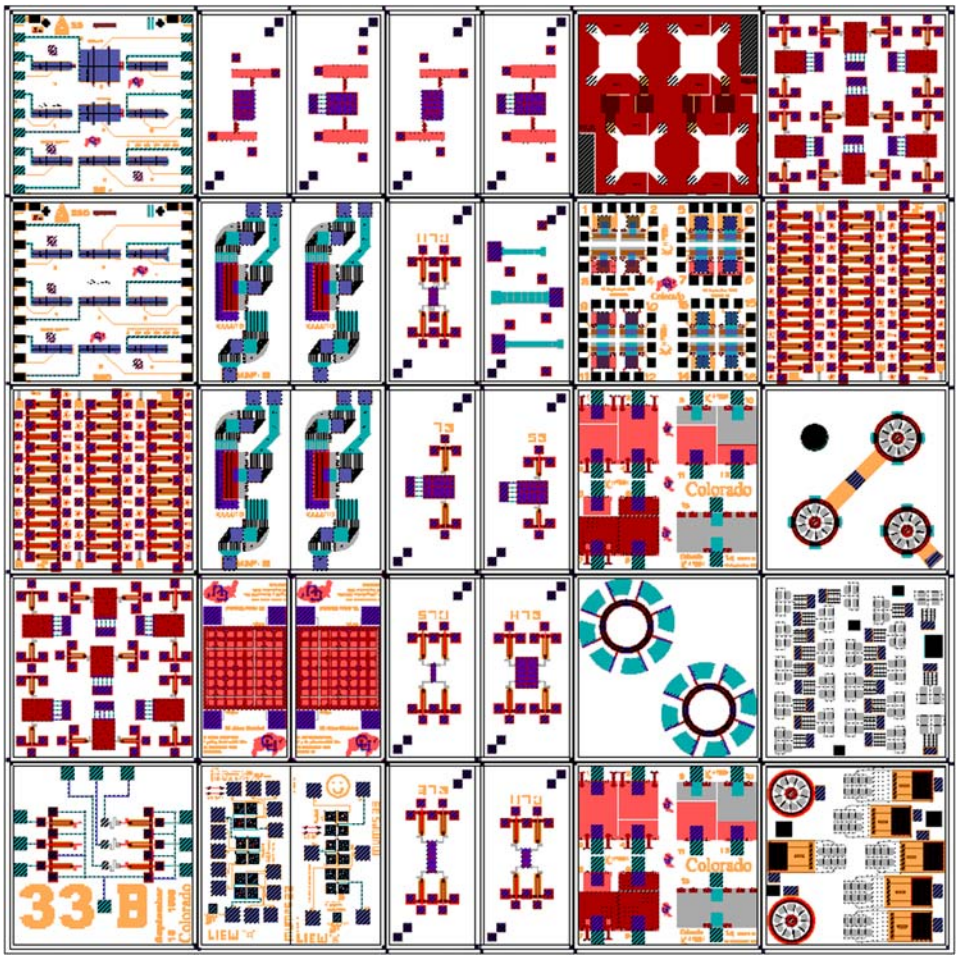


FIGURE 11.8. A typical 1 cm × 1 cm chip design using PolyMUMPs.

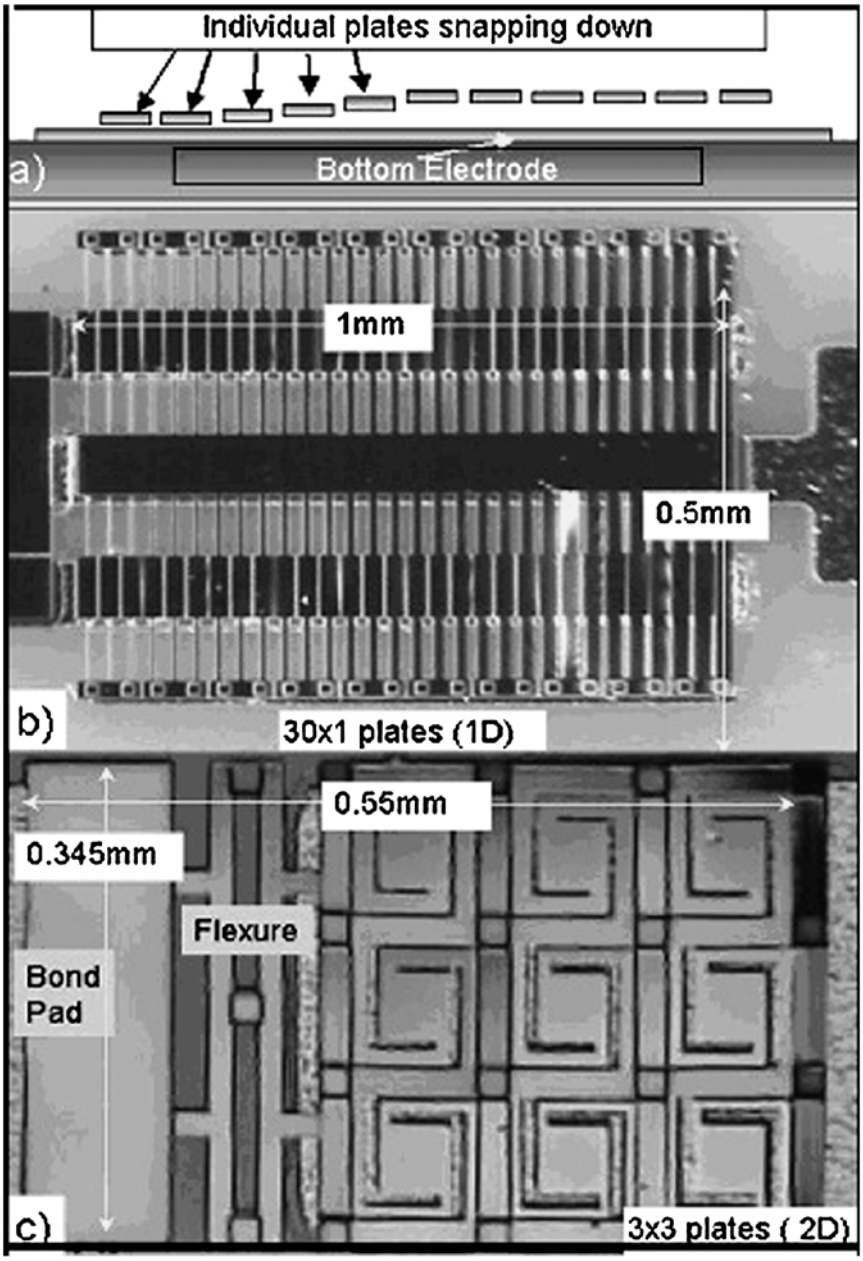


FIGURE 11.9. (a) Principle of digital pull-in in MEMS variable capacitors; (b) a 2-terminal, 1D variable capacitor (c) a 2D 3×3 MEMS variable capacitor.

the assembly process. The 2D device is shown in Figure 11.11. The device consisted of five components. Tethers connected the pre-assembly released MEMS to the silicon. The bonding pads joined the device to the new alumina substrate through solder bumps. Two compliant flexures accommodated the thermal mismatch between the silicon and alumina substrate during the flip-chip assembly. Arrays of 2×2, 3×3 or 4×4 capacitor plates

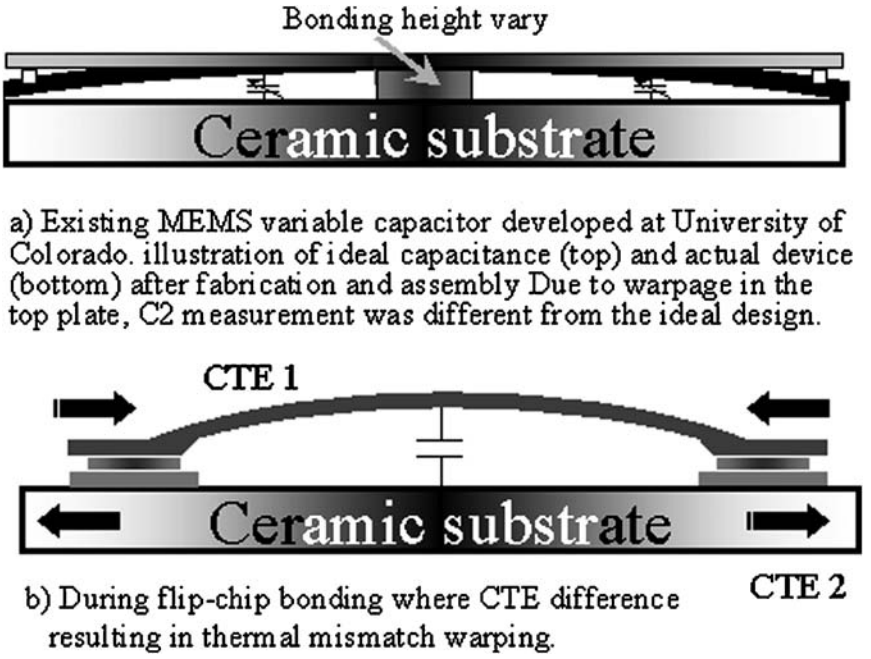


FIGURE 11.10. (a) Flip-chip assembly of MEMS with cantilever beams suffering from bond height variations and beam bending and (b) the assembly with fixed-fixed beams suffering from warpage due to a CTE mismatch.

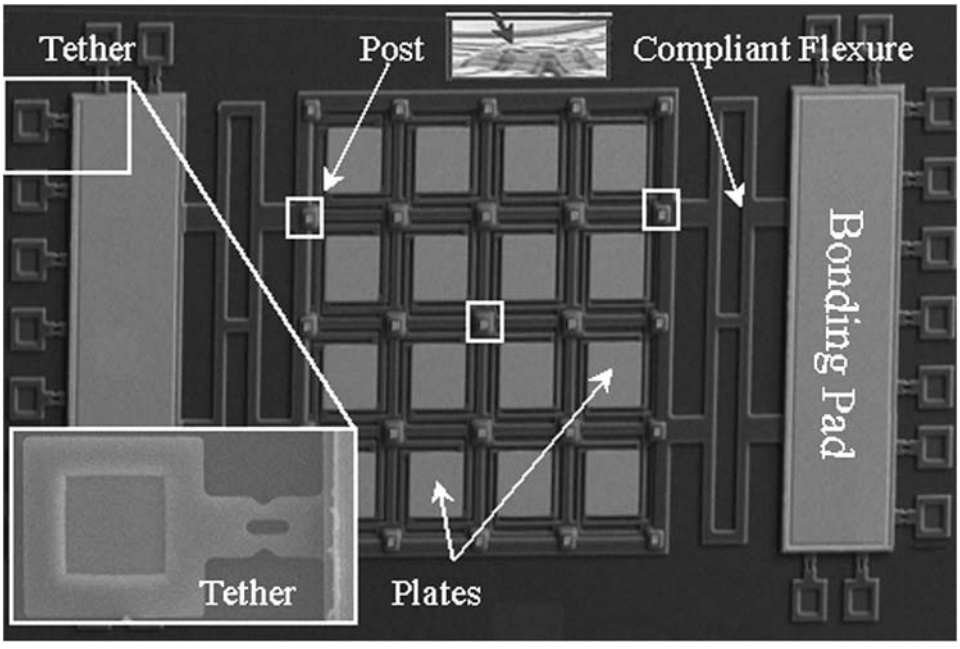


FIGURE 11.11. A variable capacitor featured with tethers, bonding pads, compliant flexures, 5×5 "posts" and 4×4 plates on the host silicon substrate.

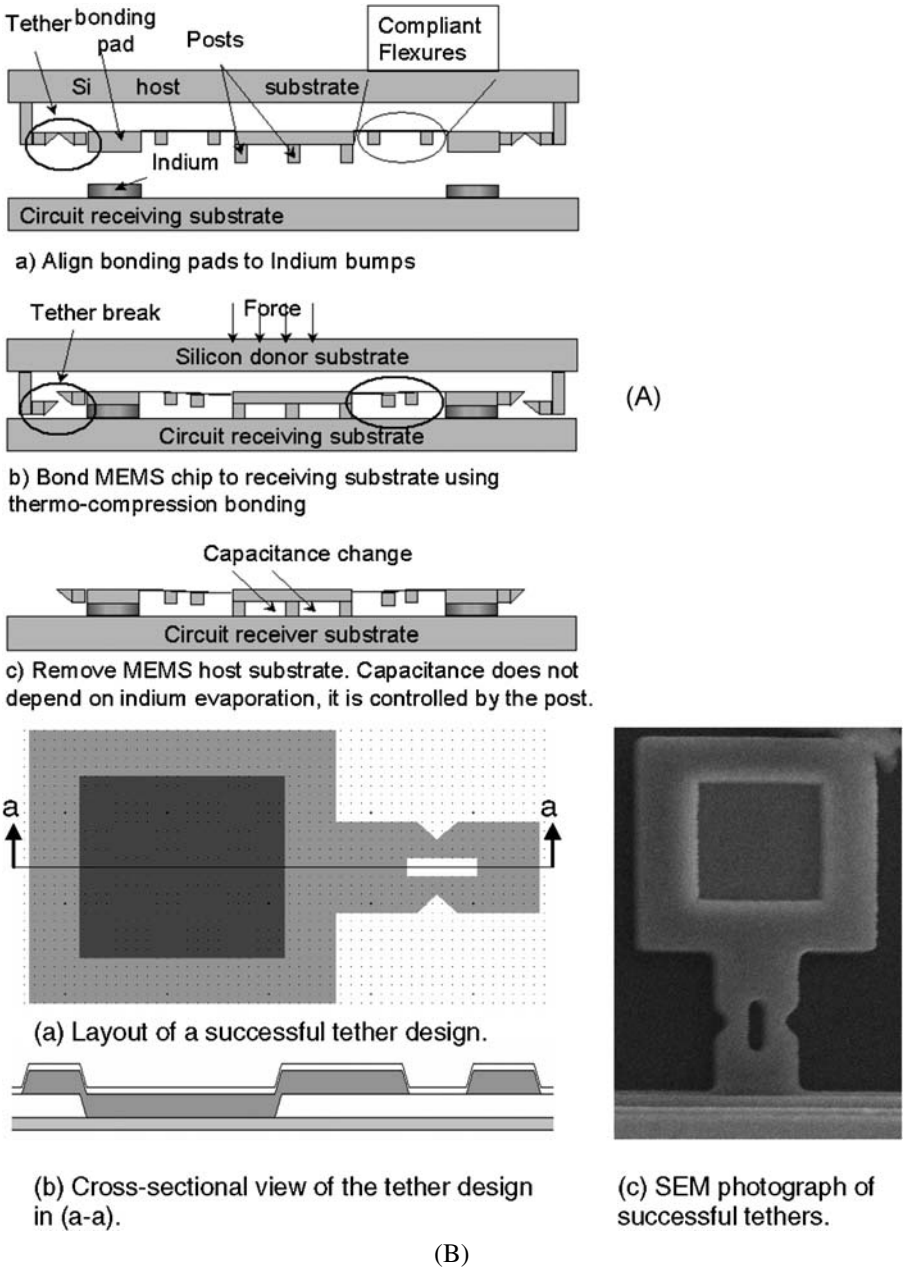
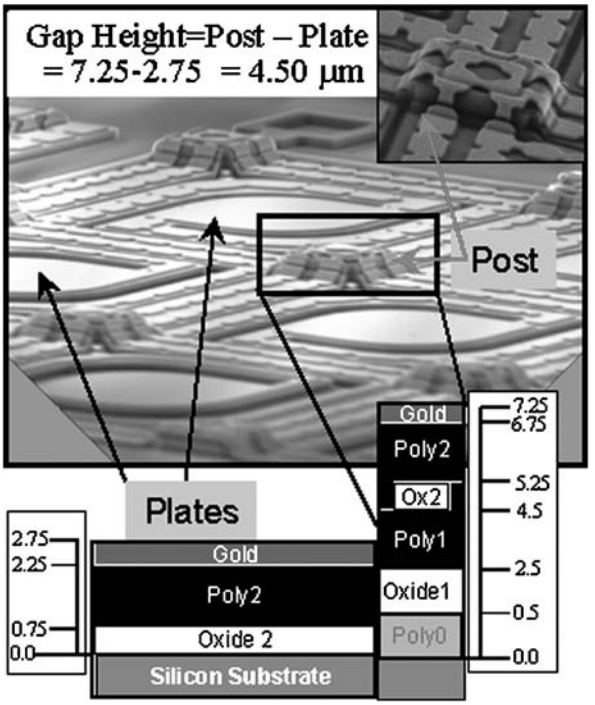


FIGURE 11.12. Illustration of the flip-chip assembly process. The upper piece shown in a) in (A) is a pre-released MEMS chip with gold pads; the lower piece shown in a) in (A) is a patterned ceramic substrate with deposited indium bumps. Tethers are used to hold pre-assembly released MEMS. Posts are used for the precision gap control after the assembly.

were designed with each plate surrounded by four "posts" (legs) to support the plate and its flexures. The corresponding flip-chip assembly process with tethers and posts is described in Figure 11.12, where the receiving substrate could be any circuit or RF substrate.



(C)

FIGURE 11.12. (Continued).

The use of tethers allowed us to transfer pre-assembly released MEMS devices onto a new substrate [26,27]. The tethers lightly connected a released MEMS device to its silicon donor substrate. They broke after delivering the device to the RF host substrate during or after the flip-chip assembly. A tether's design and photo are also shown in Figure 11.12. The use of posts enabled a precise gap control, which was critical to the operation of the capacitor plates. Before using posts, the gap was controlled by the solder joints. With evaporated indium, the gap height could vary up to ±25% [25]. To reduce such a variation, posts were created by stacking different layers during the design [15]. An example is shown in Figure 11.12. When the top plates were pulled down by the electrostatic force, each plate's pull-in voltage was controlled by the precise gap defined by the posts rather than the solder joints. In addition, posts also enabled us to design very compliant flexures to reduce thermal mismatch-induced warpage, which might degrade the electrostatic behavior of the MEMS by significantly increasing the pull-in voltage. With tethers and posts, the thermomechanical behavior of the 2D variable capacitors became controllable, and the desirable digital increments were demonstrated in the RF characterization [15].

11.3. SOLDERED ASSEMBLY FOR THREE-DIMENSIONAL MEMS

One of the most common methods for manufacturing MEMS devices is by using surface micro-machining, e.g., PolyMUMPs. Due to the nature of thin film deposition technology, a fundamental problem with surface micro-machining is its inability to produce highly

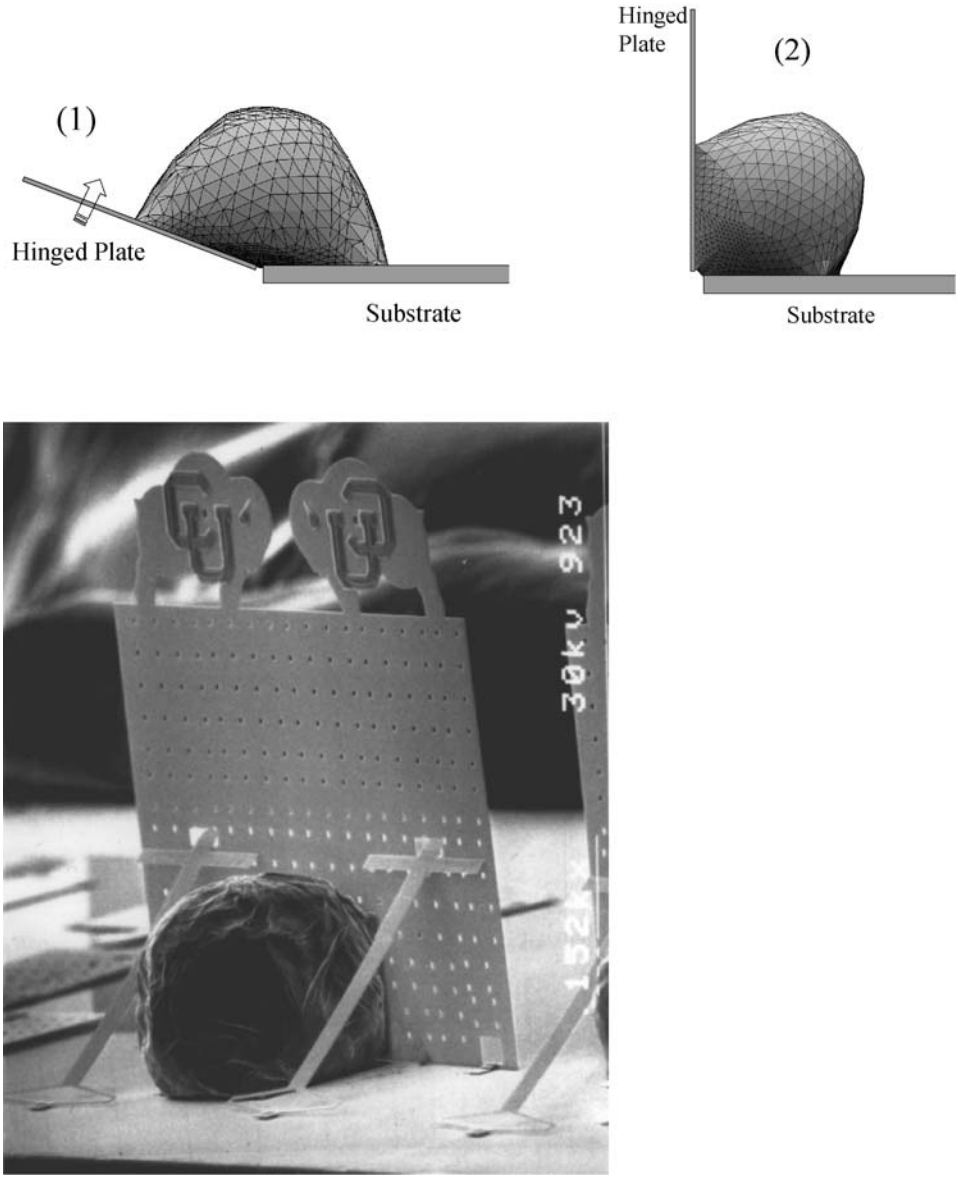


FIGURE 11.13. Illustration of solder self-assembly of a hinged MEMS plate and a solder assembled three-dimensional MEMS device with kickstands.

three-dimensional structures. A common solution is to fabricate flat, 2D hinged components that can be lifted or rotated into assembled structures [28]. Such structures are very common in many MEMS and microelectronics fields, namely micro-optics [29]. The draw back of hinged designs is that they need to be assembled after fabrication. The traditional way to perform this assembly is to do it manually or use additional MEMS mechanisms to assemble devices automatically [30,31]. Manual assembly usually consists of rotating the plates by hand using high precision micro-manipulators. This form of assembly is not

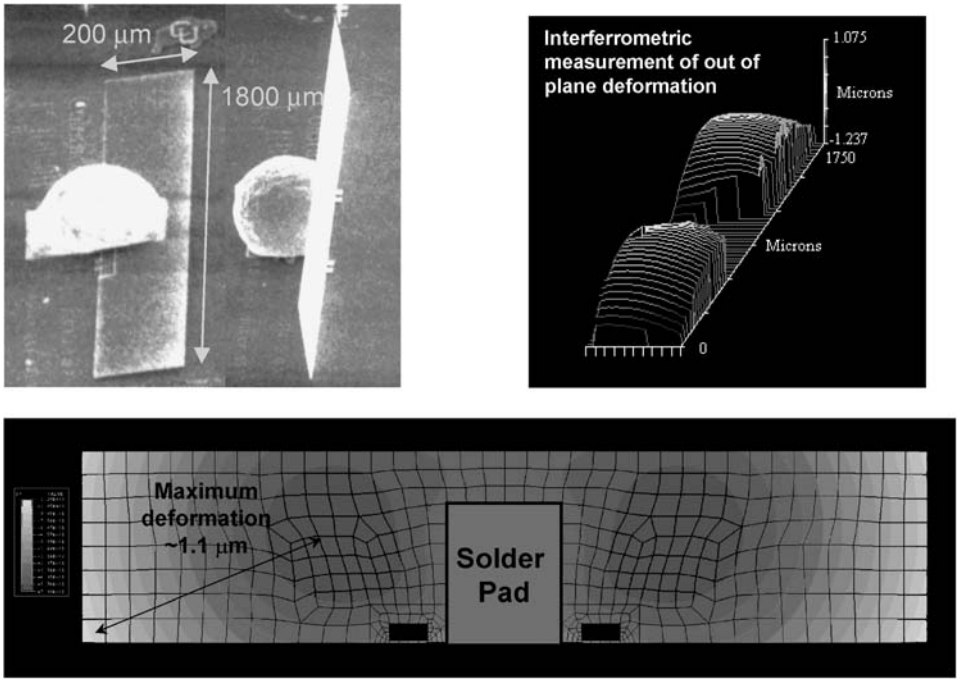


FIGURE 11.14. Validation example in which the model prediction (bottom) matches the interferometric measurement (top right).

practical for mass assembly and manufacturing though, and is rarely effective. Mechanism driven assembly is also insufficient because these MEMS mechanisms are often large and complex, and thus negate many advantages inherent in MEMS devices. An interesting solution is the use of the surface tension properties of molten solder or glass as the assembly mechanism [32–34].

The solder method involves using a standard hinged plate with a specific area metalized as solder wettable pads. Once the solder is in place, it is heated to its melting point, and the force produced by the natural tendency of liquids to minimize their surface energy pulls the free plate away from the silicon substrate (Figure 11.13). Solder is a predominant technology for electronics assembly and packaging. It is not only used for electrical connections, but also for sub-micron accuracy alignment in many packaging applications such as optoelectronic passive alignment [35]. Using solder, hundreds or thousands of precision alignments can be accomplished with a single batch reflow process, and the cost/alignment can be reduced by orders of magnitude. In addition, solder provides high quality mechanical, thermal, and electrical connections.

Figure 11.13 also shows an actual solder self-assembled plate that was 400 microns square and was assembled with an approximately 200 micron diameter solder sphere. But the plate was deformed due to process-induced stresses within the structure. Such thermomechanical deformations could be reduced by using finite element modeling and optimization algorithms.

The basic building block of a solder self-assembly structure consisted of a single solder sphere with a hinge and mechanical lock on either side (Figure 11.14). To optimize the structure, the parameters to be varied were: the contact position of the mechanical lock

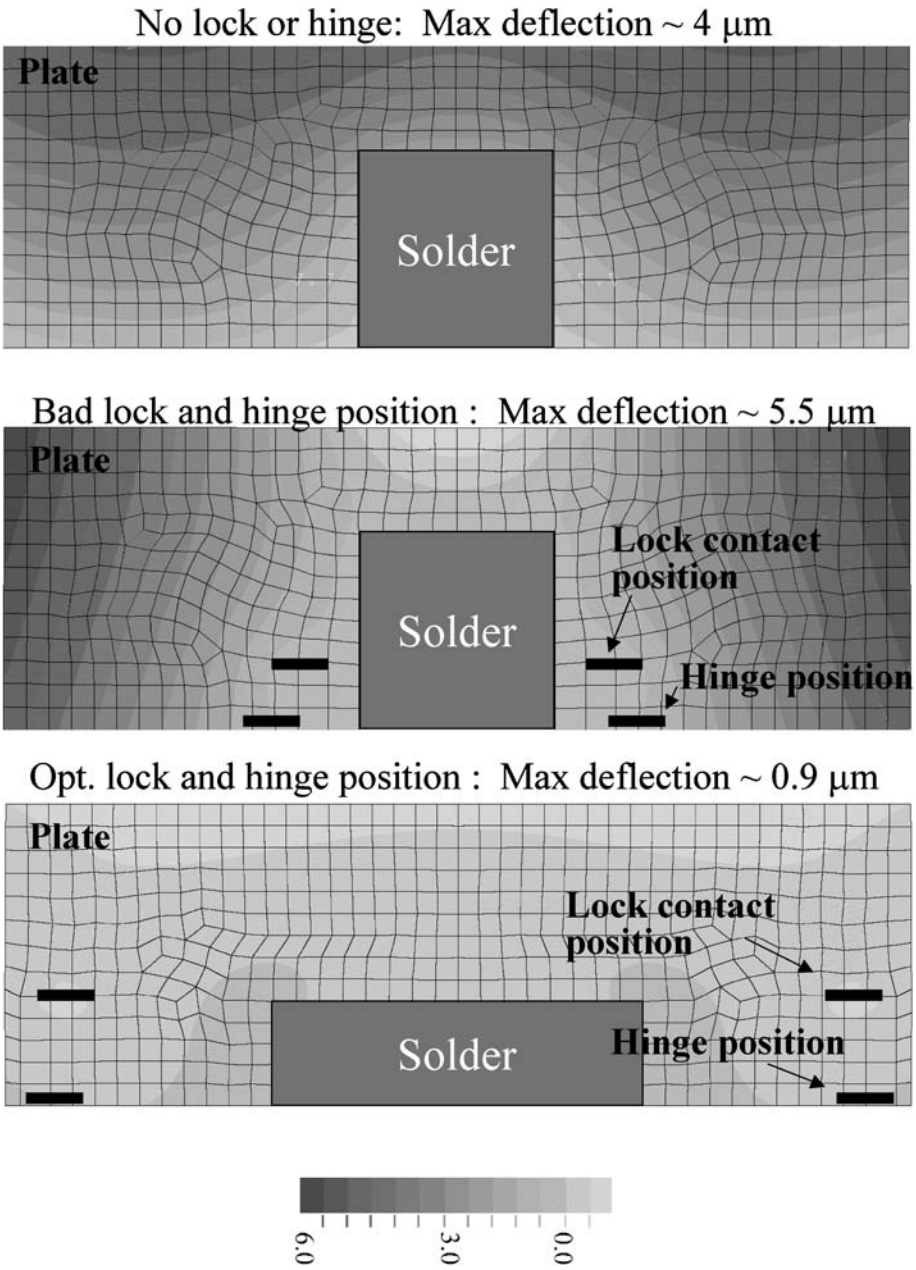


FIGURE 11.15. Comparison of results: (top) case without lock and hinges; (middle) case with lock and hinge poorly located; and (bottom) case with the optimum pad size and lock/hinge position.

and plate, the width and height of the solder pad, and the position of the hinge. The only constraint was that the solder pad should remain large enough to be practical for solder deposition. The finite element software ABAQUS was used to model the structure and extract relevant data, and the optimization algorithm NLPQL [36] was used to optimize

the variables. The plate was modeled using composite shell elements, but the solder was simulated with standard three-dimensional solid elements. The interaction of the kickstand and hinges was modeled using contact surface approximations rather than by including the actual hinge and kickstand structure into the model. The accuracy of the model was gauged by comparing the predictions to experimental data. Figure 11.14 shows one such comparison in which a 200 by 800 μm solder self-assembly plate was modeled, fabricated, and measured interferometrically. All cases that the model predicted fell within the data variation.

After validation, the optimization program was able to generate a prediction that significantly reduced the deformation in the plate. The values to be minimized were the rms, average, and maximum deflections of the plate. Figure 11.15 shows three sample cases for one design optimization problem: (a) a case in which there was no lock or hinge contact, (b) a case in which the lock contact position and hinge were poorly placed, and (c) the optimum case. Interestingly, the case in which the lock and hinge placed poorly resulted in a more severe deformation than with no lock at all. The poor lock and hinge position resulted in a maximum deflection of ∼5.5 μm and a rms deflection of ∼3.4 μm, whereas, the prediction with no lock or hinge resulted in a max deflection of ∼4 μm and an rms deflection of ∼2.1 μm. Finally, the optimized structure showed a significant improvement with a maximum deflection of ∼0.9 μm and rms deflection of ∼0.6 μm.

The reason for the reduced deformation is likely due to the lock and hinge constraints working against the deformation resulting from solder shrinkage. The shrinkage tends to cup the plate around it like a shroud. By placing the hinge and lock near the edge of the plates, they restrict the plate and force it back toward the desired position. If the lock and hinge are placed too close to the solder, they only amplify the deformation. If there are too far out, the plate will bend significantly between them and the solder joint. In addition to the plate deformation, the deviation from the desirable angles can also be controlled [28]. Advanced thermo-mechanical analysis and optimization techniques are essential to design such complex MEMS structures.

11.4. FLEXIBLE CIRCUIT BOARDS FOR MEMS

Silicon processing is not the only means to fabricate micro-scale devices. In fact, we expect to see more and more micro-devices to be fabricated using polymer materials. Here is an example for a flexible circuit-based RF MEMS [37]. Photographs of different layers and assembled prototype of X/Ku band switch designs are shown in Figure 11.16. Coplanar waveguide (CPW) lines for mounting switches and on-wafer multi-line TRL calibration were patterned on the metalization layer of a Duroid substrate. Photosensitive benzocyclobutene (BCB) dielectric layer was spin-coated and patterned on CPW lines. Adhesive spacer film was milled to create slot-openings. The switch electrode metalization was patterned on Kapton-E polyimide film, which was machined using Excimer laser to create slot-openings. These layers were aligned using a fixture and assembled using a thermo-compression bonding cycle.

These switches are manufacturable using printed circuit board (PCB) facilities, and they can be integrated with PCB-based RF circuits and antennas. We expect them to have an impact on cost-demanding applications. However, with the large size, there are concerns about their RF losses and new reliability failure mechanisms different from those of thin-film based RF MEMS. The insertion loss could be less than 0.3 dB and the isolation could

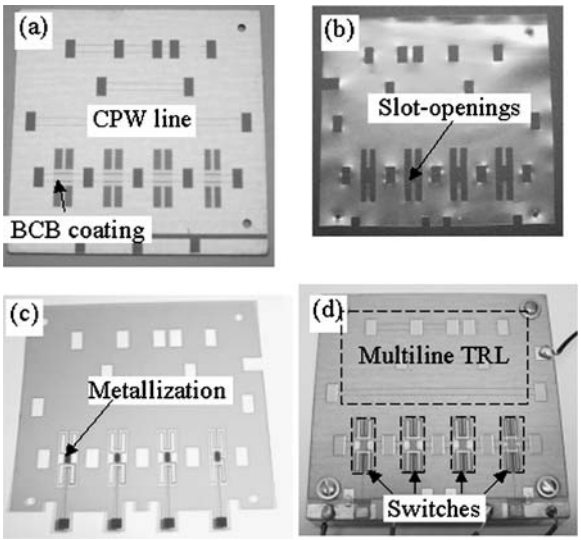


FIGURE 11.16. Photographs of (a) CPW line with BCB dielectric layer on Duroid substrate; (b) adhesive spacer film with milled slot-openings; (c) Kapton E polyimide film with switch top electrode metalization and laser machined slot-openings; and (d) assembled switch prototype.

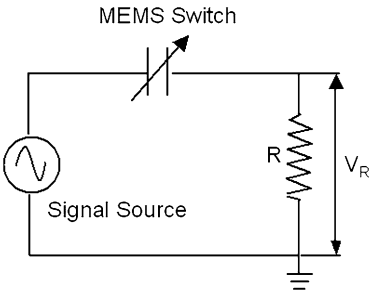


FIGURE 11.17. Circuit diagram of the reliability testing setup.

reach −50 dB at the designed frequencies [37]. Such performance is close to that achieved by thin-film based RF MEMS. The reliability has passed 75 millions of cyclic tests. This test is to be described in more details.

Figure 11.17 shows the circuit diagram for the reliability testing consists of the capacitive MEMS switch, i.e., the device under test (DUT), connected in series with a resistor. A function generator (Agilent 33250A) was used to generate the specified actuation waveform and the required amplitude was obtained by cascading the function generator with a power amplifier (Krohn-Hite 7600) [38,39]. The actuation waveform was applied to the switch and the voltage waveform across the resistor was the input to the data acquisition card (PCI-6035E, available from National Instruments) interfaced to a computer. The voltage waveform was recorded continuously using a LabVIEW program.

When the amplitude of the actuation waveform reached or exceeded the switch pull-down voltage, the switch capacitance changed from a low value (in up-position) to a high value (in down-position). On the other hand, the switch capacitance changed from a high

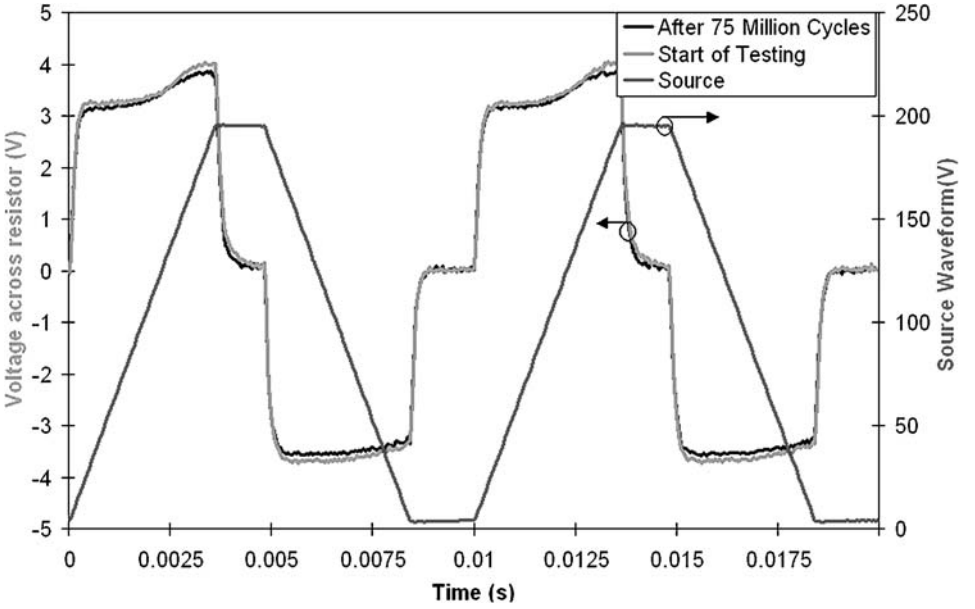


FIGURE 11.18. Reliability testing results showing the voltage waveform across the resistor after 75 millions of operations.

value (in down-position) to a low value (in up-position) when the amplitude of the waveform was less than the release voltage. This change in capacitance with time during the pull-down and release process caused a change in current passing through the test circuit. The voltage waveform across the resistor was proportional to the change in the current waveform and was observed to study the switch dynamic characteristics. Thus, this method could aid in studying the degradation and lifetime characteristics of the capacitive MEMS switches. In addition, switch failure due to mechanical failure, contact striction, fatigue, etc., could be investigated by analyzing the recorded waveform over time.

Switch reliability was studied by applying a triangular actuation waveform at a frequency and amplitude of 12 Hz and 200 V, respectively, for millions of continuous operations. The actuation waveform and the corresponding voltage waveform across the resistor were recorded simultaneously as shown in Figure 11.18. Switching speed of this switch was estimated to be in the millisecond range. This value was higher than those of other RF MEMS switches reported [20] and was due to a large switch up-position gap height of 50–70 μm compared to 2–4 μm in silicon based RF MEMS switches. A triangular wave at 12 Hz that had a rise time and fall time of 41.7 ms was used to ensure that there was enough time for the switch to respond to the actuation signal. The voltage waveform V_R was recorded every tenth of a million operations. The waveforms measured after 0.5 million and 0.7 million operations are also shown as dotted lines. These results coincide very well and indicate that the switch operated up to 1 million with no signs of degradation or failure. Such excellent responses survived after 75 millions of cyclic tests (see Figure 11.18).

Unfortunately, the switch failed right after 75 millions of cycles. After careful inspection, the failure was identified; it was caused by an electrical short resulting from the pinholes of the BCB dielectric layer. The polymer dielectric layer was not strong enough under 200 V across the 1-μm thickness. The RF MEMS device with reliability of 75 million

cycles is good enough for switching bandwidths in a cell phone. But, it should be improved further for other applications. The most viable approach to enhance the BCB strength is to deposit an inorganic alumina dielectric layer through atomic layer deposition (ALD) technology. This inorganic, pin-hole free dielectric layer is expected to solve this problem for a more reliable switch. ALD is to be described in the next section on MEMS reliability.

11.5. ATOMIC LAYER DEPOSITION FOR RELIABLE MEMS

During the last 20 years, MEMS reliability has been improved significantly [7,40,41]. Successful, reliable products have proven an important fact: mechanical devices can be switched over trillions of cycles while achieving the same reliability level as the electronic devices [7]. The contact-associated failures are strongly affected by the contact modes and materials, and the effects can be changed significantly if there are minute environmental variations due to particles, charges, and moistures. Currently, self-assembled monolayer (SAM) surface coating is used widely to protect MEMS devices from failures [42]. This organic coating layer, however, provides limited protection and has its own reliability problems. With the advancement of nano-technologies, we now have the opportunity to design and fabricate nano-scale protective coatings to assure MEMS reliability. One of such technologies is atomic layer deposition (ALD). ALD can coat thin dielectric layer to protect MEMS from electrical shorts during operations [43,44]. ALD can coat nano-scaled multilayers with conducting and dielectric materials for effective charge dissipation [45]. In addition, strong hydrophobic coating can be formed on the ALD coating to reduce moisture-induced adhesion even in a very high relative humidity environment. This technology will be introduced with an emphasis on the above-mentioned three reliability protection mechanisms.

Atomic Layer Deposition (ALD): ALD is a thin film growth technique allowing atomic-scale thickness control. ALD utilizes a binary reaction sequence of self-limiting chemical reactions between gas phase precursor molecules and a solid surface [45,46]. Films deposited by ALD are extremely smooth, pinhole-free and conformal to the underlying substrate surface. This conformity enabled successful coating to cover the entire MEMS device as shown in Figure 11.19. Furthermore, ALD is a low temperature process enabling deposition on thermally sensitive materials. For example, we can use photoresist to cover some patterned areas during deposition for selective instead of comprehensive coverage. ALD can be used to grow a variety of materials including oxides, nitrides, and metals.

Figure 11.20 illustrates the atomic layer deposition process. Reaction A deposits a monolayer of chemisorbed species on the surface. Because the resulting surface is inert to precursor A, further exposure generates no additional growth. Next, precursor B is introduced. This molecule reacts with the product surface from the A reaction in a self-passivating manner. Consequently, the B reaction terminates after the completion of one atomic layer. If reaction B regenerates the initial surface, then the two reactions can be repeated in an ABAB ... binary sequence to deposit a film of predetermined thickness.

One example of this process is the atomic layer deposition of Al_2O_3 consisting of the following binary reaction sequence in which the asterisks designate the surface species:

$$\text{(A) } Al{-}OH^* + Al(CH_3)_3 \rightarrow Al{-}O{-}Al(CH_3)_2^* + CH_4,$$

$$\text{(B) } Al{-}CH_3^* + H_2O \rightarrow Al{-}OH^* + CH_4.$$

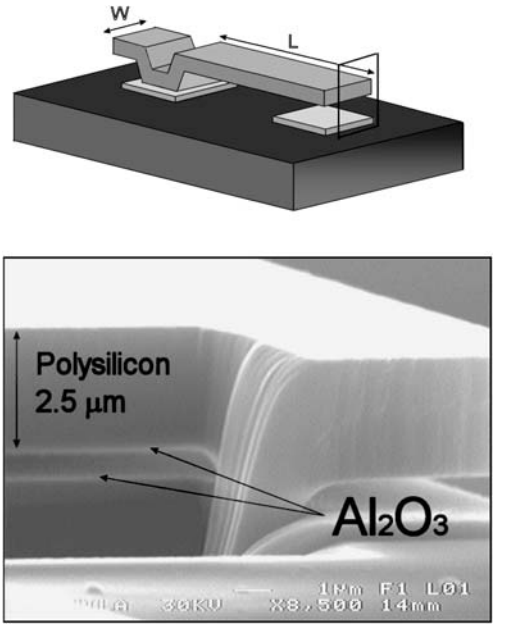


FIGURE 11.19. Illustration of beam and FIB cut section depicting deposited alumina layer.

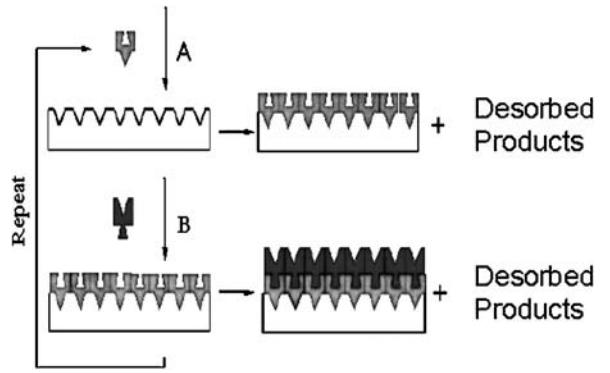


FIGURE 11.20. Description of ALD process.

In reaction (A), the $Al(CH_3)_3$ reacts with the surface hydroxyl groups to deposit a new monolayer of aluminum atoms terminated by methyl groups. In reaction (B), the methylated surface reacts with H_2O vapor, thereby replacing the methyl groups with hydroxyl groups. CH_4 is liberated in both the A and B reactions. The net result of one AB cycle is the deposition of one monolayer of Al_2O_3 onto the surface. The ALD Al_2O_3 film growth is extremely linear with the number of AB cycles performed and the growth rate is 1.29 Å/cycle. The deposition rate is about 0.12 nm (one AB cycle) in 6–10 seconds in a laboratory setup. In a manufacturing setup, the cycle time can be reduced by at least 10 times.

ALD can be used to coat many different nano-scaled, single-layer or multi-layer structures to protect MEMS from different reliability failures. In the following sections,

we will illustrate such protection with three examples associated with electrical shorts and charge-induced and moisture-induced adhesion failures.

Dielectric Coating to Prevent Electrical Shorts: Particles are the top killers to most of MEMS devices with or without surface contacts. As a result, MEMS devices should be packaged in a Class-10 environment, and the outgassing inside the package should be controlled by selecting right materials or using getters [17]. One of the particles-induced failures is shorting between conducting parts. A conformal layer of dielectric material coated can prevent this electrical short problem. As shown in Figure 11.19, a conformal layer of alumina (Al_2O_3), an excellent dielectric, was deposited onto released MEMS devices [43,44]. The ALD films cover all sides of a released MEMS device including bottom surfaces, such as underneath cantilever beams. This process was carried out at temperatures down to 150°C—significantly cooler than typical CVD temperatures. This allows for the coating of composite devices made from materials such as poly-silicon and gold without the risk of damaging the individual layers in the MEMS device. In addition, polymer based MEMS devices could also be coated at temperatures as low as 70°C. The deposition technique is compatible with integrated circuit devices as well as thermally sensitive packaged systems.

In addition to cantilever shown above, such a dielectric coating can be used for comb drive actuators and other sensors and actuators [47]. The conformable, selectable, nanoscaled coating can protect these devices from shorts caused by unexpected contacts or by particles.

Charge Dissipation for Reliable MEMS: Charge accumulation is the leading failure mode for RF MEMS switches with dielectric contacts. Figure 11.21 illustrates this charging effect: switch lifetime was about 10,000,000 cycles with 50 Volts applied, however, it reduced substantially to only 10,000 cycles when the voltage increased to 65 Volts [20]. The charge accumulation in the dielectric layer after cyclic loading with high electric fields is proportional to the voltage applied with an exponential function. The charge detrapping is governed by different mechanisms: Schottky emission, Frenkel-Pool emission, tunnel or field emission, space-charge emission, ohmic conduction, and ionic conduction. These mechanisms are very complicated and process- and material-dependent. One simple solution is to increase the effective electrical conductivity of the dielectric layer by doping. ALD can deposit a multilayer composite with hundreds of Al_2O_3 and ZnO layers. Figure 11.22 presents the resistivity values of Al_2O_3/ZnO multilayers with different Zn contents. The resistivity can be changed from 10^{16} to 10^{-3} ohm-cm by choosing a specific content [45]. This accurate control of the resistivity was proven critical to assure reliable RF MEMS switches [48].

Hydrophobic Coating for Reliable MEMS: MEMS reliability is seriously impaired by interfacial interactions. Humidity plays a key role in determining the character of interfacial adhesion. At high relative humidity, water capillary condensation in high aspect ratio micron-sized MEMS structures can cause striction and MEMS failure. To minimize water capillary condensation, the MEMS device can be coated with a hydrophobic film. These hydrophobic films are generally deposited in solution using chlorosilane attachment of alkylsilanes or perfluoroalkylsilanes onto surface hydroxyl groups [42]. Under optimum conditions, the attached alkylsilanes can form a self-assembled monolayer (SAM). However, SAM coating has its own processing challenges and reliability problems [49]. In order to achieve reliable SAM coating, ALD was used to deposit an alumina seed layer before ALD-SAM coating [50,51]. The seed layer could optimize the hydrophobic precursor attachment by: (1) covering the MEMS surface uniformly with a continu-

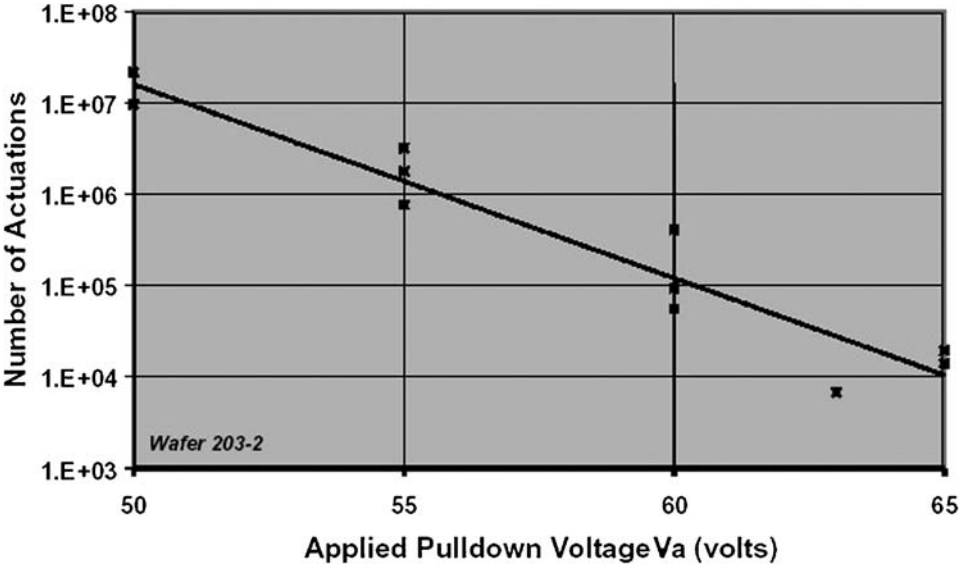


FIGURE 11.21. Lifetime as a function of applied voltage for a RF MEMS switch.

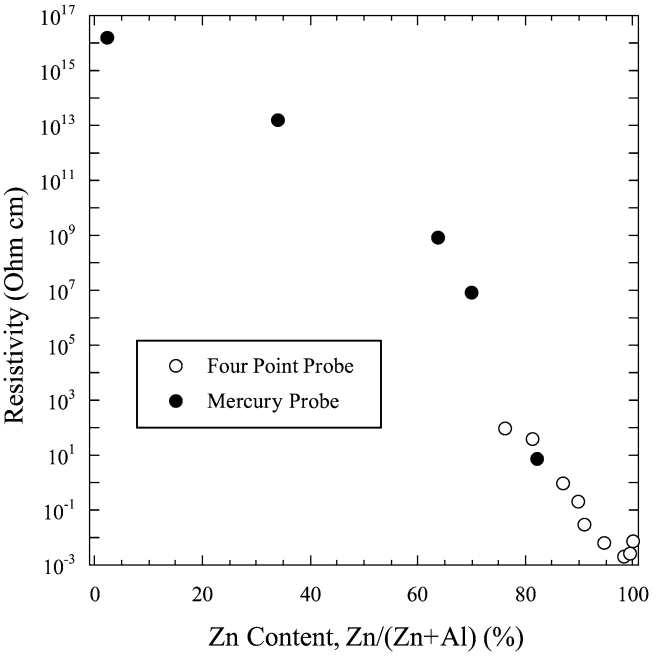


FIGURE 11.22. Electrical resistivity of ALD-coated ZnO/Al_2O_3.

ous adhesion layer; (2) providing a high surface coverage of hydroxyl groups for maximum precursor attachment; and (3) smoothing and removing nanometer-sized capillaries that may otherwise lead to moisture-induced striction problems. Additionally, polymerization was avoided by using alternative precursors, such as alkylaminosilanes, instead of the

traditional chlorosilanes. These alternative precursors reacted more completely and effectively with the surface hydroxyl groups without initial reaction with H_2O. Furthermore, non-chlorinated alkylaminosilanes would not produce HCl, a byproduct that might corrode metal surfaces. Tridecafluoro-1,1,2,2-tetrahydrooctylmethylbis(dimethylamino)silane (FOMB(DMA)S, $C_8F_{13}H_4(CH_3)Si(N(CH_3)_2)_2$) was chosen as the hydrophobic precursor. The hydrophobic film generated by ALD was proven to be more reliable [50]. Its effectiveness creates an opportunity to control the moisture-induced striction reliability problems and pave the way for non-hermetic MEMS packaging.

11.6. CONCLUSIONS

This chapter reviewed MEMS packaging and reliability. Wafer-level capping was introduced as a good example to develop a MEMS packaging technology that is compatible with microelectronic packaging. Such compatibility is essential to reduce packaging cost and size to meet the demands of large scale applications of MEMS. In addition, advanced studies were described with an emphasis on the flip-chip assembly to integrate MEMS with other components, solder assembled three-dimensional MEMS, flexible circuit-based MEMS, and atomic layer deposition for reliable MEMS. These studies are different from current practices focusing on wafer-level capping and packaging. They are good examples, however, to illustrate new packaging technologies being developed in the laboratories.

With advancement of MEMS technologies, hundreds of novel microsystems are demonstrated every year. The impressive insertion of various MEMS sensors and actuators in automobiles, cell phones, and biomedical applications represents only the beginning of a new era. MEMS can be further improved by integrating its micro-scale components with nano-scale devices, i.e., nano-electromechanical systems (NEMS). Integrated MEMS/NEMS will rival, and perhaps even surpass, the societal impact of integrated circuits (ICs). There are many opportunities for packaging engineers to make significant contributions and lead the advancement.

ACKNOWLEDGMENTS

Most of the studies reviewed in this paper were conducted at the University of Colorado—Boulder. The author would like to thank his colleagues: Professors Victor M. Bright, Steve M. George, Martin L. Dunn, and K.C. Gupta and former and current research associates/graduate students: F.F. Faheem, K.F. Harsh, R. Ramadoss, Simone Lee, N.D. Hoivik, J.W. Elam, C.F. Herrmann, F.W. DelRio, and A. Laws. The author is partially supported by a project sponsored by the DARPA (Chip-scale atomic clock program) and managed by the Department of Interior (NBCH1020008).

REFERENCES

1. D. Koester, A. Cowen, R. Mahadevan, M. Stonefield, and B. Hardy, PolyMUMPs Design Handbook, Revision 9, MEMSCAP, San Jose, CA, USA.
2. E.J. Garcia and J.J. Sniegowski, Surface micromachined microengine, Sensors and Actuators A, 48, pp. 203–214 (1995).
3. C.M. Roberts, Jr., L.H. Long, and P.A. Ruggerio, Method for separating circuit dies from a wafer, U.S. Patent #5,362,681, Nov. 1994.

4. T.W. Kenny, R.N. Candler, H.J. Li, W.T. Park, J. Cho, H. Li, A. Partridge, G. Yama, and M. Lutz, An integrated wafer-scale packaging process for MEMS, Proc. ASME International Mechanical Engineering Congress & Exposition, New Orleans, Louisiana, November 17–22, 2002.
5. K. Petersen, Bringing MEMS to Market, Proc. Solid-State Sensor and Actuator Workshop, Hilton Head Island, South Carolina, June 4–8, 2000, pp. 60–64.
6. L.J. Hornbeck, From cathode rays to digital micromirrors: a history of electronic projection display technology, TI Technical Journal, (July–September), pp. 7–46 (1998).
7. Sontheimer and M. Douglass, Identifying and eliminating digital light processing TM failure modes through accelerated stress testing, TI Technical Journal, (July–September), pp. 128–136 (1998).
8. J.P. O'Connor, Packaging design considerations and guidelines for the digital micromirror device, Proc. of IPACK'01, The Pacific Rim/ASME International Electronic Packaging Technical Conference and Exhibition, Kauai, Hawaii, July 8–13, 2001.
9. G.A. Riley, Wafer-level hermetic cavity packaging, Advanced Packaging, 13(5), pp. 21–24 (2004).
10. D. Sparks, et al., Reliable vacuum packaging using NanoGettersTM and glass frit bonding, reliability, testing and characterization of MEMS/MOEMS III, Proc. SPIE, 5343 (2004).
11. Y.T. Cheng, L. Lin, and K. Najafi, Localized silicon fusion and eutectic bonding for MEMS fabrication and packaging, J. Microelectromech. Syst., 9, pp. 3–8 (2000).
12. L.E. Felton, M. Duffy, N. Habluzel, P.W. Farrell, and W.A. Webster, Low-cost packaging of inertial MEMS devices, International Symposium on Microelectronics, Proc. SPIE, 5288, pp. 402–406 (2003).
13. L.E. Felton, N. Hablutzel, W.A. Webster, and K.P. Harney, Chip scale packaging of a MEMS accelerometer, Proc. 54th Electronic Components and Technology Conference, 2004, pp. 869–873.
14. R.N. Candler, W.-T. Park, H. Li, G. Yama, A. Partridge, M. Lutz, and T.W. Kenny, Single wafer encapsulation of MEMS devices, IEEE Transactions on Advanced Packaging, 26(3), pp. 227–232 (2000).
15. F.F. Faheem and Y.C. Lee, Tether- and post-enabled flip-chip asssembly for manufacturable RF-MEMS, Sensors and Actuators, A-114(2-3), pp. 486–495 (2004).
16. F.F. Faheem, K.C. Gupta, and Y.C. Lee, Flip-chip assembly and liquid crystal polymer encapsulation for variable MEMS capacitors, IEEE Transactions on Microwave Theory and Techniques, pp. 2562–2567 (2003).
17. A. Tseng, W.C. Tang, Y.C. Lee, and J. Allen, NSF 2000 workshop on manufacturing of micro-electromechanical systems, Journal of Materials Processing & Manufacturing Science, 8(4), pp. 292–360 (2001).
18. R. Maboudian, W.R. Ashurst, and C. Carraro, Tribological challenges in micromechanical systems, Tribology Letters, 12, pp. 95–100 (2002).
19. D.M. Tanner, Reliability of surface micromachined MicroElectroMechanical actuators, 22nd Int. Conf. Microelectronics, Nis, Yugoslavia, 2000, pp. 97–104.
20. C. Goldsmith, et al., Lifetime characterization of capacitive RF MEMS switches, 2001 IEEE MTT-S International Microwave Symposium Digest, May 2001, pp. 227–230.
21. Y.C. Lee, B. McCarthy, J. Diao, Z. Zhang, and K.F. Harsh, Computer-aided design for microelectromechanical systems (MEMS), International J. of Nano Technology, 18(4/5/6) (2003).
22. Z. Feng, H. Zhang, W. Zhang, B. Su, K.C. Gupta, V.M. Bright, and Y.C. Lee, MEMS-based variable capacitor for millimeter-wave applications, Sens. Actuator A Phys., 91, pp. 256–265 (2001).
23. M.A. Michalicek and V.M. Bright, Flip-chip fabrication of advanced micromirror arrays, The 14th IEEE Int. Conf. Microelectromech. Syst., Jan. 21–25, 2001, pp. 313–316.
24. N. Hoivik, Y.C. Lee, and V.M. Bright, Flip-chip variable high-Q MEMS capacitor for RF spplications, ASME InterPACK'01, Kauai, Hawaii, USA, July 8–13, 2001.
25. N. Hoivik, M.A. Michalicek, Y.C. Lee, K.C. Gupta, and V.M. Bright, Digitally controllable variable high-Q MEMS capacitor for RF applications, IEEE Int. MTT-S, Phoenix, AZ, USA, May 20–25, 2001, pp. 2115–2118.
26. D.A. Singh, M.B. Horsley, A. Cohn, P. Pisano, and R.T. Howe, Batch transfer of microstructures using flip-chip solder bonding, J. Microelectromech. Syst., 8(3), pp. 27–33 (1999).
27. J.Y. Chen, L.S. Huang, C.H. Chu, and C. Peizen, A new transferred ultra-thin silicon micropackaging, J. Micromech. Microeng., 12, pp. 406–409 (2002).
28. K.F. Harsh, V.M. Bright, and Y.-C. Lee, Design optimization of surface micro-machined self-assembled MEMS structures, The ASME International, Intersociety Electronic & Photonic Packaging Conference & Exhibition (InterPACK'01), Kauai, Hawaii, July 8–13, 2001.
29. N.C. Tien, M. Kiang, M.J. Daneman, O. Solgaard, K.Y. Lau, and R.S. Muller, Actuation of polysilicon surface-micro-machined mirrors, Proc. SPIE, 2687, p. 27 (1996).
30. T. Akiyama, D. Collard, and H. Fujita, Scratch drive actuator with mechanical links for self-assembly of three-dimensional MEMS, Journal of Micro-Electromechanical Systems, 6(1), pp. 10–17 (1997).

31. L. Fan, M.C. Wu, and K.D. Choquette, Self assembled micro-actuated XYZ stages for optical scanning and alignment, Transducers 97: 1997 International Conference on Solid-State Sensors and Actuators, Chicago, June 16–19, 1997.
32. K.F. Harsh, V.M. Bright, and Y.C. Lee, Solder self-assembly for three-dimensional micro-electromechanical systems, Sensors and Actuators A, 77, pp. 237–244 (1999).
33. P.W. Green, R.R.A. Syms, and E.M. Yeatman, Demonstration of three-dimensional microstructure self-assembly, Journal of Micro-electromechanical Systems, 4(4), pp. 170–176 (1995).
34. R.R.A. Syms, Rotational self-assembly of complex microstructures by surface tension of glass, Sensors and Actuators A, 65, pp. 238–243 (1998).
35. Q. Tan and Y.C. Lee, Soldering for optoelectronics packaging, IEEE Electronic Components and Technology Conference, Orlando, FL, May 28–30, 1996, p. 26.
36. K. Schittkowski, NLPQL: A Fortran subroutine for solving constrained nonlinear programming problems, Annals of Operations Research, 5, pp. 485–500 (1985/86).
37. R. Ramadoss, S. Lee, V.M. Bright, Y.C. Lee, and K.C. Gupta, Polyimide film based RF MEMS capacitive switches, 2002 IEEE/MTT-S International Microwave Symposium—MTT 2002, 2–7 June 2002, Seattle, WA, IEEE MTT-S CDROM, 2002, pp. 1233–1236.
38. S. Lee, R. Ramadoss, K.C. Gupta, Y.C. Lee, and V.M. Bright, Reliability testing of flexible circuit-based RF MEMS capacitive switches, Microelectronics Reliability, 44, pp. 245–250 (2004).
39. Advanced Design System 2001, Agilent Technologies, CA, USA.
40. MEMS Industry 2004 Report Focus on Reliability, MEMS Industry Group, Pittsburgh, PA, USA.
41. D.M. Tanner, Reliability of surface micromachined MicroElectroMechanical actuators, 22nd Int. Conf. Microelectronics, Nis, Yugoslavia, 2000, pp. 97–104.
42. R. Maboudian, W.R. Ashurst, and C. Carraro, Tribological challenges in micromechanical systems, Tribology Letters, 12, pp. 95–100 (2002).
43. N.D. Hoivik, J.W. Elam, R.J. Linderman, V.M. Bright, S.M. George, and Y.C. Lee, Atomic layer deposited protective coatings for micro-electromechanical systems, Sensors and Actuators, A-103, pp. 100–108 (2003).
44. N. Hoivik, J. Elam, S. George, K.C. Gupta, V.M. Bright, and Y.C. Lee, Atomic layer deposition (ALD) technology for reliable RF MEMS, 2002 IEEE/MTT-S International Microwave Symposium—MTT 2002, 2–7 June 2002, Seattle, WA, IEEE MTT-S CDROM, 2002, pp. 1229–1232.
45. J.W. Elam and S.M. George, Growth of ZnO/Al_2O_3 alloy films using atomic layer deposition techniques, Chem. Mater., 15, p. 1020 (2003).
46. S.M. George, A.W. Ott, and J.W. Klaus, Surface chemistry for atomic layer growth, Journal of Physical Chemistry, (100), pp. 13121–13131 (1996).
47. J.J. Yao, RF MEMS from a device perspective, J. Micromech. Microeng., 10(4), pp. R9–R38 (2000).
48. F.W. DelRio, C.F. Herrmann, N. Hoivik, S.M. George, V.M. Bright, J.L. Ebel, R.E. Strawser, R. Cortez, and K.D. Leedy, Atomic layer deposition of Al2O3/ZnO nano-scale films for gold RF MEMS, IEEE MTT-S International, Volume 3, 6–11 June 2004, pp. 1923–1926.
49. M.P. de Boer, J.A. Knapp, T.A. Michalske, U. Srinivasan, and R. Maboudian, Adhesion hysteresis of silane coated microcantilevers, Acta Mater., 48, pp. 4531–4541 (2000).
50. C.F. Herrmann, F.W. DelRio, V.M. Bright, and S.M. George, Hydrophobic coatings using atomic layer deposition and non-chlorinated precursors, 17th IEEE International Conference on MEMS, 2004, pp. 653–656.
51. U. Srinivasan, M.R. Houston, and R.T. Howe, Alkyltrichlorosilane-based self-assembled monolayer films for stiction reduction in silicon micromachines, J. MEMS, 7, p. 252 (1998).

12

Advances in Optoelectronic Methodology for MOEMS Testing

Ryszard J. Pryputniewicz

NEST—NanoEngineering, Science, and Technology, CHSLT—Center for Holographic Studies and Laser Micro-MechaTronics, Mechanical Engineering Department, Worcester Polytechnic Institute, Worcester, MA 01609, USA

Abstract Continued demands for delivery of high performance micro-optoelectromechanical systems (MOEMS) place unprecedented requirements on methods used in their development and operation. Metrology is a major and inseparable part of these methods. Optoelectronic methodology is an essential field of metrology that facilitates development of MOEMS because of its inherent advantages over other methods currently available. Due to its scalability, optoelectronic methodology is particularly suitable for testing of MOEMS where measurements must be made with ever increasing accuracy and precision. This was particularly evident during the last few years, characterized by miniaturization of devices, when requirements for measurements have rapidly increased as the emerging technologies introduced new products, especially, optical MEMS. In this chapter, a novel optoelectronic methodology for testing of MOEMS is described and its application is illustrated with representative examples. These examples demonstrate capability to measure submicron deformations of various components of the micromirror device, under actual operating conditions, and show viability of the optoelectronic methodology for testing of MOEMS.

12.1. INTRODUCTION

Advances in technology are frequently based on miniaturization of electronics while simultaneously increasing their capabilities and reducing cost. These advances have led to development of microelectromechanical systems (MEMS). Now, MEMS defines both the technologies to make these systems and the systems themselves.

One of the systems that were made possible by the MEMS technology is a micromirror system for optical applications [1]. This microsystem is a part of a group of micro-optoelectromechanical systems (MOEMS). MOEMS, fabricated using silicon and polysilicon micromachining processes, have widespread applications [2–5], including, but not limited to, optical beam steering, scanners, adaptive optical arrays, flat panel displays, optical interconnects, etc. Specific advantages that MOEMS have over their larger, conven-

tional, counterparts are lower mass, faster response speeds, lower operating power, compact design, and the potential for large arrays of micro-optical elements [6,7].

Development of MEMS, including MOEMS, and structures they interact with, requires sophisticated design, analysis, fabrication, testing, and characterization tools. These tools can be categorized as analytical, computational, and experimental. Solutions using the tools from any one category alone do not usually provide necessary information on MEMS/MOEMS and extensive merging, or hybridization, of the tools from different categories is used [8–10]. One of the approaches employed in the development of MEMS/MOEMS, as well as other complex structures of current interest, is based on a combined use of analytical, computational, and experimental solutions (ACES) methodology [11–13]. In fact, ACES methodology provides solutions where they would not otherwise be possible, or at best be difficult to obtain, while using either only analytical, or only computational, or only experimental tools alone.

In general, analytical tools are based on exact, closed form solutions. These solutions, however, are usually applicable to simple geometries for which, boundary, initial, and loading (BIL) conditions can be readily specified. Analytical solutions are indispensable to gain insight of overall representation of the ranges of the anticipated results. They also facilitate determination of the "goodness" of the results based on the uncertainty analyses. Computational tools, i.e., finite element methods (FEMs), boundary element methods (BEMs), and finite difference methods (FDMs), provide approximate solutions as they discretize the domain of interest and the governing partial differential equations (PDEs). The characteristics of discretization, in conjunction with the corresponding BIL conditions, influence degree of approximation and careful convergence studies [14] should be performed to establish correct computational solutions and modeling. It should be noted that both analytical and computational solutions depend on material properties. If material properties are well known, then solutions will give correct results, providing convergence was achieved subject to properly specified BIL conditions; if material properties are not well known, in spite of having a good knowledge of other modeling parameters, erroneous results will be obtained [15]. Experimental tools, however, in contrast to analytical and computational tools, evaluate actual objects, subjected to actual operating conditions, and provide ultimate results characterizing the objects being investigated. The experimental tools, used in this chapter, employ recent advances in optoelectronic laser interferometric microscope (OELIM) methodology [16,17].

In this chapter, hinged micromirror devices, actuated by electrostatically driven microengines, are considered, as detailed in Section 12.2.

12.2. MOEMS SAMPLES

The MOEMS considered in this chapter are micromirror devices, actuated by electrostatically driven microengines, Figure 12.1. These microsystems were fabricated at Sandia National Laboratories (SNL) using Sandia's Ultraplanar MEMS Multilevel Technology [18] (SUMMiT™ V). The entire micromirror system is made of polysilicon by surface micromachining. The process does not rely on assembly of the microsystem out of separately fabricated pieces, but produces the finished device by batch fabrication [19]. That is, at the end of the fabrication process, the microsystem is ready for use, e.g., in an optical interconnection application, Figure 12.2.

The SUMMiT™ process is based on a set of specific design tools [20,21]. These tools have been developed and validated for use with the multilayer surface micromachining,

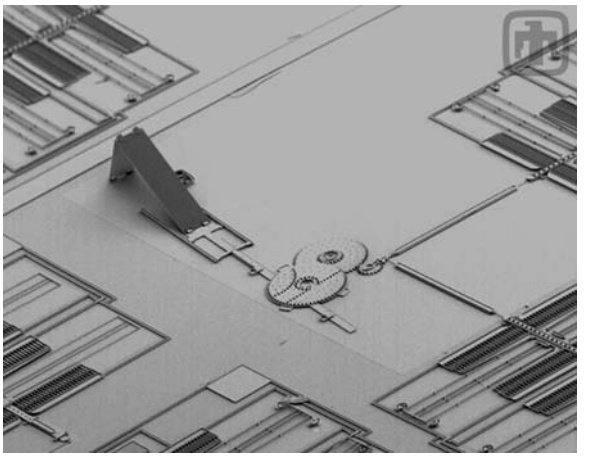


FIGURE 12.1. Sandia micromirror device actuated by electrostatically driven microengine.

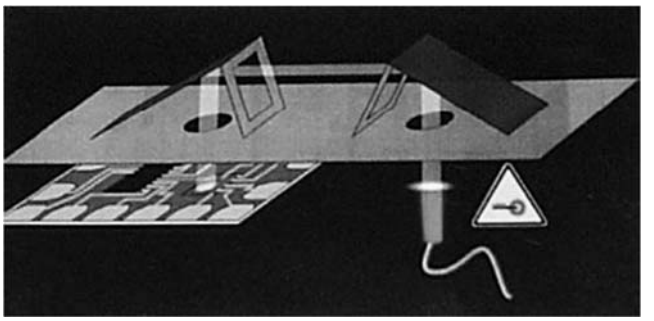


FIGURE 12.2. Interconnection concept based on the Sandia micromirror device.

Figure 12.3. Their use is facilitated by availability of standard components library. For example, to design a hinged micromirror, *Optical Components* are pulled down from the *Components Library*, Figure 12.4. Then, specific components, e.g., anchor (i.e., ground) hinge, is selected, Figure 12.5. The library component shown in Figure 12.5 contains all design details, Figure 12.6, necessary to fabricate a functional hinge using the SUMMiT™ process. Integrating other optical components, available in the library, a hinged micromirror of desired/specific dimensions can be designed, Figure 12.7, and fabricated, Figure 12.8. The micromirror integrated with other parts forms the microsystem, Figure 12.1.

The microsystem is fabricated using surface micromachining of multiple (structural) polysilicon films with intervening (sacrificial) oxide films, Figure 12.3. All structural films are made using low-pressure chemical vapor deposition (LPCVD) of structural polycrystalline silicon. The sacrificial silicon-dioxide films are also deposited by LPCVD. Fabrication of the micromirror system, including the electrostatic comb drives, gears, and the interconnecting linkages, requires four polysilicon structural-layers; in the SUMMiT™ V process up to 5 structural polysilicon layers are available if needed. In this process, the first polysilicon layer (POLY0) provides a voltage reference plane and electrical interconnections, while the remaining three (or four, as the particular case may be) polysilicon layers

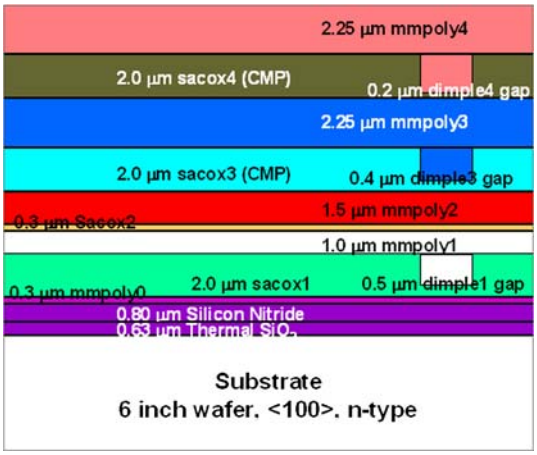


FIGURE 12.3. Film stack of the SUMMiT™ process.

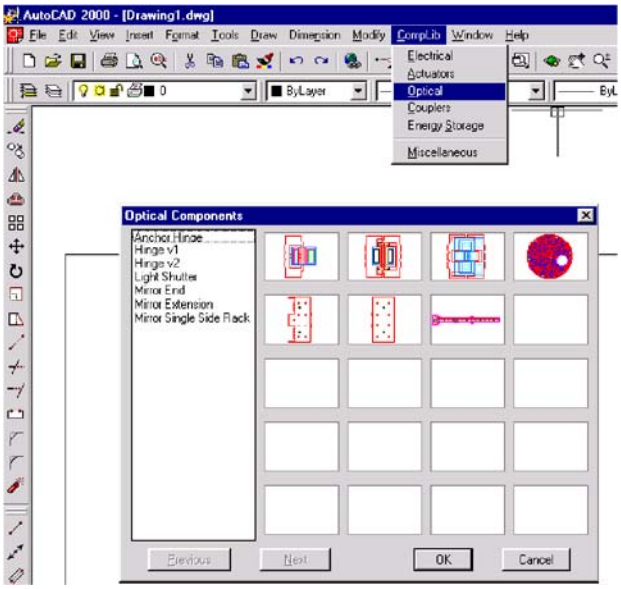


FIGURE 12.4. Optical components library of the SUMMiT™ design tools.

(i.e., POLY1, POLY2, POLY3, and POLY4) are used to form the mechanical/structural components.

In the micromirror device, the microengine converts electrical energy to kinetic energy. This microengine is controlled by two mutually orthogonal linear comb drive actuators. These comb drives consist of two sets of fingers, one stationary and the other movable. At rest, the fingers are in as fabricated position. When a voltage is induced, the movable set of fingers is attracted toward the stationary set thus producing a motion with respect to the base (or reference) while deforming the folded elastic suspension springs supporting the comb drives. When the voltage is suppressed, the elastic forces produced by the elastic springs, which are integral parts of the comb drives, restore any deflections, or movements,

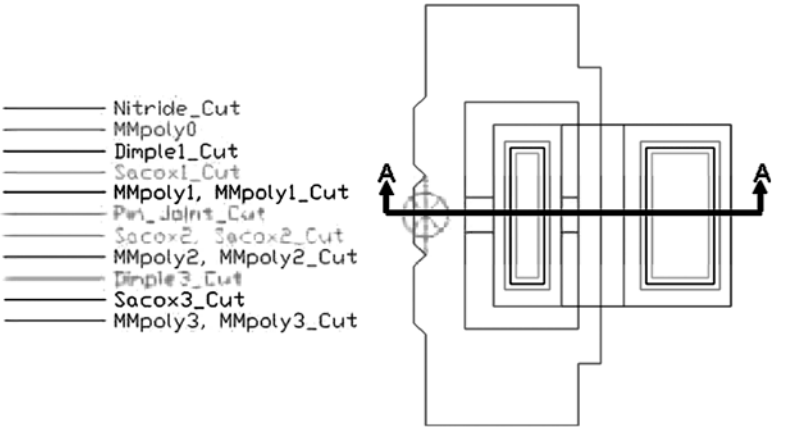


FIGURE 12.5. Anchor hinge as selected from the standard components library. Detailed cross section along A–A is shown in Figure 12.6.



FIGURE 12.6. 2D visualization of details of cross section along the line A–A of Figure 12.5.

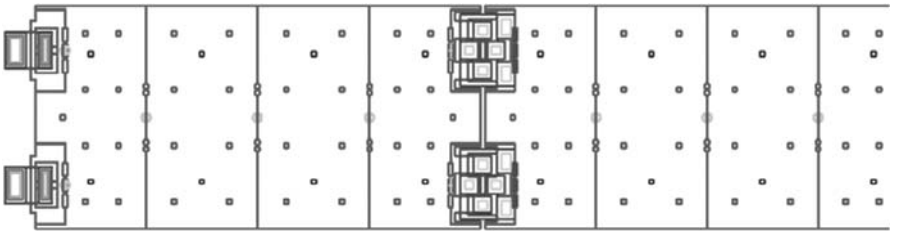


FIGURE 12.7. Hinged micromirror designed from standard components in the library of the SUMMiT™ design tools.

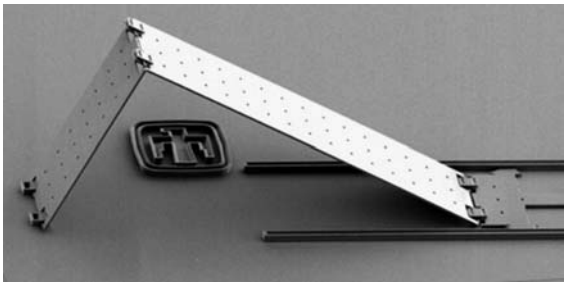


FIGURE 12.8. Hinged micromirror fabricated using the SUMMiT™ process based on the design shown in Figure 12.7.

of the actuator. The comb drives are connected to linkage arms. The linkage arms, in turn, are connected to the drive gear via pins. The drive gear is about 50 μm in diameter with perfectly formed teeth.

To understand dynamics of the micromirror device, deformations of its various components should be measured as a function of operational speeds. These deformations should then be correlated with parameters defining kinematics and kinetics of various components of the micromirror device, which are based on analysis discussed in Section 12.3.

12.3. ANALYSIS

An initial goal of the analysis is to determine accelerations of all moving parts of the microengine [22,23]. Then, using Newton's Second Law, calculate dynamic forces acting on the microengine. Once the dynamic forces are known, we can determine whether the microengine will perform as anticipated under expected operating conditions, or not?

Dynamic forces are based on accelerations, both linear and angular. In order to calculate accelerations we must first determine positions of all components in the microsystem for each increment of the input motion. Once equations defining positions are known, we differentiate them with respect to time to calculate velocities, and then differentiate again to obtain accelerations.

One way to develop equations defining positions of components is to write vector-loop-equations (VLEs) with reference to the kinematic diagram of the microengine [22]. A VLE starts at a specific point on the microengine and follows a loop, via other characteristic points, to end up at the point where it started. That is, the magnitude between the start and the stop points of a given VLE is zero, or a null vector. Thus, because of the nature of the microengine, two VLEs completely define its kinematics, based on which the corresponding equations describing displacements, velocities, and accelerations of characteristic points of the microsystem can be determined as functions of time [22]. That is, the equations defining linear and angular positions are

$$\theta_6 = \cos^{-1}\left[\left(\frac{R_2}{R_6}\right)\cos\theta_2\right], \tag{12.1}$$

$$R_5 = R_2\sin\theta_2 - R_6\sin\theta_6, \tag{12.2}$$

$$\theta_4 = \sin^{-1}\left(\frac{R_2\sin\theta_2 - R_3\sin\theta_6 - R_8}{R_4}\right), \tag{12.3}$$

and

$$R_7 = R_2\cos\theta_2 - R_3\cos\theta_6 - R_4\cos\theta_4, \tag{12.4}$$

where R and θ denote linear and angular positions, respectively, of linkages which are identified by subscripts. In Equation (12.4),

$$\theta_2 = \theta_{20} + \omega_2 t \tag{12.5}$$

with the subscript 0 denoting initial, i.e., at $t = 0$, angular position, and the angular speed ω_2 is defined as

$$\omega_2 = \frac{2\pi}{60} N \tag{12.6}$$

in which N is the *rotational speed* in *revolutions-per-minute* (rpm). It should be noted that one of the programs at SNL was dedicated to development of a microengine capable of sustained operation at 1,000,000 (i.e., one-million) rpm!

Using Equations (12.1) to (12.4), linear and angular velocities, V and ω, respectively, can be determined to be

$$\omega_6 = \left(\frac{R_2}{R_6}\right)\omega_2 \frac{\sin\theta_2}{\sin\theta_6}, \tag{12.7}$$

$$V_5 = R_2\omega_2 \cos\theta_2 - R_6\omega_6 \cos\theta_6, \tag{12.8}$$

$$\omega_4 = \frac{R_2}{R_4}\omega_2 \frac{\cos\theta_2}{\cos\theta_4} - \frac{R_3}{R_4}\omega_6 \frac{\cos\theta_6}{\cos\theta_4}, \tag{12.9}$$

and

$$V_7 = -R_2\omega_2 \sin\theta_2 + R_3\omega_6 \sin\theta_6 + R_4\omega_4 \sin\theta_4. \tag{12.10}$$

Based on Equations (12.7) to (12.10), equations for linear and angular accelerations, a and α, respectively, become

$$\alpha_6 = \frac{R_2}{R_6}\omega_2^2 \frac{\cos\theta_2}{\sin\theta_6} - \omega_6^2 \frac{\cos\theta_6}{\sin\theta_6}, \tag{12.11}$$

$$\alpha_4 = \frac{1}{R_4 \cos\theta_4}\left(-R_2\omega_2^2 \sin\theta_2 + R_3\omega_6^2 \sin\theta_6 - R_3\alpha_6 \cos\theta_6 + R_4\omega_4^2 \sin\theta_4\right), \tag{12.12}$$

$$a_5 = -R_2\omega_2^2 \sin\theta_2 + R_6\omega_6^2 \sin\theta_6 - R_6\alpha_6 \cos\theta_6, \tag{12.13}$$

and

$$a_7 = -R_2\omega_2^2 \cos\theta_2 - R_3\omega_6^2 \cos\theta_6 + R_3\alpha_6 \sin\theta_6 + R_4\omega_4^2 \cos\theta_4 + R_4\alpha_4 \sin\theta_4. \tag{12.14}$$

Kinetic analysis is based on applying Newton's Second Law of motion to the components of an operating microengine. In the foregoing analysis of the microengine, we have linkages 4 and 6 as well as horizontal and vertical comb drives that are moving. Therefore, the x and y components of the forces acting on the pin of the drive gear, due to the linkages and comb drives, can be written as

$$F_x = FD4_x + F4_x + F6_x, \tag{12.15}$$

and

$$F_y = FE6_y + F4_y + F6_y, \tag{12.16}$$

where the components of the forces $F4$, $F6$, $FD4$, and $FD6$ are determined based on the dimensions, materials, operating conditions, and relative motions of specific components with respect to the other components of the micromirror device.

Using Equations (12.15) and (12.16), magnitude of force acting on the pin of the drive gear can be computed as

$$F = \sqrt{F_x^2 + F_y^2}, \tag{12.17}$$

which is a function of time. Sample calculations of the forces acting on the pin during operation of the micromirror device follow in subsequent sections.

In addition to analytical modeling of kinematics and kinetics of the microsystem, computational modeling of its dynamics was also performed using multiphysics approach [24,25].

12.4. OPTOELECTRONIC METHODOLOGY

Optoelectronic methodology, as presented in this chapter, is based on the principles of optoelectronic holography (OEH) [10,16,17,26]. Basic configuration of the OEH system is shown in Figure 12.9. In this configuration, laser light is launched into a single mode optical fiber by means of a microscope objective (*MO*). Then, a single mode fiber is coupled into two fibers by means of a fiber optic directional coupler (*DC*). One of the optical fibers comprising the *DC* is used to illuminate the object along the direction $\mathbf{K}_1$, while the output

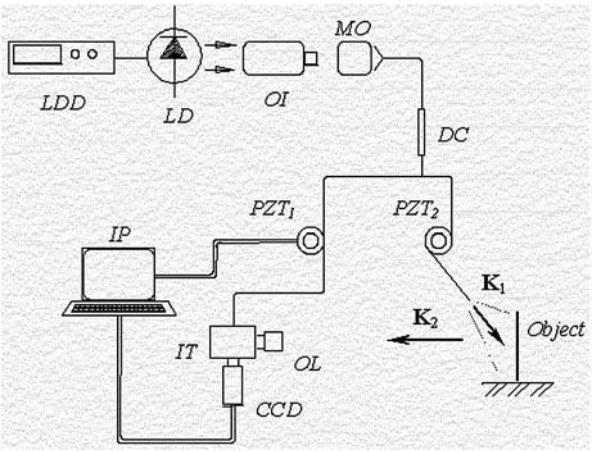


FIGURE 12.9. Single-illumination and single-observation geometry of a fiber optic based OEH system: *LDD* is the laser diode driver, *LD* is the laser diode, *OI* is the optical isolator, *MO* is the microscope objective, *DC* is the fiber optic directional coupler, PZT_1 and PZT_2 are the piezoelectric fiber optic modulators, *IP* is the image-processing computer, *IT* is the interferometer, *OL* is the objective lens, *CCD* is the camera, while $\mathbf{K}_1$ and $\mathbf{K}_2$ are the directions of illumination and observation, respectively.

from the other fiber provides reference against which signals from the object are recorded. Both, the object and reference beams are combined by the interferometer (*IT*) and recorded by the system camera (*CCD*).

Images recorded by the CCD are processed by the image-processing computer (IP) to determine the fringe-locus function, Ω, constant values of which define fringe loci on the surface of object under investigation. The values of Ω relate to the system geometry and the unknown vector $\mathbf{L}$, defining deformations, via the relationship [27]

$$\Omega = \left(\mathbf{K}_2 - \mathbf{K}_1\right) \cdot \mathbf{L} = \mathbf{K} \cdot \mathbf{L}, \tag{12.18}$$

where $\mathbf{K}$ is the sensitivity vector defined in terms of vectors $\mathbf{K}_1$ and $\mathbf{K}_2$ identifying directions of illumination and observation, respectively, in the OEH system, Figure 12.9.

Quantitative determination of structural deformations due to the applied loads can be obtained, by solving a system of equations similar to Equation (12.18), to yield [27]

$$\mathbf{L} = \left[\tilde{\mathbf{K}}^{\mathrm{T}}\tilde{\mathbf{K}}\right]^{-1}\left(\tilde{\mathbf{K}}^{\mathrm{T}}\Omega\right), \tag{12.19}$$

where $\tilde{\mathbf{K}}^{\mathrm{T}}$ represents a transpose of the matrix of the sensitivity vectors $\mathbf{K}$.

Equation (12.19) indicates that deformations determined from interferograms are functions of $\mathbf{K}$ and Ω, which have spatial, i.e., (x, y, z), distributions over the field of interest on the object being investigated. Equation (12.19) can be represented by a phenomenological equation [28]

$$\mathbf{L} = \mathbf{L}(\mathbf{K}, \Omega), \tag{12.20}$$

based on which the RSS-type (where RSS represents *the Square Root of the Sum of the Squares*) uncertainty in $\mathbf{L}$, i.e., $\delta\mathbf{L}$, which can be determined to be [28]

$$\delta\mathbf{L} = \left[\left(\frac{\partial\mathbf{L}}{\partial\mathbf{K}}\delta\mathbf{K}\right)^2 + \left(\frac{\partial\mathbf{L}}{\partial\Omega}\delta\Omega\right)^2\right]^{1/2}, \tag{12.21}$$

where $\partial\mathbf{L}/\partial\mathbf{K}$ and $\partial\mathbf{L}/\partial\Omega$ represent partial derivatives of $\mathbf{L}$ with respect to $\mathbf{K}$ and Ω, respectively, while $\delta\mathbf{K}$ and $\delta\Omega$ represent the corresponding uncertainties in $\mathbf{K}$ and Ω, respectively. It should be remembered that $\mathbf{K}$, $\mathbf{L}$, and Ω are all functions of spatial coordinates (x, y, z), i.e., $\mathbf{K} = \mathbf{K}(x, y, z)$, $\mathbf{L} = \mathbf{L}(x, y, z)$, and $\boldsymbol{\Omega} = \boldsymbol{\Omega}(x, y, z)$, respectively, when performing partial differentiations. After evaluating, Equation (12.21) indicates that $\delta\mathbf{L}$ is proportional to the product of the local value of $\mathbf{L}$ with the RSS value of the ratios of the uncertainties in $\mathbf{K}$ and Ω to their corresponding local values, i.e.,

$$\delta\mathbf{L} \propto \mathbf{L}\left[\left(\frac{\delta\mathbf{K}}{\mathbf{K}}\right)^2 + \left(\frac{\delta\Omega}{\Omega}\right)^2\right]^{1/2}. \tag{12.22}$$

For typical geometries of the OEH systems used in recording of interferograms, the values of $\delta\mathbf{K}/\mathbf{K}$ are less than 0.01. However, for small deformations, of the magnitudes encountered while studying MEMS/MOEMS, the typical values of $\delta\Omega/\Omega$ are about one order of magnitude greater than the values for $\delta\mathbf{K}/\mathbf{K}$. Therefore, the accuracy with which the fringe orders are determined influences the accuracy in the overall determination of deformations [29,30]. To minimize this influence, a number of algorithms for determination

of Ω have been developed. Some of these algorithms require multiple recordings of each of the two states, in the case of double-exposure method, of the object being investigated with introduction of a discrete phase step between the recordings [10,31,32].

For example, the intensity patterns of the first and the second exposures, i.e., $I_n(x, y)$ and $I'_n(x, y)$, respectively, in the double-exposure sequence can be represented by the following equations:

$$I_n(x, y) = I_o(x, y) + I_r(x, y) + 2\big\{[I_o(x, y)][I_r(x, y)]\big\}^{1/2} \cos\big\{[\varphi_o(x, y) - \varphi_r(x, y)] + \theta_n\big\}, \tag{12.23}$$

and

$$I'_n(x, y) = I_o(x, y) + I_r(x, y) + 2\big\{[I_o(x, y)][I_r(x, y)]\big\}^{1/2} \cos\big\{[\varphi_o(x, y) - \varphi_r(x, y)] + \theta_n + \Omega(x, y)\big\}, \tag{12.24}$$

where I_o and I_r denote the object and reference beam irradiances, respectively, with (x, y) denoting spatial coordinates, φ_o denotes random phase of the light reflected from the object, φ_r denotes the phase of the reference beam, θ_n denotes the applied n-th phase step, and Ω is the fringe-locus function relating to the displacements/deformations the object incurred between the first and the second exposures; Ω is what we need to determine. When Ω is known, it is used in Equation (12.19) to find [27] $\mathbf{L}$.

In the case of 5-phase-steps algorithm with $\theta_n = 0, \pi/2, \pi, 3\pi/2$, and 2π, the distribution of the values of Ω can be determined using [32]

$$\Omega(x, y) = \tan^{-1}\left\{\frac{2[I_2(x, y) - I_4(x, y)]}{2I_3(x, y) - I_1(x, y) - I_5(x, y)}\right\}. \tag{12.25}$$

Results produced by Equation (12.25) depend on the capabilities of the illumination, the imaging, and the processing subsystems of the OEH system. Developments in laser, fiber optics, CCD camera, and computer technologies have led to advances in the OEH metrology; in the past, these advances almost paralleled the advances in the image recording media [30]. A fiber optics based OEH system, incorporating these developments, is depicted in Figure 12.10. In addition to being able to measure static and dynamic deformations of objects subjected to a variety of BIL conditions, the system shown in Figure 12.10 is also able to measure absolute shape of the objects using multiple-wavelength optical contouring [10]. This dual-use is possible because of rapid tuning of the laser and real-time monitoring of its output characteristics by the wavelength meter (WM) and the power meter (PM)—both-integrated into the OEH system. In the configuration shown in Figure 12.10, the image-processing computer (IP) controls all functions of the OEH system.

In response to the needs of the emerging MEMS technology, an optoelectronic laser interferometric microscope (OELIM) system for studies of objects with micron size features was developed [16,33]. In the OELIM system, Figure 12.11, the light beam produced by a laser is directed into an acousto-optic modulator (AOM) and then into a single mode optical fiber. The output of the fiber is collimated by the collimating illumination lens subsystem (C). The resulting light field is then divided into reference and object beams by the beam splitter (BS). The reference beam is directed toward a PZT actuated mirror (M) and back to the beam splitter. The object beam is directed toward the MEMS under study and

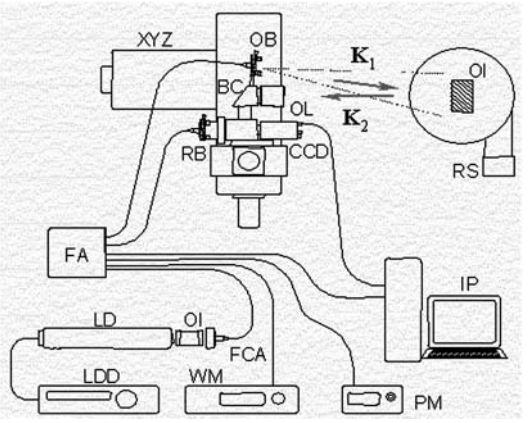


FIGURE 12.10. Fiber-optic based OEH setup arranged to perform high-resolution surface shape and deformation measurements: LDD is the laser diode driver, WM is the wavelength meter, PM is the optical power meter, LD is the laser diode, OI is the optical isolator, FCA is the fiber coupler assembly, IP is the image-processing computer, FA is the single-mode fiber optic directional coupler assembly, RB is the FC-connectorized reference beam fiber, CCD is the digital CCD camera, RS is the rotational stage, BC is the beam combiner, OL is the objective lens, XYZ is the X-Y-Z translational stage, OB is the FC-connectorized object or illumination beam fiber, OI is the object under investigation, while $\mathbf{K}_1$ and $\mathbf{K}_2$ are the vectors defining illumination and observation directions, respectively.

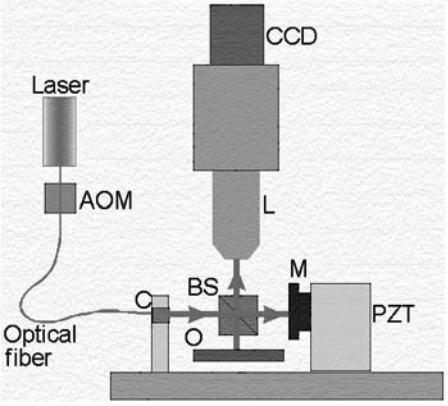


FIGURE 12.11. Optical configuration of the OELIM setup: AOM is the acousto-optic modulator, C is the collimating illumination lens subsystem, BS is the beam splitter, L is the long working distance microscope objective, M and PZT comprise the phase stepping mirror, and O is the MEMS object.

is reflected back to the beam splitter. The two beams recombine at the BS and are imaged by the long working distance microscope objective (L) onto a sensing element of the CCD camera, which records the resulting interference patterns. These patterns are, finally, transferred to the system computer for subsequent quantitative processing and display of the results.

Using the systems shown in Figures 12.10 and 12.11, issues relating to the sensitivity, accuracy, and precision, associated with application of the algorithm defined by Equation (12.25), were studied while evaluating effects that the use of high-spatial and

high-digital resolution cameras would have on the results produced [34]. In addition, development of optimum methods for driving lasers is conducted. This development is closely coupled with the development of fiber optic couplers and corresponding subsystems for efficient beam delivery.

12.5. REPRESENTATIVE APPLICATIONS

Using the analytical model presented in Section 12.3, forces acting on the microengine during its operation were calculated. Representative results of these calculations are shown in Figures 12.12 to 12.14 for the pin connecting the electrostatic comb drive linkages to the drive gear. More specifically, Figures 12.12 and 12.13 show polar plot representations of the x and y components of the forces acting on the drive gear pin and indicate that the magnitude of these forces increases from 4 nN, when the microengine operates at 6000 rpm, to 27 μN, when the microengine runs at 500,000 rpm. Figure 12.14 shows the magnitude of the drive gear pin force as a function of rotational speed of the microengine. Clearly, this force increases nonlinearly at an increasing rate as the rotational speed of the microengine increases.

The forces generated during operation of the microengine, load the drive gear and make it wobble as it rotates around its shaft. A unique capability to measure this wobble is provided by the OELIM methodology. Typical results obtained for two different positions in the rotation cycle of the drive gear are shown in Figure 12.15, where fringe patterns vividly display changes in magnitude and direction of the displacements/deformations of the microgears. Displacements of the drive gear, corresponding to the fringe patterns shown

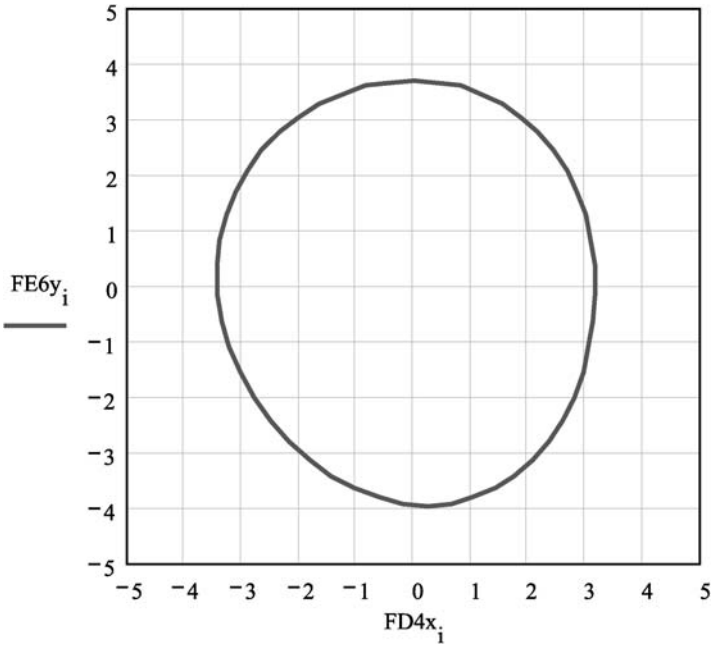


FIGURE 12.12. Polar representation of the x and y components of the force acting on the pin connecting the electrostatic comb linkages to the drive gear, for the microengine operating at 6000 rpm. Force is shown in nN.

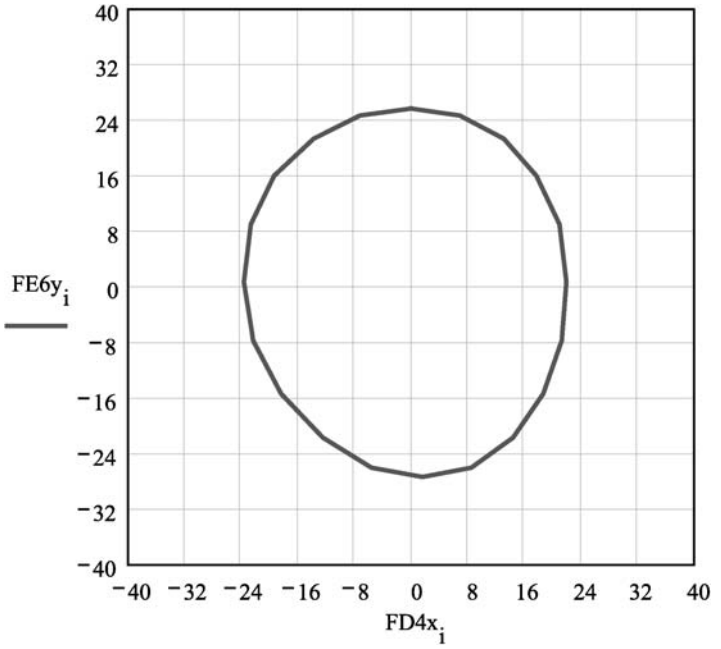


FIGURE 12.13. Polar representation of the x and y components of the force acting on the pin connecting the electrostatic comb linkages to the drive gear, for the microengine operating at 500,000 rpm. Force is shown in μN.

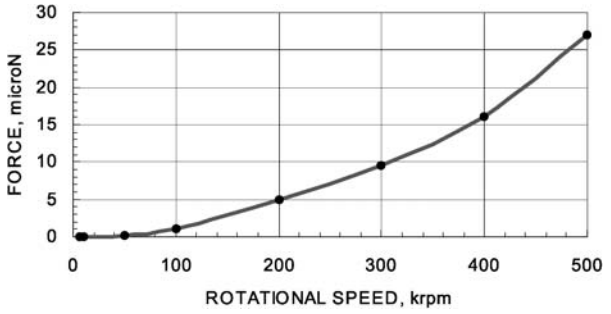


FIGURE 12.14. Force acting on the pin of the drive gear versus rotational speed of the microengine.

in Figure 12.15, are displayed in Figure 12.16 and are seen to vary in magnitude from 0.8 μm to 1.6 μm (it should be realized that thickness of the microgears in the direction of measured deformations is about 2 μm. These variations are due to kinematics and kinetics caused by operational impulsive loading forces generated by the input signals during a typical rotation cycle. In addition, the experimental results show that the wobble depends on the angular position in the rotation cycle, which can be related to the forces exerted on the drive gear by the pin during the cycle, Figures 12.12 and 12.13.

Representative deformations of the microgears, when the microengine operates at 360,000 rpm, are shown in Figure 12.17 and indicate maximum deformations of 1.8 μm.

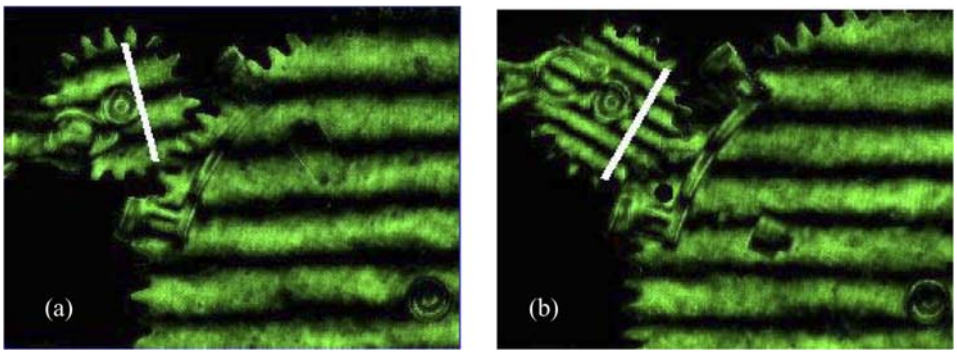


FIGURE 12.15. Representative OELIM fringe patterns recorded during a study of dynamic characteristics of microengines, at two different positions in a rotation cycle. White lines indicate locations (a) and (b) where measurements of displacements shown in Figure 12.16 were made.

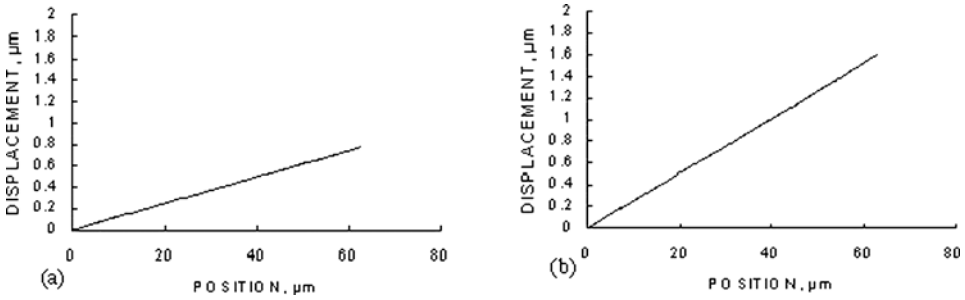


FIGURE 12.16. Displacements of the drive gear of the microengine along the white lines: (a) of Figure 12.15(a), and (b) of Figure 12.15(b).

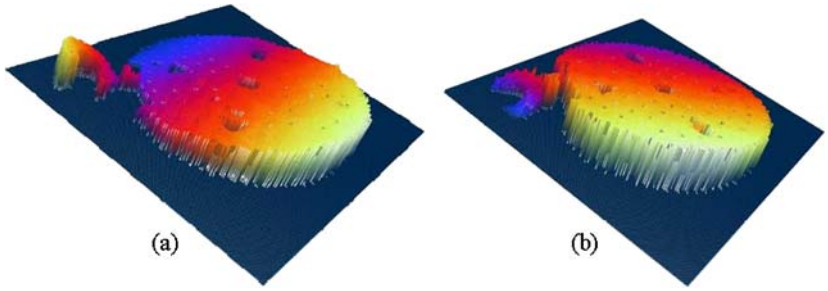


FIGURE 12.17. Deformations of the microgears of the microengine operating at 360,000 rpm: (a) 0° position, (b) 90° position.

Operational functionality of the micromirror system depends also on the quality of motions of section AB of the hinged micromirror, Figure 12.18. OELIM was used to measure these motions by recording fringe patterns, Figure 12.19, which were, in turn, interpreted to determine displacements/deformations of the section AB of the micromirror, Figure 12.20.

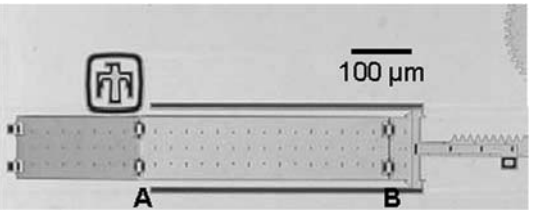


FIGURE 12.18. Measurements were made on the AB section, 100 µm wide and 400 µm long, of the hinged micromirror.

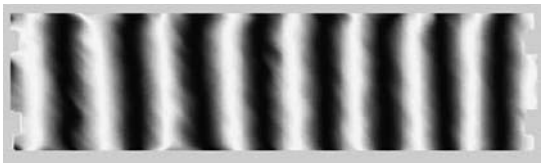


FIGURE 12.19. Representative OELIM fringe pattern of the AB section of the hinged micromirror, shown in Figure 12.18.

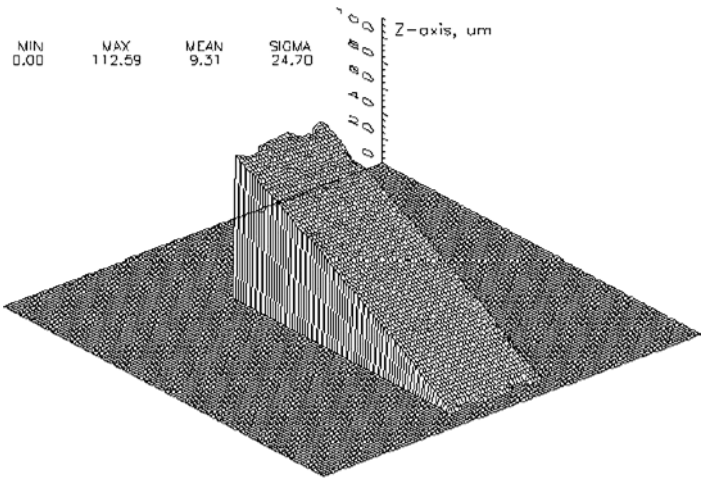


FIGURE 12.20. Wireframe representation of absolute shape and deformations of the section AB of the hinged micromirror, corresponding to the upright position displayed in Figure 12.1. Measurements show that the micromirror displacements range from 0 µm at hinge B to 113 µm at hinge A at which there is also a tilt of 18 mrad.

Figure 12.20 shows that, for the operating conditions for which the fringe pattern of Figure 12.19 was recorded, the out of plane displacement of the section AB of the micromirror was about 113 µm.

Using Figure 12.20, detailed information about deformations of the micromirror can be obtained, Figure 12.21. In this figure, traces (made parallel to the long edges of section AB of the micromirror) are shown. Noticeable differences between the two traces were observed and led to calculation of a tilt amounting to about 18 mrad at the hinge A.

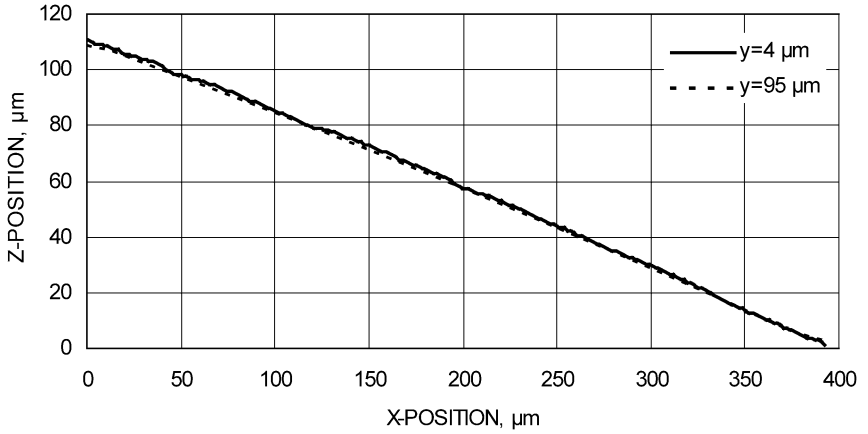


FIGURE 12.21. Vertical displacements (Z-POSITION) of section AB as a function of position along the length (X-POSITION) of the micromirror as determined from traces parallel to the long edges of the wireframe section shown in Figure 12.20.

Following procedures used to obtain representative results shown in this section, deformations and motions of other MEMS/MOEMS can also be determined. Results of these future studies will be reported on in subsequent publications.

12.6. CONCLUSIONS AND RECOMMENDATIONS

Novel optoelectronic methodology for testing of MOEMS was presented. This methodology is based on the optoelectronic laser interferometric microscopy (OELIM) and provides remote submicron measurements in near real-time in full-field-of-view and under actual operating conditions.

Representative results indicate that various components of the micromirror device deform up to 1.8 µm, depending on the position in the rotation cycle and the corresponding force system acting on the component. Although on the micrometer-scale, these deformations are rather large when compared with a nominal thickness of 2 µm of the gears and other moving components of the microengines considered herein.

To complement measurements, a vector based analytical model was developed to determine forces acting on various components of the micromirror device. Using this model, it was shown that the forces acting on the pin connecting the comb drive linkages to the drive gear range from 4 nN, when the microengine is running at 6000 rpm, to 27 µN, when the engine is operating at 500,000 rpm. Furthermore, polar representations of the Cartesian components of the forces acting on the pin while the engine is operating at a constant speed, show variation in magnitude of these components. This, together with fabrication tolerances, gives rise to wobble of the gears as they rotate around their hubs, which is not desirable for sustained operations at any speed.

At even higher speeds, up to 1,000,000 rpm as dictated by requirements in specific programs, forces will substantially increase and will lead to even larger motions and instabilities than measured thus far. Therefore, work of the type presented in this chapter is very timely and will contribute to further advances of the MOEMS.

The results presented in this chapter indicate that vector mechanics approach combined with noninvasive OELIM methodology is a viable hybrid-tool (consisting of the experimental and analytical/computational methods) for characterization of dynamic effects in the micromirror devices as well as in other MOEMS. By understanding the details of MOEMS performance in three-dimensions, we can make specific suggestions for improvements in their design and fabrication based on the SUMMiT™ technology [35,36].

Need for remote and noninvasive measurements in full-field-of-view providing data in three-dimensions and in real-time that optoelectronic methodology is capable of will be ever increasing as the emerging technologies (ET) evolve into mature technologies. This need will continue to be over multiscales ranging from milliscale to nanoscale and even down to picoscale as the "building blocks" out of which large structures will be made in the future, as the ET evolve with advances in Nanotechnology, will be shrinking in size. To be ready to satisfy the testing and characterization demands that ET will generate, development of metrology, specifically optoelectronic methodologies, should be continued.

ACKNOWLEDGMENTS

The micromirror devices used in this study were fabricated at and provided by Sandia National Laboratories. Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Company, for the United States Department of Energy under Contract DE-AC04-94AL85000.

REFERENCES

1. D.L. Hetherington and J.J. Sniegowski, Improved polysilicon surface-micromachined micromirror device using chemical-mechanical polishing, Proc. Internat. SPIE Symp. on Optical Sci., Eng., and Instrumentat., San Diego, CA, 1998.
2. S.C. Gustafson, G.R. Little, D.M. Burns, V.M. Bright, and E.W. Watson, Microactuated mirrors for beam steering, Proc. SPIE, 3008, pp. 91–99 (1997).
3. M. Ikeda, H. Goto, H. Totani, M. Sakata, and T. Yada, Two-dimensional miniature optical scanning sensor with silicon micromachined scanning mirror, Proc. SPIE, 3008, pp. 111–122 (1997).
4. R.L. Clark, J.R. Karpinski, J.A. Hammer, R. Anderson, R. Lindsey, D. Brown, and P. Merritt, Micro-optoelectromechanical (MOEM) adaptive optic system, Proc. SPIE, 3008, pp. 12–24 (1997).
5. S. Kurth, R. Hahn, C. Kaufman, K. Keher, J. Mehnerm, U. Wollman, W. Dotzel, and T. Gessner, Silicon mirrors and micomirror arrays for spatial laser beam modulation, Sensors and Actuators, A66, pp. 76–82 (1998).
6. T. Gessner, W. Dotzel, D. Billlep, R. Hahn, C. Kaufmann, K. Kehr, C. Steiniger, and U. Wollman, Silicon mirror arrays fabricated using bulk- and surface-micromachining, Proc. SPIE, 3008, pp. 296–305 (1997).
7. J.B. Sampsell, Digital micromirror device and its application to projection displays, J. Vac. Sci. Technol., B12, pp. 3242–3246 (1994).
8. R.J. Pryputniewicz, A hybrid approach to deformation analysis, Proc. SPIE, 2342, pp. 282–296 (1994).
9. C. Furlong and R.J. Pryputniewicz, Hybrid computational and experimental approach for the study and optimization of mechanical components, Opt. Eng., 37, pp. 1448–1455 (1998).
10. C. Furlong, Hybrid, experimental and computational, approach for the efficient study and optimization of mechanical and electro-mechanical components, Ph.D. Dissertation, Worcester Polytechnic Institute, Worcester, MA, 1999.
11. D.R. Pryputniewicz, ACES approach to the development of microcomponents, MS Thesis, Worcester Polytechnic Institute, Worcester, MA, 1997.
12. R.J. Pryputniewicz, P. Galambos, G.C. Brown, C. Furlong, and E.J. Pryputniewicz, ACES characterization of surface micromachined microfluidic devices, Internat. J. of Microelectronics and Electronic Packaging (IJMEP), 24, pp. 30–36 (2001).

13. D.R. Pryputniewicz, C. Furlong, and R.J. Pryputniewicz, ACES approach to the study of material properties of MEMS, Proc. Internat. Symp. on MEMS: Mechanics and Measurements, Portland, OR, 2001, pp. 80–83.
14. P.J. Saggal, V. Steward, C. Furlong, and R.J. Pryputniewicz, Analytical and experimental study of dynamics of a MEMS accelerometer, MRS Proc. Nano- and Micro-Electromechanical Systems (NEMS and MEMS) and Molecular Machines, Boston, MA, 2002.
15. C. Furlong and R.J. Pryputniewicz, Computational and experimental approach to thermal management in microelectronics and packaging, J. Microelectronics Internat., 18, pp. 35–39 (2001).
16. G.C. Brown, Laser interferometric methodologies for characterizing static and dynamic behavior of MEMS, Ph.D. Dissertation, Worcester Polytechnic Institute, Worcester, MA, 1999.
17. R.J. Pryputniewicz, M.P. de Boer, and G.C. Brown, Advances in optical methodology for studies of dynamic characteristics of MEMS microengines rotating at high speeds, Proc. IX Internat. Congress on Exp. Mech., SEM, Bethel, CT, 2000, pp. 1009–1012.
18. M. Rogers and J.J. Sniegowski, 5-level polysilicon surface micromachining technology: application to complex mechanical systems, Tech. Digest of the Solid State Sensor and Actuator Workshop, Hilton Head Island, SC, 1998.
19. E.J. Garcia and J.J. Sniegowski, Surface micromachined microengine, Sensors and Actuators, A48, pp. 203–214 (1995).
20. M.S. Rogers, S.L. Miller, J.J. Sniegowski, and G.F. LaVigne, Designing and operating electrostatically driven microengines, Proc. 44th Internat. Instrumentation Symp., Reno, NV, 1998, pp. 56–65.
21. V.R. Yarberry, Meeting the MEMS "design-to-analysis" challenge: the SUMMiT™V design tool environment, Paper No. IMECE2002–39205, Am. Soc. Mech. Eng., New York, NY, 2002.
22. E.J. Pryputniewicz, ACES approach to the study of electrostatically driven MEMS microengines, MS Thesis, Worcester Polytechnic Institute, Worcester, MA, 2000.
23. E.J. Pryputniewicz, S.L. Miller, M.P. de Boer, G.C. Brown, R.R. Biederman, and R.J. Pryputniewicz, Experimental and analytical characterization of dynamic effects in electrostatic microengines, Proc. Internat. Symp. on Microscale Systems, Orlando, FL, 2000, pp. 80–83.
24. R.J. Pryputniewicz, Integrated approach to teaching of design, analysis, and characterization in micromechatronics, Paper No. IMECE2000/DE-13, Am. Soc. Mech. Eng., New York, NY, 2000.
25. A.J. Przekwas, M. Turowski, M. Furmanczyk, A. Hieke, and R. J. Pryputniewicz, Multiphysics design and simulation environment for microelectromechanical systems, Proc. Internat. Symp. on MEMS: Mechanics and Measurements, Portland, OR, 2001, pp. 84–89.
26. C. Furlong and R.J. Pryputniewicz, Characterization of shape and deformation of MEMS by quantitative optoelectronic metrology techniques, Proc. SPIE, 4778, pp. 1–10 (2002).
27. R.J. Pryputniewicz, Quantitative determination of displacements and strains from holograms, in Holographic Interferometry, Vol. 68 of Springer Series in Sciences, Springer-Verlag, Berlin, 1995, Ch. 3, pp. 33–72.
28. R.J. Pryputniewicz, Engineering Experimentation, Worcester Polytechnic Institute, Worcester, MA, 1993.
29. R.J. Pryputniewicz, High precision hologrammetry, Internat. Arch. Photogramm., 24, pp. 377–386 (1981).
30. R.J. Pryputniewicz, Hologram interferometry from silver halide to silicon and ... beyond, Proc. SPIE, 2545, pp. 405–427 (1995).
31. C. Furlong and R.J. Pryputniewicz, Absolute shape measurements using high-resolution optoelectronic holography methods, Opt. Eng., 39, pp. 216–223 (2000).
32. R.J. Pryputniewicz, P. Hefti, A.R. Klempner, R.T. Marinis, and C. Furlong, Hybrid methodology for the development of MEMS, J. Strain Analysis for Engineering Design, 41, pp. 708–718 (2006).
33. C. Brown and R.J. Pryputniewicz, Holographic microscope for measuring displacements of vibrating microbeams using time-average electro-optic holography, Opt. Eng., 37, pp. 1398–1405 (1998).
34. C. Furlong, J.S. Yokum, and R.J. Pryputniewicz, Sensitivity, accuracy, and precision issues in optoelectronic holography based on fiber optics and high-spatial and high-digital resolution cameras, Proc. SPIE, 4778, pp. 216–223 (2002).
35. R.J. Pryputniewicz, MEMS design education by case studies, Paper No. IMECE2001/DE-23292, Am. Soc. Mech. Eng., New York, NY, 2001.
36. R.J. Pryputniewicz, E. Shepherd, J.J. Allen, and C. Furlong, University—National Laboratory alliance for MEMS education, Proc. 4th Internat. Symp. on MEMS and Nanotechnology (4th-ISMAN), Charlotte, NC, 2003, pp. 364–371.

13

Durability of Optical Nanostructures: Laser Diode Structures and Packages, A Case Study

Ajay P. Malshe[a] and Jay Narayan[b]

[a]*Department of Mechanical Engineering, MEEG 204, University of Arkansas, Fayetteville, AR 72701, USA*
[b]*Department of Materials Science and Engineering, North Carolina State University, Raleigh, NC 21695-7907, USA*

Abstract

Durability is a synergistic reliable response of subsystems in integrated (packaged) systems, which in this case under discussion are nanostructured integrated optical systems. Understanding science and engineering aspects of these optical nanostructures integrated systems through design, fabrication, packaging and reliability testing are of paramount importance to obtain durable optical nanostructured packaged systems. To communicate specific aspects, this chapter addresses durability of optical quantum structures through carefully selected case studies in two parts. The part one includes novel design and deposition of quantum structures and the part two includes discussion of reliability of packaged quantum layered laser diode structures.

In the first case study, it is demonstrated that, $In_xGa_{(1-x)}N$ based multiquantum well (MQW) light emitting diodes and lasers (LEDs and LDs) have been fabricated and it is shown that high optical efficiency in these devices is related to thickness variation (TV) of $In_xGa_{(1-x)}N$ active layers. The thickness variation of active layers is found to be as important as In composition fluctuation in quantum confinement of excitons (carriers) in these devices. In this work, MQW $In_xGa_{(1-x)}N$ layers are produced with a periodic thickness variation, which results in periodic fluctuation of bandgap for the quantum confinement of excitons. Detailed STEM-Z contrast analysis, where image contrast is proportional to Z^2 (Z = atomic number), was carried out to investigate the spatial distribution of In. It is discovered that there is periodic variation in thickness of $In_xGa_{(1-x)}N$ layers with two periods, one short range (SR-TV, 30 to 40 Å) and other long-range thickness variation (LR-TV, 500 to 1000 Å). It is envisaged that LR-TV is the key to quantum confinement of the carriers and enhancing the optical efficiency and at the same time offering excellent reliability. The SR-TV is caused by In composition fluctuation. It was also found that the variation in In concentration is considerably less in the LED and LD structures which exhibit high optical efficiency. A comparative microstructural study between high and low optical efficiency MQW structures is presented to show that thickness variation (SR-TV) of $In_xGa_{(1-x)}N$ active layers is the key to their enhancement in optical efficiency.

Once quantum structures are engineered and devices are fabricated packaging and its reliability become important. Hence, in the second case study presented, authors discuss the laser diode package reliability for as applications continue to demand increasingly higher optical output power and longer lifetime, thermo-mechanical stresses on dissimilar materials interfaced for packaging pose an ever-growing challenge for the realization of a durable system. Particularly important for an epitaxy-down configuration is the die-attachment interface, which is desired to be defect-free and stress-managed for reliable optical alignment. A knowledge of the changes in the physical defect density and magnitude of the thermo-mechanical stress present in the active region as a function of the fabrication process and aging is crucial to an understanding of the influence of the process parameters and operating conditions on device performance and reliability. In this case study, we discuss investigation of high power laser diode array packages aged under various conditions. Microscopic defect analyses of the die attachment interface and device stress were carried out using primarily metallography, scanning electron microscopy (SEM), scanning acoustic microscopy (SAM), micro-hardness, and micro-Raman spectroscopy. It was noted that the intermetallic compounds and microscopic physical defects at the die attach interface are detrimental to transient heat transfer, and thus, overall package reliability. Using micro-Raman spectroscopy, we found that tensile stress near the bar-package interface increases with aging for the first few hundred hours and then decreases with further aging.

In conclusion, this chapter discusses synergistic engineering of nano structures along with micro interfaces in a macroscopic packaging which is essential for realizing a durable nanostructured integrated optical system.

13.1. HIGH EFFICIENCY QUANTUM CONFINED (NANOSTRUCTURED) III-NITRIDE BASED LIGHT EMITTING DIODES AND LASERS

13.1.1. Introduction[*]

The III-nitrides and their alloys have assumed a special importance due to their tremendous potential for fabricating the light emitting diodes and lasers (LEDs and LDs) operating in the red to ultraviolet (UV) energy range. The active layer in these devices with a composition of $In_xGa_{(1-x)}N$ has been the key to obtaining a high optical efficiency. The alloying with In is considered to be important, however, its role has not been clarified. Some studies have suggested In composition fluctuation leading to a phase separation to be responsible for high optical efficiency. Since the In content controls the bandgap in $In_xGa_{(1-x)}N$ alloys, it is envisaged that the composition fluctuation leads to quantum confined (QC) regions whose size is smaller than the dislocation separation (DS). These QC regions trap the bound excitons which recombine to produce photons, and thus recombination of excitons is not affected by the presence defects such as dislocations.

The evidence for indium composition fluctuation in $In_xGa_{(1-x)}N$ layers in MQW structures has been largely circumstantial. There is some evidence for phase separation (indium rich and indium poor phases) in $In_xGa_{(1-x)}N$ ($x > 0.3$) only in relatively thick layers (300–400 nm), which were grown by ECR-MBE. However, no phase separation is observed in GaN/$In_xGa_{(1-x)}N$/GaN double heterostructures with $x > 0.3$, grown under similar conditions. Similarly, in MOCVD grown samples, phase separation has been reported only in thick $In_xGa_{(1-x)}N$ layers for $x > 0.28$. In the case of multiple-quantum-well structures (sapphire/4000 nm GaN:Si/10 period 2 nm InGaN/4 nm GaN/200 nm GaN:Mg),

[*]J. Narayan et al., Appl. Phys. Lett. 81, 841 (2002); U.S. Patent #US 6,881,983 B2, April 19, 2005, and references there in.

phase separation was observed only after prolonged (40 h) post annealing above 950°C. Some authors have claimed the formation of indium-rich and indium-poor regions (2–5 nm size) in MOCVD grown MQW structures using diffraction contrast transmission electron microscopy (TEM) techniques. Since the image contrast in these techniques is sensitive to diffraction of atomic planes, these observations do not provide reliable information on composition fluctuation.

The LEDs and LDs based upon $In_xGa_{(1-x)}N$ multiquantum well (MQW) structures, which exhibit high optical efficiency, are found to show subband emissions. These emissions have been explained on the basis of In composition fluctuation in $In_xGa_{(1-x)}N$ layers. The energy separation of each subband emission in these samples is typically about 2 meV which is much smaller than expected for transition between $n = 1$ and $n = 2$ energy levels in 2–5 nm quantum dots. Thus, there is a urgent need to clarify the role of In composition fluctuation or any other effects leading to the formation of quantum confined regions. In this study, we have used high-resolution TEM and STEM-Z contrast techniques to investigate the In composition fluctuation and thickness variation and correlate them with optical efficiencies.

The $In_xGa_{(1-x)}N$ multiquantum well structures (10 period $In_xGa_{(1-x)}N$/GaN// 2/10 nm) were grown by the MOCVD technique at a temperature of 800°C and the GaN capping layer at 950°C for less than 30 min. The details of growth of these structures are reported elsewhere. These wafers were used to prepare cross-section specimens by a standard ion milling procedure. A special care was necessary to used in terms of low-temperature, low voltage and shallow angle thinning to minimize the surface damage for STEM-Z studies. For STEM-Z (scanning transmission electron microscopy-atomic number contrast studies, we used atomic resolution JEOL 2010 field emission electron microscope with GIF (Gatan Image Filter attachment. In the STEM-Z mode, a small electron probe (1.6 nm) is scanned across the thin cross-section specimen and the Z-contrast image results from mapping the intensity of electrons reaching the annular detector. The detector performs the function of Lord Rayleigh's condenser lens. It enforces high scattering angles, so that Rutherford scattering dominates and atoms contribute to the image with a brightness determined by their mean square atomic number (Z) and with a resolution of the probe size (1.6 nm). Since the atomic number of In (49) is much higher than that of Ga (31), the image contrast is dictated by In concentration.

Thus thickness variations can result in the formation of QC (quantum confined) regions. The InGaN/GaN quantum-well structure of high-efficiency LEDs is shown in Figure 13.1. The details of quantum-well structure and associated thickness variation to produce quantum-confined (QC) regions are shown in Figures 13.2 and 13.3 as a cross-section STEM-Z contrast image of $In_{0.2}Ga_{0.8}N$/GaN layers are shown in Figures 13.2 and 13.3. In the STEM-Z images, the contrast is proportional to Z^2 (Z = atomic number). Our contrast analysis reveals that the variation in In concentration is not as significant, and that the enhanced efficiency results from the thickness variation. The $In_{0.2}Ga_{0.8}N$ layers with similar composition but with uniform thickness resulted in considerably less optical efficiency.

The contrast due to indium is enhanced by two and half times compared to the gallium concentration. The SR-TV (short range thickness variation) period in Figures 13.2 and 13.3 is estimated to be 30 to 40 Å. Figures 13.2 and 13.3 show a STEM-Z (transmission electron microscopy) micrographs from specimens, which consistently showed higher optical efficiencies. These specimens showed a short range (30 to 40 Å period) and a long range (500 to 1000 Å period) thickness variation. In contrast to the high optical efficiency from specimens (shown in Figures 13.1– 13.3), the specimens with relative low optical efficiencies are shown in Figure 13.4. In these specimens, where optical efficiencies are lower

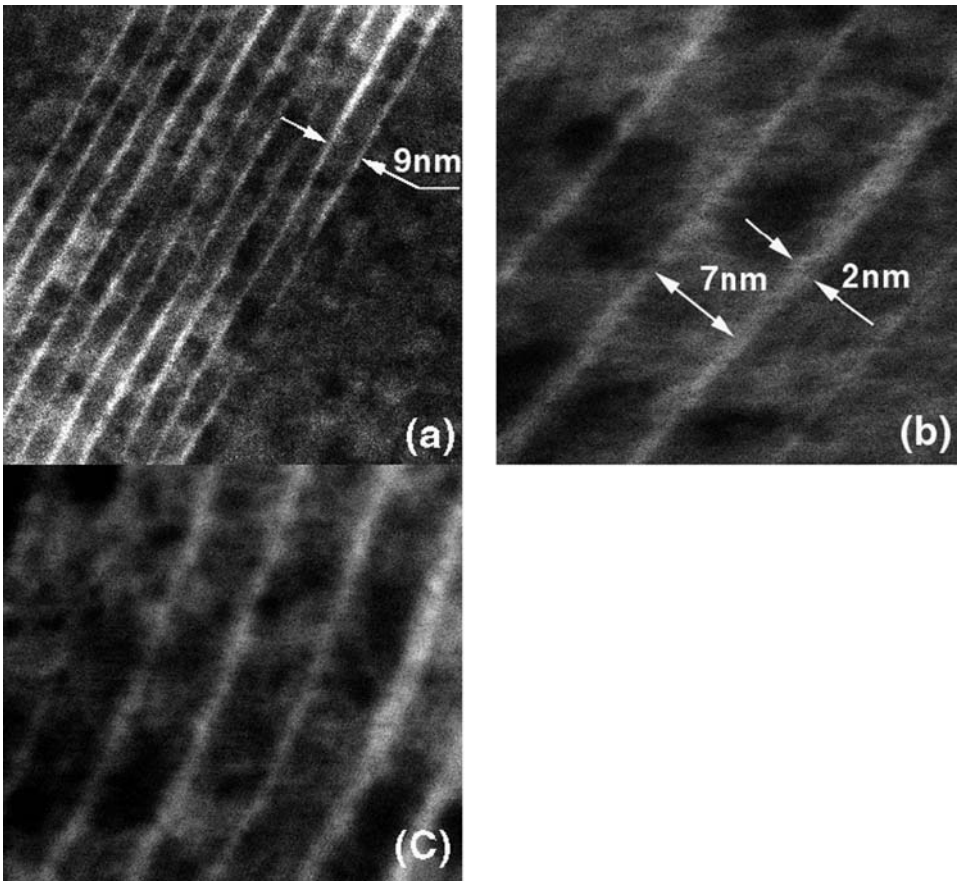


FIGURE 13.1. Quantum well structure of novel high-efficiency InGaN/GaN LEDs.

as much as a factor of two and three, superlattice thickness as well as indium concentration are quite uniform. Thus our experimental results on comparative study of high- and low-efficiency LEDs and LDs (as shown in Figure 13.5) clearly demonstrate that thickness variation coupled with indium concentration variation is the key to enhancing the optical efficiencies in LEDs and LDs.

In these studies, it is envisaged that the In composition fluctuation results in quantum-dot like structures from which subband emission occurs. The quantum well trap excitons whose radiative recombination is responsible for efficient spontaneous emission in MWQ LEDS and LDs. High optical efficiency results despite high dislocation density ($\sim 10^{10}$ cm^{-2}) because the localization of excitons is within a region less than the dislocation separation (DS) in these structures. The DS is given by $\rho^{-1/2}$, where ρ is the density of dislocations (number/cm^2). Thus, the loss of excitons due to nonradiative recombination at the dislocations is avoided resulting in high optical efficiency of LEDs and LDs. In our investigation, we have produced periodic thickness variations (short-range, ST-TV; and log-range, LR-TV) which result in the formation of QC regions for the excitons which recombine without being affected by the presence of dislocations. Our detailed STEM-Z contrast analysis shows that thickness of $In_xGa_{(1-x)}N$ layers are equally important as In composition fluctuation in producing quantum confined regions for excitons leading to en-

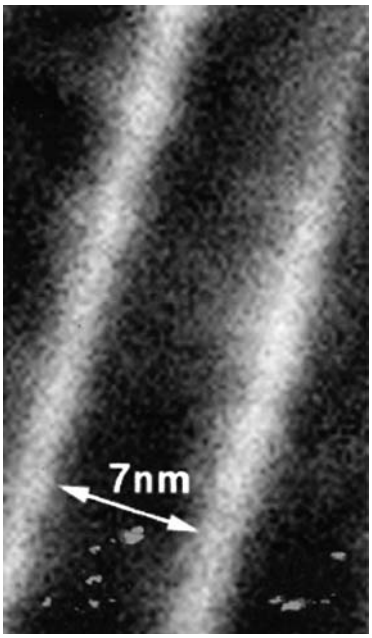


FIGURE 13.2. Thickness variation in InGaN/GaN quantum wells to confine the carriers.

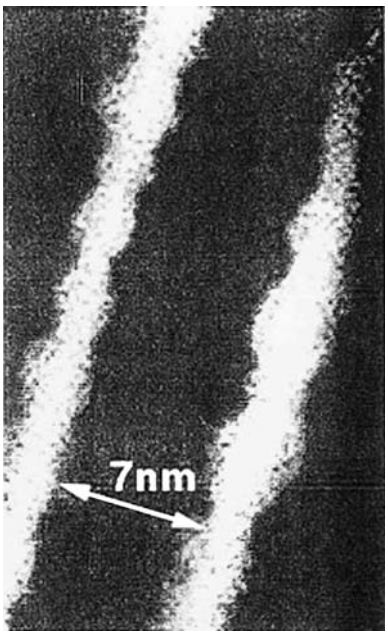


FIGURE 13.3. Another example of high resolution (STEM-Z) micrograph showing details of quantum confinement of carriers in InGaN nanopockets.

hanced optical efficiency of LEDs and LDs. In the previous studies, the fluctuation in In concentration $In_xGa_{(1-x)}N$ layers has been investigated by cross-section TEM (using con-

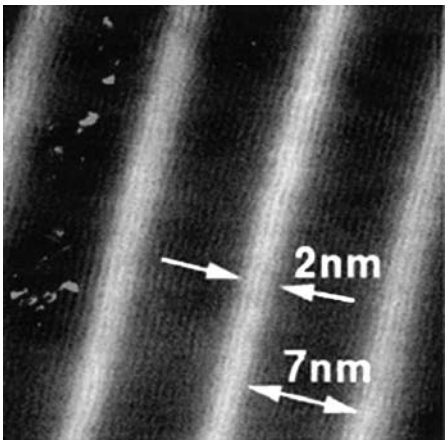


FIGURE 13.4. Uniform structure of InGaN/GaN quantum wells.

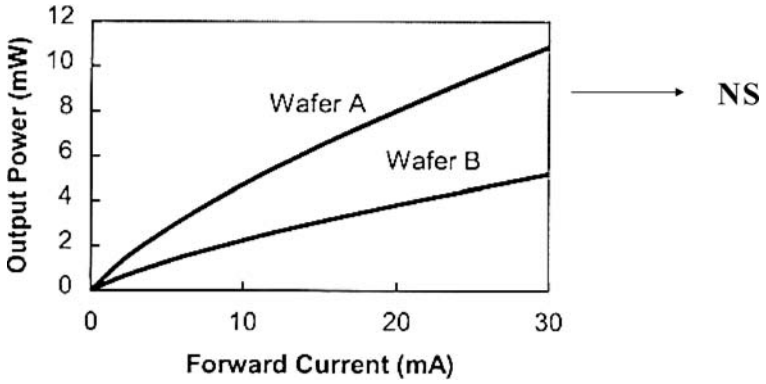


FIGURE 13.5. Comparison of LED efficiencies: (A) nonuniform quantum wells; (B) uniform quantum wells.

ventional phase contrast transmission electron microscopy), photoluminescence (PL) and Raman spectroscopy techniques. From these studies, the size of these regions was estimated to be 2–5 nm. The time-resolved photoluminescence spectroscopy and electroreflectance studies have revealed that a small In addition the GaN active layers plays a key role in suppressing the nonradiative recombination processes. The Raman Stokes shift between the exciting and emission in the range of 100 to 250 meV at RT was attributed to energy depth of localized states of the carriers (excitons) in the $In_xGa_{(1-x)}N$ layers.

The change in bandgap of $In_xGa_{(1-x)}N$ alloys can occur as function of the composition "x" and the thickness "L_z" of the superlattice. For a typical active layer composition ($x = 0.2$) the change in bandgap is estimated to be as follows: $x = 0.2$, bandgap $= 3.0$ eV; $x = 0.1$, bandgap $= 3.2$ eV; $x = 0.3$, bandgap $= 2.8$ eV. This amounts to a 50% change (from $x = 0.2$) in active layer composition. Experimentally observed composition fluctuations are less than 10% which should lead to a less than 0.07 eV change in the bandgap. (For $x = 0.35$, bandgap $= 2.70$ eV; $x = 0.40$, bandgap $= 2.60$ eV; $x = 0.45$, bandgap $= 2.55$ eV; $x = 0.50$, bandgap $= 2.47$ eV.) Since the bandgap is dictated by the thickness (L_z) via: $E_n = h^2n^2/(8m^*L_z^2)$, where E_n is the allowed energy level (n), h is Planck's constant, and m^* is effective mass. Thickness variations lead to changes in ΔE

proportional to L_z^{-2}. Thus, 10 to 20% in thickness variation can result in 20 to 40% change in ΔE (using $m_e^* = 0.11m_0$ for InN, $m_e^* = 0.20m_0$ for GaN, Refs. [8,9]). Experimentally observed thickness variations (long-range, LR-TV) are in the range of 10 to 50%, and short-range variations SR-TV are less than 10%.

It is proposed that the thickness variation is caused by two-dimensional strain in the $In_xGa_{(1-x)}N$ layer below its critical thickness. Since strain energy increases with thickness, the uniform thickness breaks into a periodic variation by which the free energy of the system can be lowered. Since the strain also increases with In concentration, some fluctuation in In concentration is also expected. However, this phenomenon of thickness variation has been well documented for pure germanium thin film growth on (100) silicon below its critical thickness where no composition fluctuation is involved. We have modeled the thickness variation and derived the following relation for TV period (λ)

$$\lambda = \pi\gamma(1-\nu)/\left[2(1+\nu)^2\mu\varepsilon^2\right], \tag{13.1}$$

where γ is the surface energy, ν is the Poisson's ratio, μ is the shear modulus of the film, and ε is the strain normal to the film surface. To avoid nonradiative recombination at the dislocations (density ρ), we derive the optimum structure to be

$$\rho^{-1/2} > \pi\gamma(1-\nu)/\left[2(1+\nu)^2\mu\varepsilon^2\right] \text{ or } \rho < \left\{\pi\gamma(1-\nu)/\left[2(1+\nu)^2\mu\varepsilon^2\right]\right\}^{-2}. \tag{13.2}$$

It is estimated that a typical value of λ using the following parameters for our growth conditions. For $In_{0.4}Ga_{0.6}N$, shear modulus is estimated to be 82 GPa, Poisson's ratio to be 0.3, surface energy 4000 ergs/cm^2, strain 2%, this results in λ of 793 Å or 800 Å, which is in good agreement with observed LR-TV.

The high efficiency MQW structured LEDs exhibit characteristic subband emission separated by 1–2 meV. In addition, cathodoluminescence measurements show Stokes like shift between the exciting and emission in the range of 100 to 250 meV. A spherical potential treatment suggested by Brus was used to make an estimate of the QC region corresponding to confinement energy (ΔE) expressed as $\pi^2\hbar^2/(2m^*R^2)$, where m^* is the reduced effective mass of electron-hole pairs, $\hbar$ is the Planck's constant, and R is the dot radius. Using this model we estimated the transition energy ΔE to be 2 meV corresponding to the 50 nm quantum dot radius. This is consistent with experimentally observed LR-TV period of about 80 nm. Similarly, Stokes like shift in the CL measurements of 200 meV is expected to arise from the quantum confined region of 5 nm, which is closer to SR-TV of 3–5 nm observed in our STEM-Z contrast experiments. The indium composition fluctuation can also result from surface diffusion flux of vacancies and lead to short range thickness variation of the order of 2–5 nm.

In summary, $In_xGa_{(1-x)}N$ based multiquantum well (MQW) light emitting diodes and lasers (LEDs and LDs) are fabricated and it is shown that high optical efficiency in these devices is related to thickness variation (TV) of $In_xGa_{(1-x)}N$ active layers. It is discovered that there is a periodic variation in thickness of $In_xGa_{(1-x)}N$ layers with two periods, one short range (SR-TV, 30 to 40 Å) and other long-range thickness variation (LR-TV, 500 to 1000 Å). It is envisaged that LR-TV, which may be related 2 meV subband emission is the key to quantum confinement of the carriers and enhancing the optical efficiency. It is envisaged that the SR-TV is caused by In composition fluctuation. It was also found that the variation in In concentration is considerably less in the LED and LD structures which exhibit high optical efficiency. The reliability in forming these nanostructures is the key to obtain sustained high-efficiency of LEDs and LDs.

13.2. INVESTIGATION OF RELIABILITY ISSUES IN HIGH POWER LASER DIODE BAR PACKAGES

13.2.1. Introduction

Increasing optical efficiency, new package designs, better optical coupling methods, a growing number of applications, and steadily declining prices have accelerated the transition of high power laser diodes from research and development into mainstream applications. As applications continue to demand increasingly higher optical output power and longer lifetime, thermo-mechanical stresses on dissimilar materials interfaced for packaging pose an ever-growing challenge to the realization of a durable system. Thus, it has become increasingly important to analyze the root causes of specific degradation modes at optimum laser operating conditions so as to manufacture reliable systems.

An edge emitting high power laser diode packaged system studied in this section is a combination of quantum-well laser diode arrays and multilayered integrated metalization schemes that combine dissimilar materials for die attachment on a copper heat sink, all functioning under high transient temperature conditions. Typically, failure of a packaged system is due to interrelated electro-thermo-mechanical-material reasons. For example, temperature cycling of a packaged laser bar attached using a soft solder results in creep and stress relaxation at the die attach interface causing mechanical deformation of the laser diode array. Such deformation causes variations in the optical emission across the bar. During product development and manufacturing, understanding the influence of packaging parameters and operating conditions on optical device performance, thermo-mechanical stresses on the device at the die attach interface, and physical defect density and microstructural changes in the die attachment material under continuously evolving/degrading conditions is crucial.

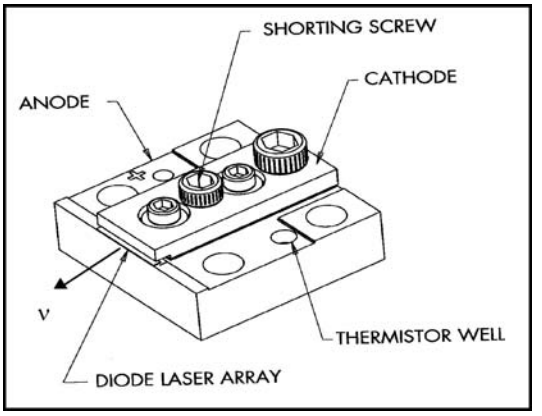
Over the years, many analytical and experimental studies have been performed to assess die-attachment joint integrity from a physics-of-failure perspective [1–12]. However, detailed analyses, correlating device performance and failure modes, packaging and operating conditions, and thermo-mechanical stresses and material behavior are necessary in order to realize reliable, application-specific, durable systems. Many inconsistencies result from a lack of knowledge of the unique properties of each solder, such as age and cycle softening, hardening due to intermetallic grain-growth, dynamic recrystallization, strain-rate hardening, superplasticity, etc. Many analytical inconsistencies are traced to differing interpretations of the effects of the temperature, the current, the cycle frequency, and the period [13].

Various kinds of problems are associated with thermo-mechanical stress, which appears in the active region of a device during its growth, packaging, and also, during its operation. Stress may directly trigger the nucleation and propagation of dislocations and the formation of voids and cracks. The absorption of photons by dislocations and the migration of carriers toward such dislocations in the active region of a diode result in additional thermo-mechanical stresses, and hence, can be a major reliability problem. Furthermore, even when stress is not severe enough to destroy the functionality of a diode array, its presence can influence routine performance by modifying the semiconductor band structure, which, in turn, affects the output wavelength, threshold current, and quantum efficiency. The problem becomes more acute with shrinking device size, increasing complexity of the integrated system, and increasing output power [14–23].

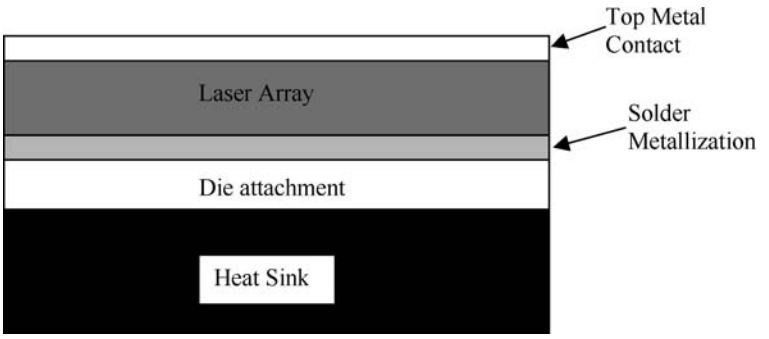
Consequently, the objective of this research work was to explore and understand various failure mechanisms affecting packaged high power laser diode bars under different testing/ operating conditions.

13.2.2. Preparation of Packaged Samples for Reliability Testing

The laser diode array packages investigated are continuous wave (CW) laser diode bars. A schematic diagram of a packaged bar is shown in Figures 13.6(a) and (b). The GaAs multiple quantum-well bars (19 emitters; 1 mm × 1 cm × 100 μm) were mounted on a polished copper heat sink in an *epitaxy side down* configuration using soft indium solder and Ti-Pt-Au interface metallurgy. A copper heat sink provided anode contact to the device, while a metal foil mounted on top of the bar provided cathode contact. Ridges of 150 μm widths were formed in the p-side of the bar. During die attachment reflow, a load of about 50 grams was applied in the direction normal to the plane of the bar. The reflow process was carried out in a vacuum furnace. The bars were then aged for 0, 96, 744, 1000 and 7000 hours at 40 A operating current for an output power of 30 W. Some bars succumbed to infant mortality, while most of the bars survived and operated satisfactorily for over 7000 hours.



(a)



(b)

FIGURE 13.6. (a) Schematic view of the package. (b) Schematic cross-sectional view of the package.

In order to identify and understand the defects and related failure modes, samples were characterized using the following complementary analytical techniques: optical microscopy and scanning electron microscopy (SEM, Hitachi) to observe interface microstructure and micro-defects, scanning acoustic microscopy (SAM, Sonoscan) to explore void distribution at the device bar and die-attachment interface, energy dispersive spectroscopy (EDS, Kevex) for chemical analysis of the interface, and micro-hardness (Buehler) to test the mechanical response as a function of thermal degradation. The purpose of this analysis was to determine the relationship between material microstructure, micro-hardness, and chemical changes in the die-attachment as a function of aging of the package.

Thermo-mechanical stresses present in the active region of the device were measured using micro-Raman spectroscopy (Renishaw) by observing the shift of the characteristic GaAs peak (the stress measurement error was $\pm 5\%$) [24–29]. Micro-Raman measurements were performed using a 488 nm Ar^+ laser with a spatial resolution of about 1 μm. Observations were made at various locations along and across the active emitter regions of the laser bars. Precisely scanning of the quantum-well along the 1 cm width with a spatial resolution close to 1 μm was practically difficult. Thus, the package-aging induced stress profile was measured across the width of the active emitter region of the bar. Observations were made on the front [110] facet of the device. In this configuration, scattering from the transverse optical (TO) phonons was allowed in accordance with the symmetry selection rules [15]. The shift in the TO peaks was studied as a function of aging. Positive and negative peak shifts correspond to compressive and tensile stress in the GaAs material, respectively. Since the peak shifts are small, peak fitting was employed systematically. The Gaussian peak fit was used for the laser peak and mixed Gaussian and Lorenztian peak fits were used for the Raman peak. The change in the full width at half maximum (FWHM) of the Raman peak was measured and plotted as a function of spatial position [15]. Effects of laser heating of the sample, in our case, were minimal because of the confocal Renishaw Raman spectroscopy system, which facilitates measurements with low laser power at high speed, thereby reducing the laser heating effects. It is reported that lattice heating caused by a cw excitation laser does not exceed 10 K [18].

13.2.3. Finding and Model of Reliability Results

13.2.3.1. Physical Defect and Morphological Observations The following is a detailed discussion of the various micro-defect structures observed in the aged laser bars.

Non-uniform physical contact between the top cathode metal foil contact and the laser diode bar, the bar and the multilayer metalization, the metalization and the die-attachment, and the die-attachment and the copper heat sink were observed at various locations across the width of a bar. This non-uniform contact varied in morphology from delamination (Figure 13.7) to crack propagation in metalization (Figure 13.8) to cavitation at the ridges to crack propagation near the ridge structures. Figure 13.9 provides evidence that non-uniform physical contact, resulting from non-uniform microloading on the GaAs bar along the metalization and die-attach interface during the die-attachment process, particularly at the ridges, can cause cavitations and stress gradients. This results in cracking of the bar.

Coefficient of thermal expansion (CTE) mismatches between the metalization, die attachment, and copper heat sink are responsible for delaminations and cracks that run along the interface. Such non-uniform physical contact between a high power laser diode

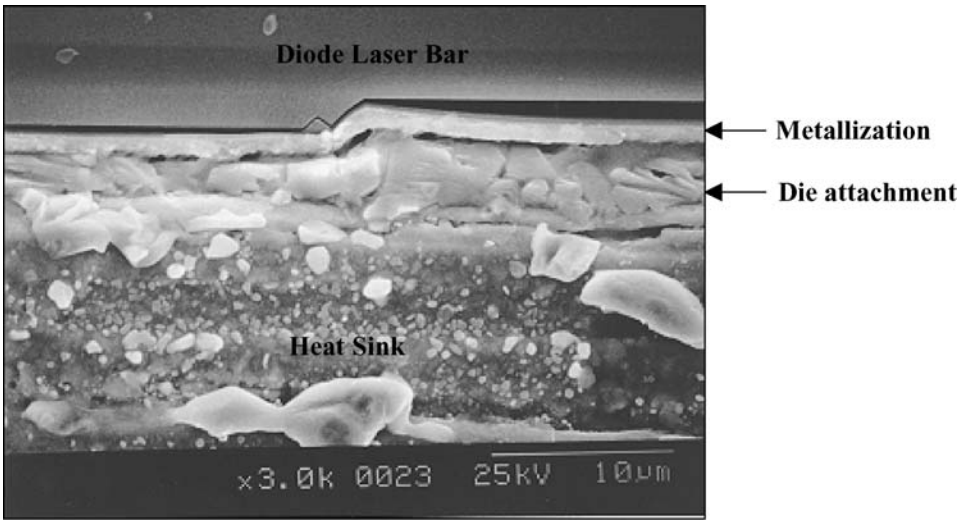


FIGURE 13.7. Delamination of metalization.

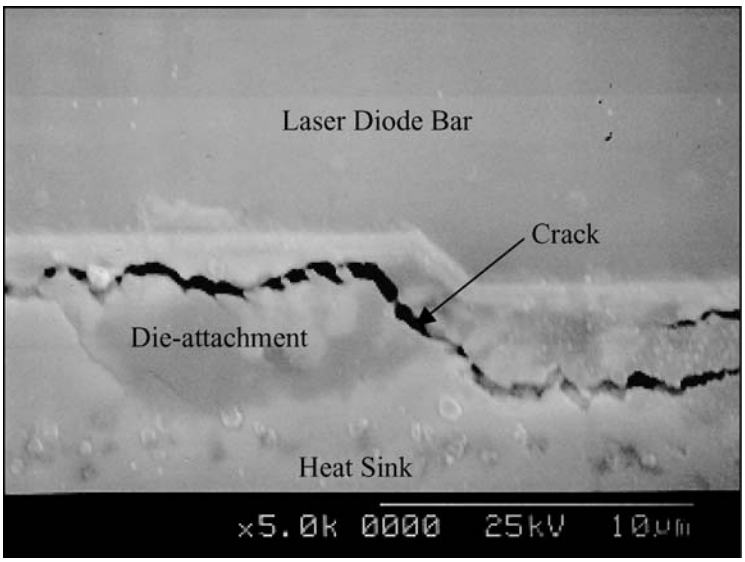


FIGURE 13.8. Crack running along the die-attachment.

bar and a copper heat sink causes excessive heating at the device junction, resulting in occasional burning at the bar emitter surface.

Figure 13.10 shows the output of the laser emission analyzer for a mounted bar, aged for 1000 hrs. Physical bending/displacement of the bar can be clearly seen. This is a representative case of various observations on different samples for which random bending, along with non-uniform optical emission from the bars across their widths, was observed. This problem, commonly known as the "smile problem," was observed to initiate during the die attachment process.

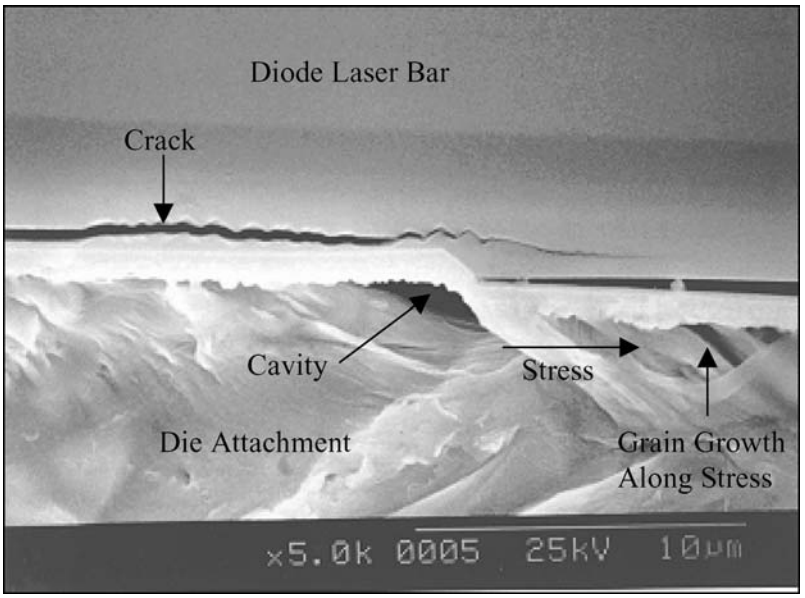


FIGURE 13.9. Higher concentration of physical defects and stress induced grain growth at ridges.

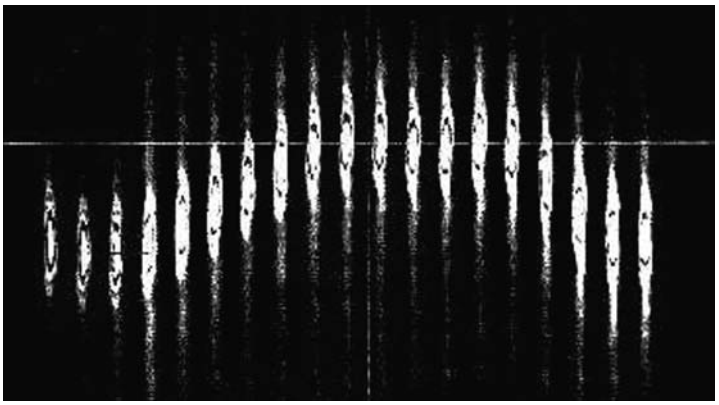


FIGURE 13.10. Beamview analyzer picture of the "Smile Problem" (cross-sectional view).

Interestingly, dynamic recrystallization was observed to present in the die-attachment at the interface, particularly in the 7000 hours aged samples. Since the heat generated is concentrated in narrow regions, significant thermal gradients contribute to the nucleation and grain boundary motion required for dynamic crystallization. Figure 13.9 shows grain growth in the direction of the shear stress. The crack seen in the bar is due to excessive stress.

Figures 13.11(a) and 13.11(b) show typical distributions of voids in a good and a defective bar, respectively. It was observed that voids present in the die-attachment region are distributed randomly, but typically, void density is relatively higher at the center of the bars (along length). The presence of voids directly under the facet region was observed to be responsible for the emitter burn spots, and occasionally, for vertical cracking of the bar.

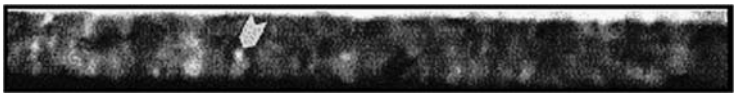


FIGURE 13.11a. SAM picture of randomly distributed (light regions) lesser voids in solder region in a good bar.

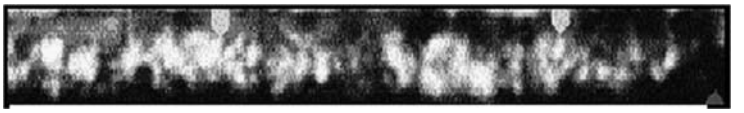


FIGURE 13.11b. SAM picture of randomly distributed (light regions) larger voids in solder region in a defected bar.

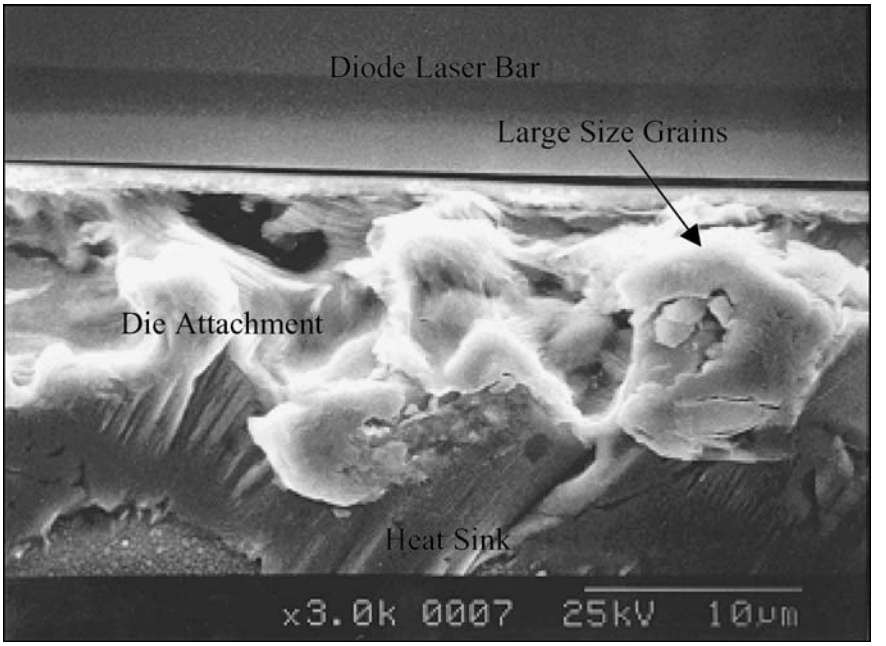


FIGURE 13.12. Large size grain growth in 7000 hours aged sample.

After an important synchronized set of observations, we conclude that both the increase in void density and the occurrence of delamination at the bar and die attachment interface increase with aging. Presence of voids is detrimental since they are the major barriers for efficient heat transfer and can cause accelerated failure of the packaged device. However, after the first few hundred hours of aging, few regions in the die-attachment appeared stable. We believe that this may be due to the plastic flow of die-attachment as a result of local heating.

Figure 13.12 shows the large grain growth for a sample aged for 7000 hours. The grain size at the interface of the copper (Cu) and indium (In) increased with aging time. We suggest that excessive heating due to insufficient physical contact between the device and heat sink gives rise to grain growth at the interface as a function of aging time.

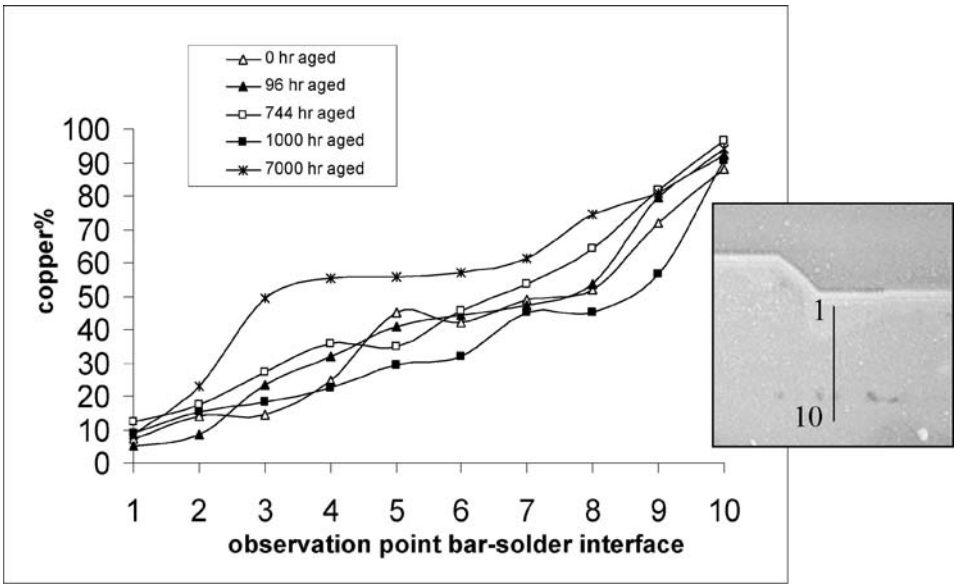


FIGURE 13.13. Copper concentration is highest near die-attach interface for 7000 hours aged sample.

As expected, physical defects were observed more frequently at the ends of the laser diode bars (along width), owing to higher stresses near the ends of the bars. Further, out of total 40 samples, in more than 90% of the samples it was observed that the die-attachment layer and the heat sink front surface were misaligned up to few microns along the length. This misalignment varied from sample-to-sample, and along the width of a bar.

13.2.3.2. Chemical and hardness observations Figure 13.13 provides energy dispersive X-ray analysis (EDS) data for aged samples. The measurements were performed on the die-attachment region. Observation points 1–10 on the x-axis represent sampling points where point 1 is nearest to the device-metalization interface and point 10 is close to the copper heat sink, within approximately 0.5 μm. The measurements were performed on various emitter regions across the bar to collect better statistics. We observed copper diffusion from the copper heat sink into the indium solder. We further observed that the diffused copper concentration and profile varies along the width of the bar (data not shown). Based on the Cu-In phase diagram and the previously discussed grain growth, we conclude that, at the interface, there is intermetallic formation, which is known to be hard and brittle and can contribute to crack development under stress conditions. We also observed that copper diffusion occurs at higher rates near the metalization-die attach interface during the initial hours of aging, and that the change is relatively small, though gradual, over the remaining aging period. A similar study was performed for gold diffusion into the die-attachment region. Gold diffusion into die-attachment was also observed to increase with aging (data not shown). However, due to the small thickness of the die-attach layer, X-ray diffraction (XRD) could not be employed to identify the various intermetallic phases in the die-attachment layer.

To confirm the increase in hardness caused by the intermetallic formation, we performed microhardness analysis on the interface metallurgy. Table 13.1 gives Knoop microhardness data, measured at 50 g load, as a function of aging. The hardness of the die-

TABLE 13.1.
Knoop micro-hardness as a function of aging.

Sample description	Average Knoop micro-hardness (at 50 g)
0 hr aged	157
96 hr aged	160
744 hr aged	196
1000 hr aged	203
7000 hr aged	214

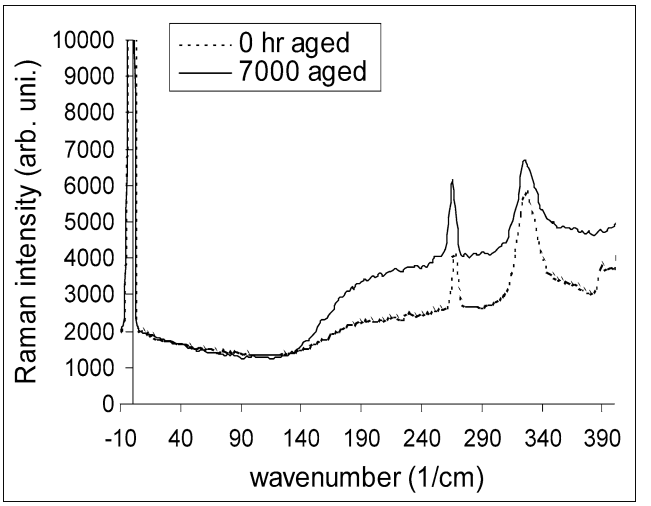


FIGURE 13.14. Typical profile of Raman peak shift of laser diode.

attachment region increased with aging time, which agrees with the observation of intermetallic phase formation as discussed previously.

13.2.3.3. Micro-Raman Observations Figure 13.14 shows a typical Raman spectrum in the active region of a high power laser diode studied in the present work. Micro-Raman measurements were performed on the optically sensitive GaAs device, across the width of the bar, at about 1 μm spacing. Sampling points 1 to 6 on the x-axis of the plot in Figures 13.15 and 13.16 represent measurement locations across the width of the front face of the active region of the GaAs bar. Point 1 is the point nearest the device-heat sink interface and point 6 is the farthest away. For analysis, we used the GaAs TO Raman peak at 269.3 cm^{-1} [16].

Figure 13.15 is a typical graph of micro-Raman shift and stress profile as a function of aging time. The graph clearly shows a shift in the peak, particularly toward lower wavenumbers for regions of the device near the die-heat sink interface, and high wavenumbers for the region away from the interface. The negative and positive shifts are a result of tensile and compressive stresses, respectively, in the active region of the laser diode array. Using the known pressure dependence of the TO-mode shift (0.5 cm^{-1}/100 MPa), these experimental mode shift profiles can be converted to stress profiles, and this is shown in Figure 13.15 [15,18]. A maximum tensile stress of ∼800 MPa and a compressive stress

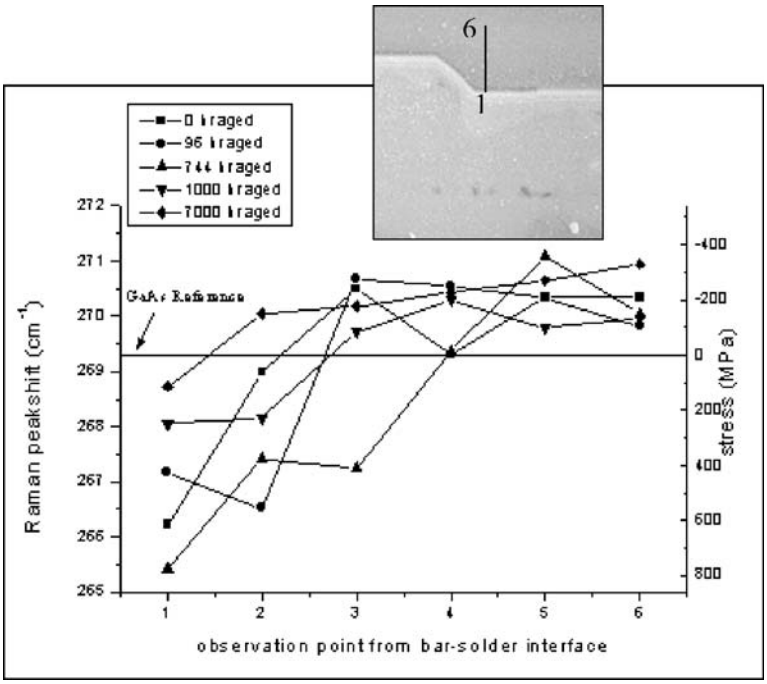


FIGURE 13.15. A typical profile of Raman peak shift and stress as a function of distance from bar–solder interface.

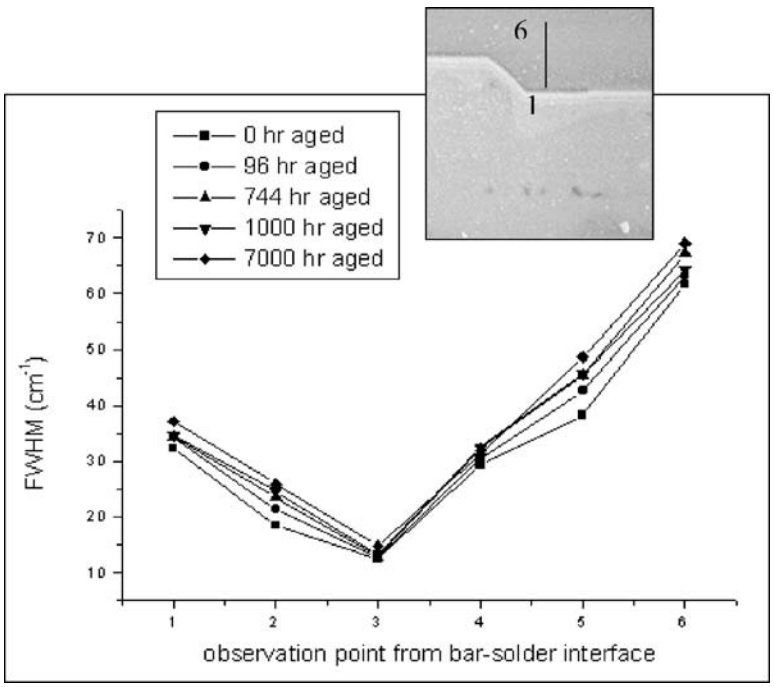


FIGURE 13.16. A typical graph of FWHM as a function of distance from bar–solder interface.

of ∼420 MPa are observed. Two stress zones, one under compressive and the other under tensile stress, can be seen. The zone near the GaAs bar-heat sink interface is under tensile stress, while the zone near the quantum-well active layer is under compressive stress. As a function of aging there is measurable change in the tensile stress value, unlike little or no change in the compressive stress value near active layer. This significant change in the tensile stress value has clearly affected reliability of the packaged bar. Further the randomness the variation, we believe is caused by the multiple defects that can arise during aging and needs more detail investigation. Also, we believe that, under the influence of thermal cycles during laser diode operation, the hardness of the die-attachment increases, which causes the tensile stress to decrease [30].

The FWHM of the GaAs peak is an indicator of the amount of disorder present in the lattice. The disorder can be due to rearrangement of atoms under the influence of high temperatures and increasing threshold current. Hence, the magnitude of the FWHM is related to lattice degradation caused by thermo-mechanical stress. From Figure 13.16, we can see that the FWHM is high ($\sim$35 cm^{-1}) at the bar-die-attachment interface, decreases as we move further into the bar, and is smallest ($\sim$12 cm^{-1}) near the active layer, which is the quantum-well structure. We can infer that the quantum-well structure is more robust. This observation is also in contrast to the usual notion that in the quantum-well structure, where maximum intermixing is present, disorder, and hence, FWHM should be maximum [27,31,32]. We believe that the lower FWHM at the quantum well region is an artifact of confinement. It is observed that, although the stress decreases as we go from bar-die-attachment interface into the substrate, the FWHM first decreases and then increases. This observation is surprising since it is contrary to the commonly held belief that aging related defects in the crystal are predominantly observed near the bar-package interface and decrease into the substrate. Although the change is small, it is clear from Figure 13.16 that the FWHM, and hence, the amount of disorder in the lattice increases with aging.

It is important to note that both the stress-induced peak shift and the FWHM increase as we approach the ridge edges from both sides. The camel hump-like stress profile, which is characteristic of ridge lasers, was not observed [15,18]. This is due to a very high width to height ratio of the emitter. It is also worth noting that the shape and strength of the stress profile curve and the FWHM change from one emitter to another in an array.

13.3. CONCLUSIONS

In summary, it is demonstrated that, in high power laser diode systems, various materials, optical, mechanical, and thermal parameters are inter-related and form the basis of a synergistic analysis of the failure modes. Complementary analytical techniques have been applied to evaluate various unique nanostructures, physical defects at the die attach-device interface and aging-induced grain growth and change in the chemistry of the interface metallurgy. It is demonstrated that both physical defects and microstructural changes in the die-attach, during device operation, affect the reliability of packaged high power laser bars. It is observed significant copper diffusion into the die attach layer, which caused the formation of hard and brittle intermetallics. Further, it is concluded that stress relief during aging and excess heating caused by inadequate physical contact between the bar and the heat sink result in uneven bending of the bar causing non-uniform optical alignment. It was demonstrated that micro-Raman spectroscopy is a very useful tool for studying packaging-induced local mechanical stress. Although there is certain failure related randomness in the interface and active regions, a definite failure trend exists as the package undergoes aging.

Physical defects, optical response, and chemical composition analysis, along with micro-hardness measurements, provide valuable insight into packaging and operation-induced defects affecting the reliability of high power laser diodes. It is believed that the results of the work discussed here should ultimately provide assistance in predicting packaging failure modes that will lead to better processing and packaging schemes in the future. The ability to map the distribution of stress within the device and then correlate the resulting data with points of failure should lead to more reliable design and manufacturing.

A more fundamental study of the evolution of die-attachment metallurgy, in which intermetallic formation, together with softening effects and the finite size under the thermal fatigue conditions are considered, is essential. Such a study should provide information that lead to a much better understanding of the random and unpredictable behavior of some of the failure mechanisms observed in this work. In addition, *in situ* micro-Raman analysis, along with photoluminescence measurements, and numerical modeling during the aging process are essential to a better understanding of the gradual changes in the stresses as a function of aging time.

ACKNOWLEDGMENTS

One of the authors acknowledge Ajit Dhamdhare (previously at the University of Arkansas) for his MS contributions, and Dr. John Nightingale, Mr. Robert Miller, and Mr. John Morales of Coherent, Inc. for valuable technical discussions. We he wishes to acknowledge Dr. Richard Bormett and Ms. Diane Allen of Renishaw Inc. and Dr. John Shultz of Arkansas Analytical Laboratory for micro-Raman and other analytical support work.

REFERENCES

1. C.E. Ho, W.T. Chen, and C.R. Kao, Interactions between solder and metallization during long-term aging of advanced microelectronic packages, Journal of Electronic Materials, 23(7), pp. 379–385 (2001).
2. P.L. Tu, Y.C. Chan, and J.K.L. Lai, Effect of intermetallic compounds on the thermal fatigue of surface mount solder joints, IEEE Transactions on Components Packaging and Manufacturing Technology, Part B, 20(1), pp. 87–92 (1997).
3. S.A. Merritt, P.J.S. Heim, S.H. Cho, and M. Dagenais, Contrlled solder interdiffusion for high power semiconductor laser diode die bonding, IEEE Transactions on Components Packaging and Manufacturing Technology, Part B, 20(2), pp. 141–145 (1997).
4. J.M. Parsey, S. Valocchi, W. Cronin, J. Mohr, B.L. Scrivner, and K. Kyler, A metallurgical assessment of SnPbAg solder for GaAs power devices, Journal of Materials, 31(3), pp. 28–31 (1999).
5. W.L. Phillipson, and D.J. Diaz, Microstructural examination and failure analysis of microelectronic components, Materials Developments in Microelectronics Packaging Conference Proceedings, 1991, pp. 289–297.
6. A. Zubelewicz, R. Berriche, L.M. Keer, and M.E. Fine, Lifetime prediction of solder materials, Journal of Electronic Packaging, 111, pp. 179–182 (1989).
7. N. Strifas, and A. Christou, Die attach adhesion and void formation at the GaAs substrate interface, Mat. Res. Soc. Symp. Proc., 356, pp. 869–874 (1995).
8. N. Zhu, Thermal impact of solder voids in the electronic packaging of power devices, IEEE 15th SEMI-THERM Symposium, 1999, pp. 22–29.
9. W. Nakwaski, An additional temperature increase within GaAs/(AlGa)As diode lasers caused by the deterioration of an indium solder, Electron Technology, 23(1/4), pp. 33–38 (1990).
10. V.D. Bo, F. Bartels, A. Scandurra, Ch. Luchinger, and S. Radeck, More insights into the soft die attach, 11th European Microelectronics Conference, 1997, pp. 308–319.
11. Q. Tan and Y.C. Lee, Soldering technology for optoelectronic packaging, IEEE Electronic Components and Technology Conference, 1996, pp. 26–36.

12. A.Y. Kuo and K.L. Chen, Effect of thickness on thermal stresses in a thin solder or adhesive layer, Conference Proceedings of ASME Winter Annual Meeting, 1991, pp. 1–6.
13. L. Wen, G.R. Mon, and R.G. Ross, Design and reliability of solders and solder interconnections, Symposium Proceedings of TMS Annual Meeting, 1997, pp. 219–226.
14. M.E. Polyakov, Mechanical stresses in AlGaAs/GaAs quantum-well heterolasers, Sov. J. Quantum Electronics, 19(1), pp. 26–29 (1989).
15. P.W. Epperlein, Temperature, stress, disorder and crystallization effects in laser diodes: measurements and impacts, Proc. SPIE, 3001, pp. 13–28 (1997).
16. P. Puech, G. Landa, R. Carles, and C.J. Fontaine, Strain effects on optical phonons in <111> GaAs layers analyzed by Raman scattering, J. Appl. Phys., 82(9), pp. 4493–4499 (1997).
17. E. Anastassakis, Stress measurements using Raman scattering, Analytical Techniques for Semiconductor Materials and Process Characterization: Proceedings of the Satellite Symposium to ESSDERC 1989, Berlin, 1989, pp. 298–326.
18. P.W. Epperlein, G. Hunziker, K. Datwyler, U. Deutsch, H.P. Dietrich, and D.J. Webb, Mechanical stress in AlGaAs ridge lasers: its measurement and effect on the optical near field, Proceedings of the 21st International symposium on Compound Semiconductors, Vol. 21, 1994, pp. 483–488.
19. R. Puchert, J.W. Tomm, A. Jaeger, A. Barwolff, J. Luft, and W. Spath, Emitter failure and thermal facet load in high power laser diode arrays, Appl. Phys. A, 66, pp. 483–486 (1998).
20. I. De Wolf, J. Chen, M. Rasras, W.M. van Spengen, and V. Simons, High-resolution stress and temperature measurements in semiconductor devices using micro-Raman spectroscopy, Proc. SPIE, 3897, pp. 239–252 (1999).
21. I. DeWolf and H.E. Maes, Mechanical stress measurements using micro-Raman spectroscopy, Microsystems Technologies, 5, pp. 13–17 (1998).
22. A. Barwolff, J.W. Tomm, R. Muller, S. Weiss, M. Hutter, H. Oppermann, and H. Reichl, Spectroscopic measurement of mounting-induced strain in optoelectronic devices, IEEE Transactions on Advanced Packaging, 23(2), pp. 170–175 (2000).
23. J.W. Tomm, R. Muller, A. Barwolff, T. Elsaesser, A. Gerhardt, J. Donecker, D. Lorenzen, J. Daiminger, S. Weiss, M. Hutter, E. Kaulfersch, and H. Reichl, Spectroscopic measurement of packaging-induced strains in quantum well laser diodes, J. Appl. Phys., 86(3), pp. 1196–1201 (1999).
24. D. Wood, G. Cooper, D.J. Gardiner, and M. Bowden, Raman spectroscopy as a mapping tool for localized strain in microelectronics structures, Journal of Materials Science Letters, 16(14), pp. 1222–1223 (1997).
25. P.W. Epperlein, Raman spectroscopy of semiconductor lasers, Proceedings of Conference on Lasers and Electro-Optics, Vol. 9, 1996, pp. 108–109.
26. B. Dietrich and K.F. Dombrowski, Experimental challenges of stress measurements with resonant micro-Raman spectroscopy, Journal of Raman Spectroscopy, 30, pp. 893–897 (1999).
27. P.S. Pizani, F. Lanciotti, Jr., R.G. Jasinevicius, J.G. Duduch, and A.J.V. Porto, Raman characterization of structural disorder and residual strains in micromachined GaAs, Journal of Applied Physics, 87(3), pp. 1280–1283 (2000).
28. K. Iizuka, T. Yoshida, I. Matsuda, H. Hirose, and T. Suzuki, Micro-Raman study of the residual stress in molecular-beam-epitaxy-grown $Al_xGa_{1-x}As$/GaAs multilayer structures, Materials Science and Engineering, B5, pp. 261–264 (1990).
29. J.P. Landesman, A. Flore, J. Nagle, V. Berger, E. Rosencher, and P. Puech, Local stress measurements in laterally oxidized GaAs/$Al_xGa_{1-x}As$ heterostructure by micro-Raman spectroscopy, Appl. Phys. Lett., 71(17), pp. 2520–2522 (1997).
30. R.R. Varma, Bonding induced stress in semiconductor laser, Proceedings of IEEE Electronic Components and Technology Conference, 1993, pp. 482–484.
31. A.S. Helmy, A.C. Bryce, C.N. Ironside, J.S. Aitchison, and J.H. Marsh, Raman spectroscopy for characterizing compositional intermixing in GaAs/AlGaAs heterostructures, Appl. Phys. Lett., 74(26), pp. 3978–3980 (1999).
32. G. Attolini, L. Francesio, P. Franzosi, C. Pelosi, S. Gennari, and P.P. Lottici, Raman scattering study of residual strain in GaAs/InP heterostructures, J. Appl. Phys., 86(11), pp. 6425–6430 (1994).

14

Review of the Technology and Reliability Issues Arising as Optical Interconnects Migrate onto the Circuit Board

P. Misselbrook[a,c], D. Gwyer[b], C. Bailey[b], P.P. Conway[c] and K. Williams[c]

[a]*Celestica, Kidsgrove, Stoke-on-Trent, UK*
[b]*Centre for Numerical Modelling and Process Analysis, University of Greenwich, London, UK*
[c]*Interconnection Group, Loughborough University, Loughborough, UK*

Abstract

Light has the greatest information carrying potential of all the perceivable interconnect mediums; consequently, optical fiber interconnects rapidly replaced copper in telecommunications networks, providing bandwidth capacity far in excess of its predecessors. As a result the modern telecommunications infrastructure has evolved into a global mesh of optical networks with VCSEL's (Vertical Cavity Surface Emitting Lasers) dominating the short-link markets, predominately due to their low-cost. This cost benefit of VCSELs has allowed optical interconnects to again replace bandwidth limited copper as bottlenecks appear on VSR (Very Short Reach) interconnects between co-located equipment inside the CO (Central-Office).

Spurred by the successful deployment in the VSR domain and in response to both intra-board backplane applications and inter-board requirements to extend the bandwidth between IC's (Integrated Circuits), current research is migrating optical links toward board level USR (Ultra Short Reach) interconnects. Whilst reconfigurable Free Space Optical Interconnect (FSOI) are an option, they are complicated by precise line-of-sight alignment conditions hence benefits exist in developing guided wave technologies, which have been classified into three generations. First and second generation technologies are based upon optical fibers and are both capable of providing a suitable platform for intra-board applications. However, to allow component assembly, an integral requirement for inter-board applications, 3rd generation Opto-Electrical Circuit Boards (OECB's) containing embedded waveguides are desirable.

Currently, the greatest challenge preventing the deployment of OECB's is achieving the out-of-plane coupling to SMT devices. With the most suitable low-cost platform being to integrate the optics into the OECB manufacturing process, several research avenues are being explored although none to date have demonstrated sufficient coupling performance. Once in place, the OECB assemblies will generate new reliability issues such as assembly configurations, manufacturing tolerances, and hermetic requirements that will also require development before total off-chip photonic interconnection can truly be achieved.

14.1. BACKGROUND TO OPTICAL INTERCONNECTS

Whilst there are many important characteristics of data carrying interconnects such as security, speed, and reliability, independent of link distance: the fundamental link property is unquestionably information-carrying capacity. As demonstrated by the Shannon-Hartley theorem, capacity is proportional to the channel bandwidth, which in turn is proportional to the frequency of the carrier [4]. This formula, drawn from information theory, is true regardless of specific technology and highlights the issue that the bandwidth capacity of transmission mediums is ultimately limited, not by technological advances, but by frequency, a physical property of the medium itself. Defined by this, copper, the foremost transmission medium, has the lowest potential bandwidth, increasing with twisted pairs, RF, and microwaves (satellite channels), through to light that has the highest frequency and therefore the greatest bandwidth potential [4].

The vast bandwidth potential offered by optical links over more traditional electrical links first became a commercial reality during the 1970s as technological advances, such as the development of edge-emitting diodes and single mode fibers, enabled optical links to supersede copper in long distance telecommunication links [5–7]. Through the 1980s and 90s, as demand on link capacity increased across the network, copper became increasingly redundant over reducing distances as it struggled to provide for the bandwidth explosion generated by three driving factors: the increasing base of global end-users, popularity of technologies such as the Internet, and the emergence of data intensive multimedia services such as video conferencing [5]. The modern telecommunications infrastructure is a global mesh of optical networks offering a plethora of multimedia services. Therefore to ensure global compatibility, various transmission standards have been adopted such as SONET, ATM, and Ethernet. These standards govern specifications from data protocol to loss budgets for each of the individual interconnecting networks, which are typically organized by function and link distance into three market-segments: long-haul, Metropolitan Access Networks (MAN), and local access networks. This hierarchy allows for the aggregation of the lower bandwidth access traffic, generated by the user, through the regional MAN to the corresponding MAN CO, also called the Point of Presence (POP) [8,9]. Each CO contains switches and routers that interconnect with other POP's through the long-haul network to provide complete inter-networking of all end users.

Although the link distances reduce, each market segment cannot simply be a scaled down version of the larger due to the varying requirements based on the traffic each network handles. During their evolution, each market segment has therefore developed specific requirements on the cost and performance of the transmission equipment utilized.

14.2. TRANSMISSION EQUIPMENT FOR OPTICAL INTERCONNECTS

Long-haul networks span both regional and extended geographic distances, connecting MANs to extend global connectivity between regional domains [10,11]. Due to the long distances involved and the deployment of Dense Wavelength Division Multiplexing (DWDM) systems, to provide for the huge bandwidth required at each link, high performance single-mode transmitters are required (Figure 14.1). For this reason, edge emitting Distributed Feedback (DFB) laser diodes transmitting around the 1550 nm wavelength have established themselves as the technology of choice for long-haul applications, giving unrivaled performance in areas such as: single mode stability, power output, line-width, and

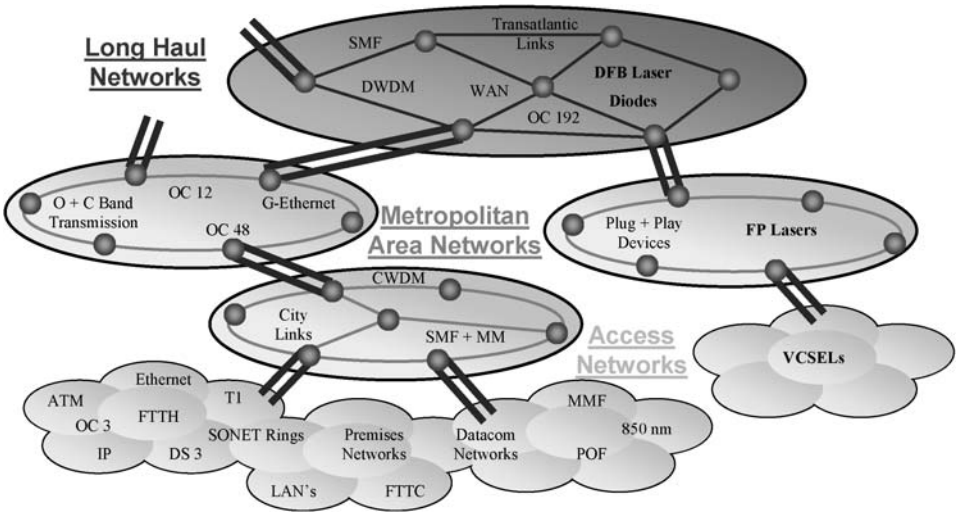


FIGURE 14.1. Optical network hierarchy: The three market segments, long-haul, MAN, and access networks interconnect at CO switches to route traffic between global end users.

wavelength selection [11]. Achieving such performance comes at a trade-off with price, with components being dominated by performance considerations rather than cost.

Metro networks operate over much reduced link distances and provide the connection between Local Area Networks (LAN) or access networks to each other and to the backbone for global communications between end users. Although DFB lasers meet many of the performance requirements of the metro space, they are uneconomical due to the cost of packaging the components into industry standard 14-pin butterfly devices. Consequently cheaper Fabry-Perot (FP) lasers transmitting at the shorter 1310 nm wavelength are more commonly used as link distances decrease [12].

As access networks have predominately the shortest link-lengths of the three telecommunication network segments, between several hundred to several thousand meters, interconnects deployed here are much more sensitive to cost [13]. Thus, a high premium for additional performance would not persuade operators to switch to optical channels from simpler, and cheaper copper-based interconnects with proven reliability. From this standpoint the most important breakthrough in allowing optical interconnects into the domain of shorter link lengths was the development of the VCSEL in the late '90s. Operating in the 850 nm window, short-wavelength multimode VCSEL's provide a very cost effective solution with ample performance density to operate over the short link lengths involved in access networks. The fundamental difference between edge emitting diodes and surface emitting VCSEL's is just that; VCSEL's emit the light beam from the surface of the wafer where as previously the light was generated in the plane of the wafer, only becoming accessible after wafer dicing.

VCSEL construction has seen significant development [14–16] since their inception from the first design with the active area sandwiched between two Distributed Bragg Reflector (DBR) mirror stacks, fabricated predominately on gallium arsenide (GaAs) wafers. The light beam is created when electrical current is applied across the active layer, via intra-cavity contacts, generating photons that are then reflected by the DBR mirrors before being emitted from a circular aperture, about 14 μm in diameter [17], on the top surface

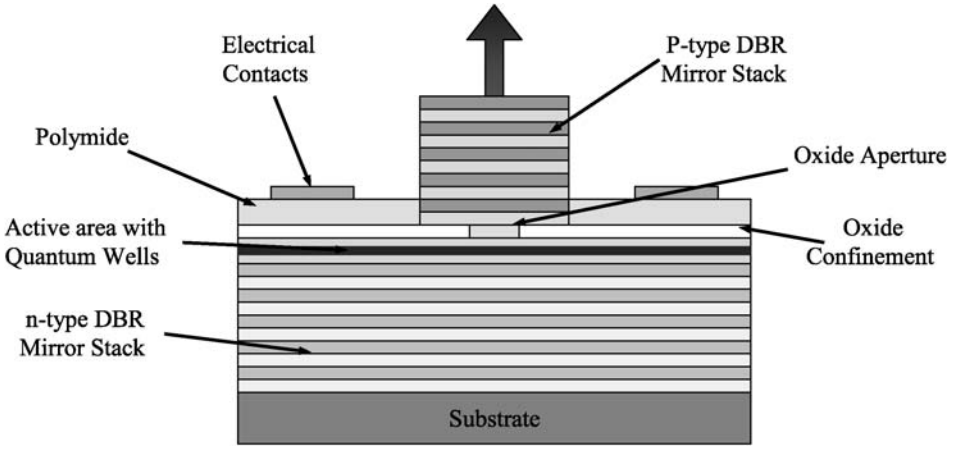


FIGURE 14.2. VCSEL cross-section: Layers of material are thinly deposited onto GaAs wafers to form DBR mirrors sandwiched either side of an active area. The quantum wells convert electrical energy into photons, emitting a light beam perpendicular to the surface of the substrate.

of the wafer. Before the wafer is diced a polyimide coating is applied between the VCSEL structures to protect the sides from oxidization [4]. A single VCSEL die is typically 250 μm square although wafers can be diced into any pattern or array. The electrical contacts can be created to enable the die to be wire bonded or flip-chip bonded with the aperture on the top or bottom side (Figure 14.2). The unique manufacturing process of VCSEL devices has enabled them to dominate the optical access market in recent years, their many inherent advantages include: low-cost due to their ability for high volume manufacture and in-production testing at the wafer level, ultra-high modulation rates, low power consumption with threshold currents less than a milliamp, and low coupling tolerances due to the circular beam output.

Although DFB, FP, and VCSEL transmission lasers dominate their respective markets, the downturn in the telecommunications industry, just after the turn of the century, has resulted in a necessary rationalization in the performance and packaging of components. This has opened the door for new technologies and the spread of the developed sources into new, previously inhibited sectors. The most important issue that this rationalization addressed was an effort to reduce costs across all parts of the manufacturing process. It is widely accepted that of the final laser module's cost, packaging constitutes between 85%, for laser diodes in butterfly packages [11], and 33%, for lower specification VCSEL's [18], hence laser module packaging has been most affected by cost reduction initiatives. This has primarily seen an underlying trend across the market segments toward vertical integration into standardized transceiver packages, which incorporates the transmitter, receiver, and electronics into a standard Small Form Factor Pluggable (SFP) module [19].

This drive to reduce the cost of optical components has also had additional benefits, allowing VCSEL's to become competitively priced offering an alternative to copper in VSR interconnects, between 10–300 m, as copper yet again creates bottlenecks in the COs and POPs of telecommunication networks. As explained further below, the successful emergence of VCSEL's in the VSR arena has also renewed the long anticipated wait for optical solutions to USR interconnects, distances less than 10 m and predominantly based on circuit boards.

14.3. VERY SHORT REACH OPTICAL INTERCONNECTS

The expanding demand on networking capacity has meant service providers need to connect core routers and optical transport equipment with multiple high-speed links to prevent bottlenecks developing at the POP. Traditionally, these interconnects have been deployed across copper based-interconnects, but the move toward data rates of 10 Gbps and beyond is pushing these links to their limitations, so increasingly optical interconnects are being considered. Since the majority of POP equipment tends to be physically located within the same building, a significant proportion of these links are less than 300 m, where it is uneconomical to deploy optical interconnects operating over standards optimized for longer distances [3]. Subsequently, a set of VSR optical interconnection standards has been developed by the Optical Networking Forum (OIF) aimed at low-cost interconnects between co-located equipment (<300 m) [20].

VSR leverages technology developed for Gigabit-Ethernet, to transport OC-192 (10 Gbps) traffic. The first implementation agreement developed for the VSR arena was OIF-VSR4-01, which employs parallel optics with the signal distributed over twelve 1.25 Gbps VCSEL channels operating over multimode fiber (MMF). As an alternative to the 12 fiber parallel optics standard, OIF-VSR4-03 a four-fiber standard was implemented with the VCSEL's operating at 2.5 Gbps. For VSR applications where single mode fiber (SMF) has already been laid, a significant cost advantage is made by utilizing this, so a standard was also developed for 1310 nm VCSEL's and FP lasers over SMF, which has now been developed as OIF-VSR4-05 (Figure 14.3). Finally, in appreciation of the advances being made in VCSEL technology, the set of OC 192 VSR implementation agreements was completed by OIF-VSR4-04, based on a single 850 nm MM fiber channel, for distances up to 85 m.

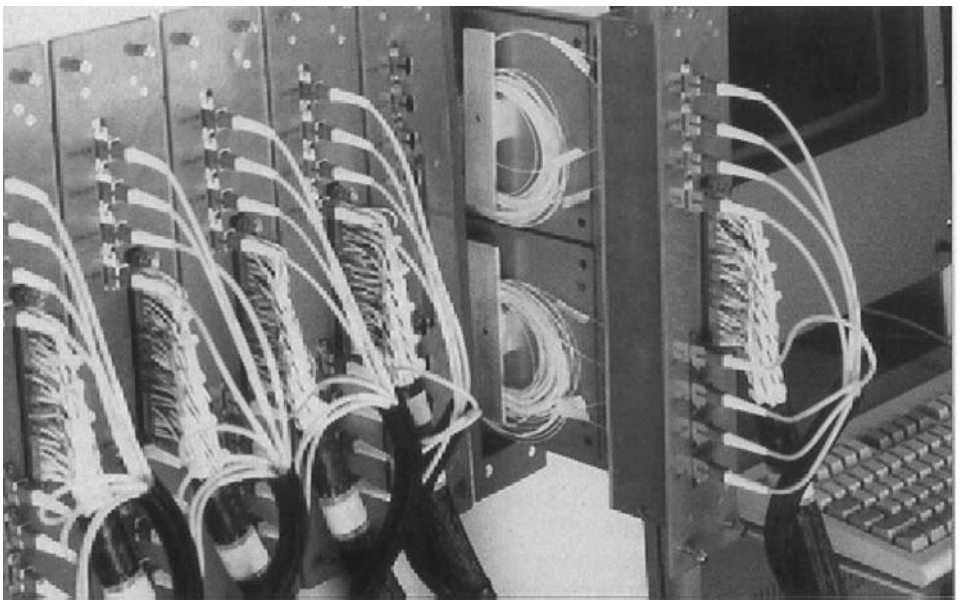


FIGURE 14.3. VSR optical interconnects: Discrete optical fibers connect co-located equipment over distances up to 300 m, using OIF implementation agreements. Figure courtesy of BPA [2].

This comprehensive set of VSR standards has enabled low-cost optical interconnects, based on VCSEL technology, to replace copper as the technology of choice for high-speed connections inside the CO and POP. However, this migration from long haul to VSR, despite repeated predictions, has taken over two decades. The main reasons for this delay have been in over-coming the technological challenge of manufacturing optical interconnects with strong manufacturability, greater performance, and at a lower cost than electrical interconnects. Now that this has been achieved down to a few hundred meters, the push is on to produce board-level USR optical interconnects for distances less than 10 m. The drive toward this goal is currently coming from two main directions, although the end impact of low-cost on-board optical interconnects is expected to have a much broader impact across the electronics market. The first factor stems from the successful implementation of VSR optical links between telecommunications equipment, which has now lead to a bottleneck forming on the backplanes. Low-cost inter-board optical links to transport OC 192 traffic between daughter-cards in a rack configuration are required.

The second driving force toward board-level USR optical interconnects is concerned with intra-board applications, specifically the increase in bandwidth capacity between IC's, which is being identified as an issue as chip manufacturers commit to the extension of Moore's law well into the future [21]. Based on this trend, the doubling of transistors on IC's every eighteen months, it is predicted that within five years CMOS-based transistors will be fast enough for transceivers to operate at clock speeds of around 14 GHz with data-transfer rates in the region of 20 Gbs [22]. Since copper interconnects on FR4 based Printed Circuit Boards (PCB's) become bandwidth limited beyond 10 GHz, primarily due to frequency dependent losses such as the skin effect in conductors and dielectric losses in the substrate, manufacturers have sought to address the bandwidth limitation of copper traces. However, the potential copper-trace based solutions such as new board substrates and sophisticated encoding techniques tend to be expensive making optical solutions increasingly attractive.

History shows that the barrier to implementing optical links over established copper-based interconnects is overcoming the technical challenges to achieve higher performance for lower cost. These technology issues are similar for both the driving forces behind USR optical interconnects and can currently be divided into two research areas: free space optics, and guided waves.

14.4. FREE SPACE USR OPTICAL INTERCONNECTS

The most basic, although by no means the simplest method of applying optical interconnects at the board level is to use lenses and collimators to expand a VCSEL beam suitably so that it can be sent through the air to a corresponding configuration, for detection by a Photo Detector (PD) [23–26]. So called FSOI's can boost the bandwidth between chips due to the combination of both high data rates and the fact that FSOI's can be densely packed on the circuit board. The addition of diffractive optics to the configuration allows the signal to be routed to different detectors, so FSOI's can benefit from being reconfigurable [27–29].

Complications to this basic principal develop due to the requirement to maintain a precise line-of-sight rule, in that if the VCSEL's and PD's become misaligned or blocked for any reason the signal is lost. Although research is ongoing into ways of tracking and actively maintaining the links as the transmitter and detector arrays move relative to one

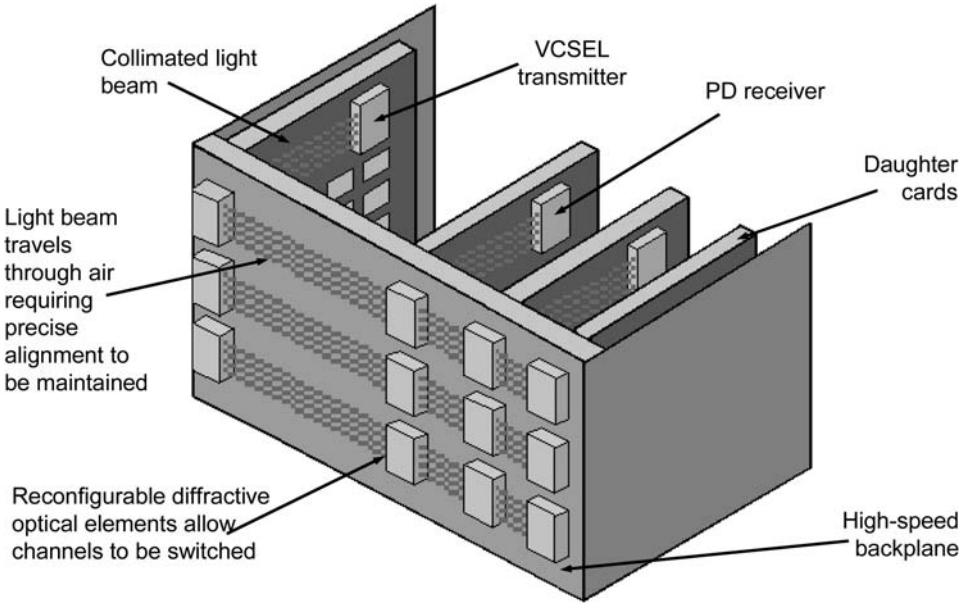


FIGURE 14.4. FSOI based backplanes: VCSEL's and PD's on separate daughter cards communicate across an optically enabled high-speed backplane via free space channels.

another [30,31], the occurrence of catastrophic signal loses due to environmental issues can not be eliminated without hermetically sealing the entire backplane, which would certainly compromise the cost advantage. Greater control of the signal between the transmitter and receiver is beneficial and can be achieved by guiding it between geometric boundaries such as a fiber or waveguide, as is the case in telecommunications systems. Consequently, there is an increased interest in developing PCB based guided wave USR optical interconnect solutions that permit greater freedom in the routing between devices, without strict line-of-sight rules, and using fewer elements than FSOI systems (Figure 14.4).

14.5. GUIDED WAVE USR INTERCONNECTS

Optical fiber and planar waveguides are two examples of guided wave interconnects. Although the end product and manufacturing processes are markedly different, with fibers having a circular cross section and waveguides square, the physics behind light transmission is similar. The basic principal consists of the light signal propagating down a core that is surrounded by a cladding material with a slightly lower refractive index. For multimode systems the light is retained in the core material due to Total Internal Reflection (TIR) at the core-cladding interface.

To date, the most successful uptake of board-level guided wave USR interconnects has been on high-speed backplanes inside CO telecommunication equipment. As identified by BPA [32], there are three methods for the implementation of these interconnects over backplanes. The 1st generation consists of discrete optical fiber interconnects. Each fiber is individually routed at the back of the switch for point-to-point optical links, a direct descendant of the VSR links that are currently being deployed to interconnect backplanes. However, although separate fiber interconnects have been widely implemented across VSR

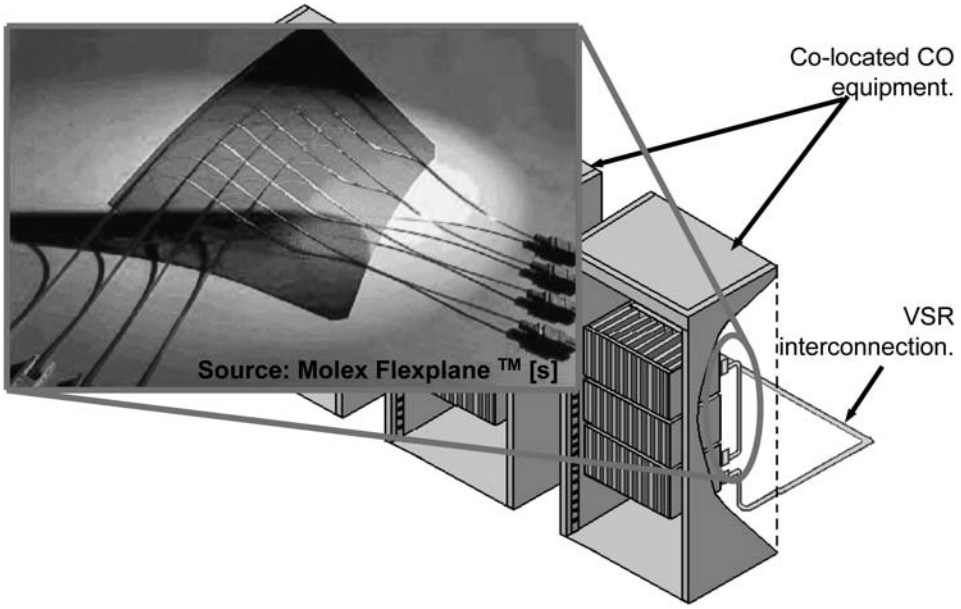


FIGURE 14.5. FOB USR technology: Inter board fiber management problems are reduced by combining the fiber into a flexible harness or PCB. Inset, an example of Molex's Flexplane™.

links between systems and backplanes, attempting to extend 1st generation technology to address links across the backplane have been thwarted by fiber management problems generated by the resulting increase in the density of fiber terminations. This greater density creates a "rat's nest" of fibers at the backplane and also requires significant resources to manually route the fibers, outside of the manufacturing environment.

To enable greater integration of the USR optical interconnects into the board and to counter some of the problems associated with discrete fiber interconnects, various institutions and companies [1,33,34] are developing so called 2nd generation technologies which combine the fiber into a rigid board, termed fiber-on-board (FOB). The essence of FOB technology is to embed standard optical fiber into a PCB harness (Figure 14.5). Whilst fiber handling and its associated costs are clearly still involved in the manufacturing process, the method is capable of yielding low-loss connections with relatively few technology hurdles. Fiber management still creates problems though, particularly in maintaining the minimum bend radius of the fiber, typically a few centimeters, which if exceeded would result in unacceptable attenuation due to the conditions for TIR being negated. Exceeding the minimum bend radius will mostly occur when routing the fibers around the board and prevents more than two fibers from crossing on the same plane and also fibers from being bent out-of-plane to the surface of the board, which is required if optical Surface Mount Technology (SMT) components are to be assembled. Coupling light into the embedded fibers is further complicated due to issues concerned with mounting and joining the output of transmitter and receivers to non-connectorized fiber terminations. Predominately non-connectorized fibers are spliced together using a fusion process, however this process requires a reasonable amount of accessible fiber preventing splices in close proximity to the board edges. This also limits the number of rework options available, since if a fiber exiting the board is

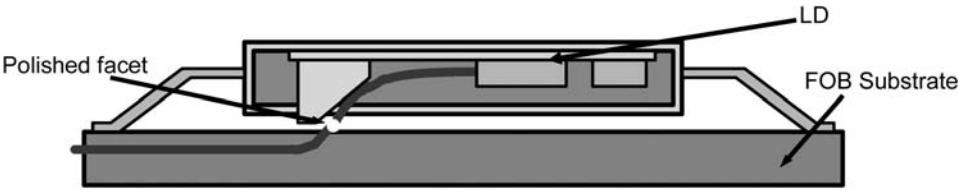


FIGURE 14.6. FOB component coupling issues: Maintaining the minimum bend radius inhibits the in laid fiber from being bent upward. The assembly also requires precise cleave and polish location of fiber termination.

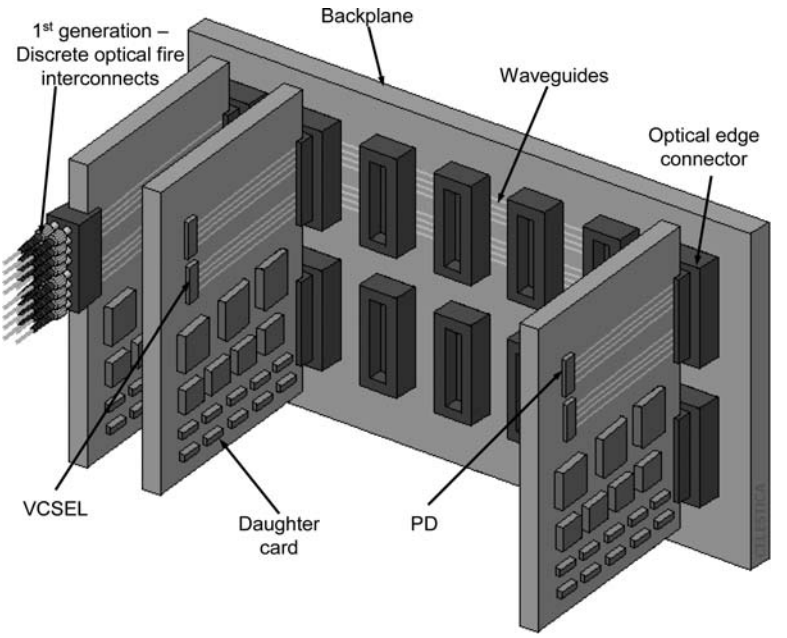


FIGURE 14.7. OECB based optical backplane architecture: SMT assembled VCSEL's emit light into embedded waveguides, which route onto the backplane to be distributed onto other daughter cards or in to fiber for aggregation over larger areas.

broken within a close proximity to the PCB harness the entire board must be re-fabricated (Figure 14.6).

Due to the associated costs and problems with fiber management and component coupling, a significant cost advantage can be realized by eliminating fibers from the heart of the manufacturing process. This is the main advantage in 3rd generation technologies, which look to replace fibers with planar optical waveguides. Waveguides allow the optical links to be fully integrated into the PCB, with easier coupling routes for SMT devices, which allows the solution to address inter-board as well as intra-board interconnects. It is anticipated that most 3rd generation technologies will sandwich an optical layer, containing the waveguide, into a standard FR4-stack. The board will remain FR4 based since not all interconnects are required to be high speed so some copper tracks will be retained thus creating a hybrid OECB with both electrical and optical interconnects on one board. OECB fabrication builds on a core FR4 board produced using current fabrication methods with the copper tracks being routed through layers in the board by vias and blind vias (Figure 14.7).

There are currently two main candidate materials being investigated to form the optical waveguide layers: glass, and polymer. BPA's backplane report suggested that to achieve the necessary attenuation (less than 5 dB/cm), glass waveguides would have to be adopted due to their reduced material absorption when compared to polymers [2]. However, against this prediction, glass manufacturing issues have prevented the production of high quality waveguides with low attenuation levels, whilst developments in the manufacture of polymer based layers have produced attenuation figures under 0.03 dB/cm [35] and are therefore becoming the technology of choice. Polymer waveguide processing techniques also have the advantage of being compatible with current PCB fabrication methods, allowing low-cost manufacture, an important specification for OECB's.

With the addition of optical edge connectors, OECB based interconnects enable high-speed point-to-point optical links to be established between transmitting and receiving equipment satisfying the specification for both inter-board optical backplanes and high bandwidth inter-board connections between ICs. Due to their inherent advantages, the technology set of choice is anticipated to comprise of readily available VCSEL and PD pairs, with the light signals being routed between the pairs via embedded waveguide structures, which are becoming commercially available through companies such as [35,36].

The major OECB technology challenge remaining, currently surrounds coupling between the waveguides and optical SMT components. With the principal board fabrication issue being the 90° out of plane coupling, that would enable light incident normal to the board's surface to be coupled and transmitted along the waveguide. Subsequent research into assembly configurations, manufacturing tolerances, and hermetic requirements, will also be required before OECB systems can be commercially exploited.

14.6. COMPONENT ASSEMBLY OF OECB'S

The major OECB fabrication issue currently under investigation is the assembly, and 90° out-of-plane interconnection between the SMT components and the embedded waveguides. Research is also proceeding into the possibility of opening up pockets in the optical layers, sinking edge emitting components to emit directly into the waveguide facet therefore negating the requirement for the 90° change in direction. However, methods that couple signals to the waveguides without altering the direction of the light are limited in their application to other components, including low-cost VCSEL's, and therefore do not provide the overriding advantage of deploying 3rd generation interconnects, and the low-cost assembly of components. It is therefore essential to develop a low-cost method to change the direction of the light from being parallel with the board to projecting 90° out of plane to the top surface, thereby enabling compatibility with SMT devices and fiber ferrules. Literature suggests that there are two main distinguishable research trends for coupling out of plane to OECB's: assembled techniques, and integrated mirrors.

Assembled techniques form the 90° optical interconnection by first creating a trench in the OECB to expose the vertical facet of the waveguides. An optical sub-assembly based on a Ball Grid Array (BGA) interposer is then placed onto the OECB with a carrier inserted into the trench. Light is then coupled to and from the waveguides either by direct alignment of a VCSEL or by using a mirror to direct the light upward to the surface using either a free space or guided wave approach. Whilst assembled techniques provide a simple process route with relatively few technological issues to resolve, the optical coupling relies on manufacturing tolerances that are currently unable to provide sufficient repeatability. This

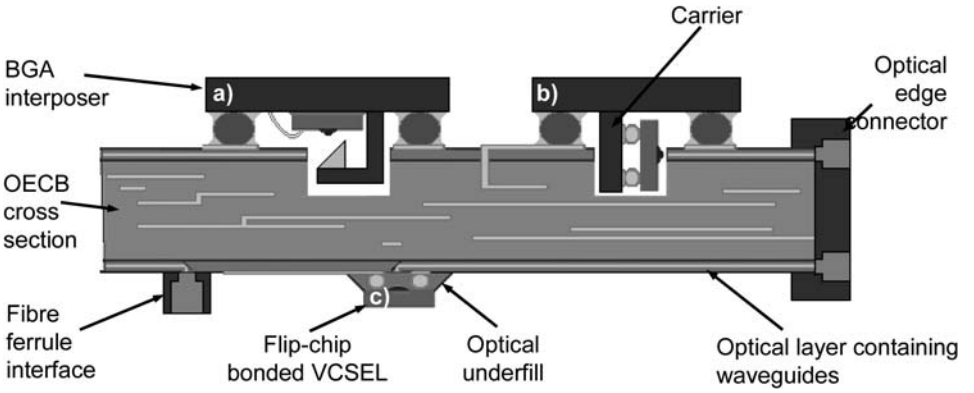


FIGURE 14.8. OECB component coupling techniques: Typical configurations for coupling components to OECB's showing, on the top, two examples of assembled mirrors with (a) the VCSEL mounted on an interposer allowing easier progression for other components such as fiber ferrules and (b) the VCSEL aligned to the waveguide supported by a carrier providing electrical contact. The bottom side of the OECB diagram shows (c) an example of an integrated mirror with routes for connecting a wide variety of components.

is due to the reliance on tensions in the solder joints to self align the components to the waveguides, assuming that the bond pads are correctly positioned in the outset. The depth to which the carrier is inserted is also a critical alignment factor that cannot be guaranteed since the reference plane is FR4.

Integrated mirror techniques combine the 90° deflecting mirrors on the ends of the waveguides, allowing active devices to be flip-chip bonded directly on top of the OECB. This technology not only removes the need for additional assemblies and their associated part costs and manufacturing processes, but also enables the OECB to drive toward low-cost volume manufacture through SMT assembly. It is this move toward traditional EMS (Electronic Manufacturing Service) competencies, coupled with the perceived benefits in the re-work and test areas, which makes integrated mirrors the solution that is the most likely to break the significant USR market barriers of cost and performance (Figure 14.8).

To date, several organizations have reported the development of integrated mirror solutions that allow the direct coupling of components into the waveguides, with examples being published work from Intel and Optical Cross-Connect (OXC) [36]. These two similar demonstrators are manufactured based on a standard core board containing electrical contacts for the components to which a section containing the waveguides is laminated. The waveguide section is pre-fabricated with 45° metalized Input/Output (IO) mirrors on the ends. The section is then aligned on the core board using fiducial marks to ensure the precise position of the mirrors to the VCSEL pads and bonded using an Ultra Violet (UV) curable epoxy. To enable SMT assembled devices to stand clear of the waveguide section it is necessary to sink the layer into a trench formed in the original board so as to sit the top surface of the waveguide insert within tolerance of the top surface of the core substrate. Using this approach Intel have reported that when coupling VCSEL's and PD's to the waveguides a misalignment in excess of ± 10 μm still yielded 80% efficiency from the maximum coupling. This tolerance range allows passive alignment, crucial in achieving the low cost manufacturing required for the systems.

With no set standards defining USR optical interconnect performance the important characteristic of the link is to maintain the signal integrity over the distances involved,

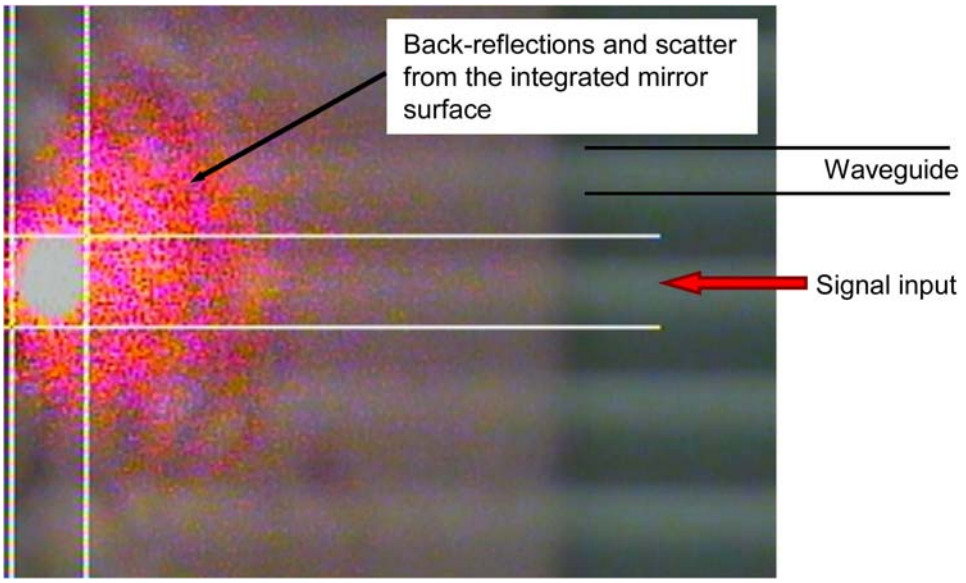


FIGURE 14.9. Plan view of an OECB demonstrating capability of a novel integrated mirror manufacturing process: For display purposes a visible laser is used to illuminate a waveguide with an integrated mirror formed on one end. Backscatter and divergence need to be reduced to improve attenuation.

thereby ensuring all the bits sent are correctly received. Although bit errors are generated by many factors, large link attenuation is a key contributor. Intel's demonstrator reported 1 Gbts error free transmission over link distances of approximately 20 cm with a loss budget of between 7 and 12 dB. Of this loss budget the IO coupling between waveguide and component constituted signal attenuation of between 1.5 and 3 dB. Despite demonstrating significant performance capability, the separate alignment and bonding steps required to attach the waveguide layer to the board contributes to additional tooling and process costs since the assembly methods are not aligned to that of current PCB fabrication. It is therefore necessary to further integrate the manufacture of the waveguides and IO mirrors into the substrate (Figure 14.9).

To that end, research by the authors is investigating a novel manufacturing approach utilizing current PCB fabrication techniques to manufacture OECB's with integrated mirrors. The substrates require no post processing allowing VCSEL's and PD's to be directly bonded to the OECB substrate. Although the integrated mirror OECB samples have not yet reliably yielded insertion losses of less than 3 dB, this is caused by the prototype nature of the manufacturing method and research is ongoing. However, once reliably achieved, significant research and development of the OECB is required to ensure sufficient reliability and performance measures are met. The remaining uncertainties include assembly issues, manufacturing tolerances, and hermetic requirements.

Although it would be advantageous to keep the assembly processes similar to those currently used for high volume SMT production, some differences are necessary. For example, it will probably be necessary to assemble the components using fluxless processes due to the risk of the flux coating the lenses and optical elements; this is common practice in optical component assembly. It is also envisaged that to remove the risk of particles permeating under the SMT devices, an optical underfill will be required. The underfill's

primary function will be to add hermetic protection to the signal by removing any free-space transmission rather than to aid mechanical reliability of the package. Removing the air-waveguide interface by index matching the underfill to the waveguide may also decrease optical losses by removing Fresnel reflections.

14.7. COMPUTATIONAL MODELING OF OPTICAL INTERCONNECTS

The issues met during the manufacturing and operation of optical interconnects can be aided by the use of the finite element modeling. This valuable tool can be used for the prediction of thermal, thermo-mechanical, and optical behavior during the manufacturing processes. Thermal simulations can show how much heat diffuses through a VCSEL package, and the rate at which it does so. Thermo-mechanical models predict the build up of stress and strain in the package due to CTE (Coefficient of Thermal Expansion) mismatches in the various materials present, which gives an insight into how close the CTE values of local materials must be to avoid undesirable stress values. Finally, optical simulations can help calculate the potential optical losses in the package by modeling a range of scenarios to see their effects on the transmitted signal.

Modeling work has been undertaken to aid the analysis of the OECB assemblies described above. A thermo-mechanical model was constructed to investigate misalignment and stress caused by machine component placement tolerance and CTE mismatches in the OECB, VCSEL's, and bonding materials [38]. Secondly, an optical model was used to predict the coupling efficiency of the optical interconnect as a signal is transmitted through it. The optical signal will be subject to attenuation losses when the VCSEL to waveguide coupling deteriorates. Computational modeling can help in identifying optimum process parameters that should be used to assemble the OECB's, giving an appropriate compromise between mechanical reliability and optical efficiency [37].

The thermo-mechanical model is constructed from the proposed OECB assembly, and dimensions used are accurate to manufacturing specifications. The model is comprised of a 1×4 VCSEL array, which is flip-chip mounted onto an OECB, bonded by Au stud bumps. The GaAs die of the VCSEL array, copper contact pads, and the Au stud bumps are all surrounded by an underfill material (Figure 14.10).

The thermo-mechanical model simulates the four VCSEL's in the array heating up due to normal operation. Localized heating in the area surrounding each VCSEL device is seen at around the expected temperature of 85°C. With the correct heating profile, the resultant stresses can be seen due to these thermal effects and the CTE mismatch between the various materials present (Figure 14.11). For all the simulations the greatest stress and deformation occurred in the outermost VCSEL's in the 1×4 array, and so attention was focused on these when measuring the in-plane misalignment (between VCSEL aperture and waveguide entrance).

It should be noted that all but one material used in the thermo-mechanical model is isotropic. FR4 material is orthotropic, having material properties in the plane different to that out of the plane. In the plane the CTE value is 21 ppm compared to that of 60 ppm out of the plane (Figure 14.12).

Two underfill types are used in the simulations to compare the impact each one has on the model. From Figure 14.11, "Underfill 1" is standard underfill and "Underfill 2" is an optical underfill which does not contain filler particles as would be found with a traditional underfill. Such an optical underfill is similar to the so-called no-flow underfill. The CTE

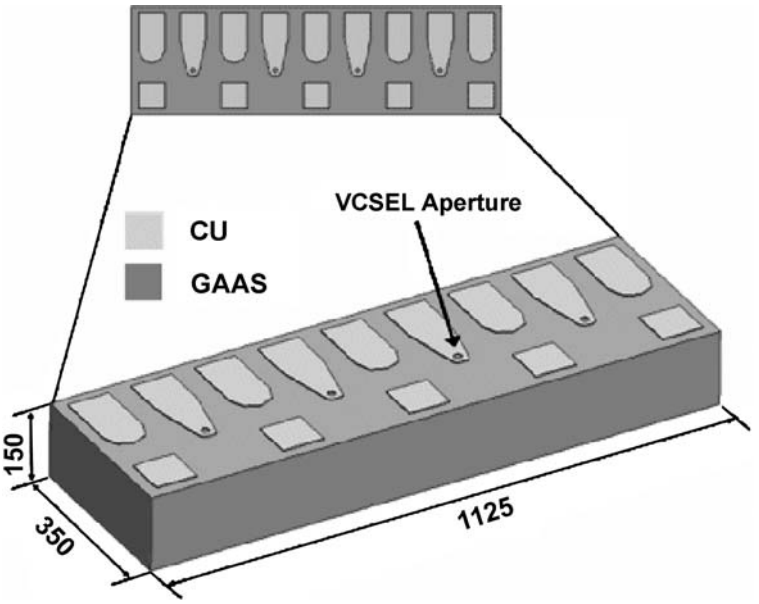


FIGURE 14.10. VCSEL array geometry: 1×4 VCSEL array geometry with copper pad detail.

	Thermal Cond. (W/mK)	Specific Heat (J/kgK)	Density (kg/M^3)	CTE ($(K^{-1})*10^{-6}$	Young's Mod ($N\,m^{-2}$) GPa	P. Ratio
FR4	0.3	100	1900	21	21	0.3
Copper	400	384	8960	16.5	130	0.34
Gold	8	129	19280	14.2	78	0.44
GaAs	45	350	5320	5.8	85	0.31
Underfill 1	250	900	8000	30	6	0.35
Underfill 2	250	900	8000	75	2	0.35
Epoxy	29	150	2100	60	0.179	0.32
Waveguide	29	150	2100	60	0.179	0.32

FIGURE 14.11. Material properties: Thermo-mechanical material properties used for the VCSEL array computational model.

value for the optical underfill is much higher and is less stiff (due to a lower Young's modulus). As expected, much more movement and deformation was seen with the no-flow underfill material during simulations, due to the variation in material properties between the two underfill types. For this reason the optical underfill material was used for all further simulations (where underfill is required), as this shows the worst case scenario for using an underfill from the two choices available.

The thermo-mechanical model has boundary conditions in place to allow for natural deformation and stress build up to occur. For a single VCSEL array, by fixing two whole side faces and the entire bottom face it is assumed that the array is one out of a block of four. The fixed side faces define the simulation symmetry planes. This has the inherent advantage that only one VCSEL array is actually needed to modeled, significantly reducing pre-processing model development and simulation run time. The computational mesh is comprised of 46,659 individual elements and is unstructured (Figure 14.13).

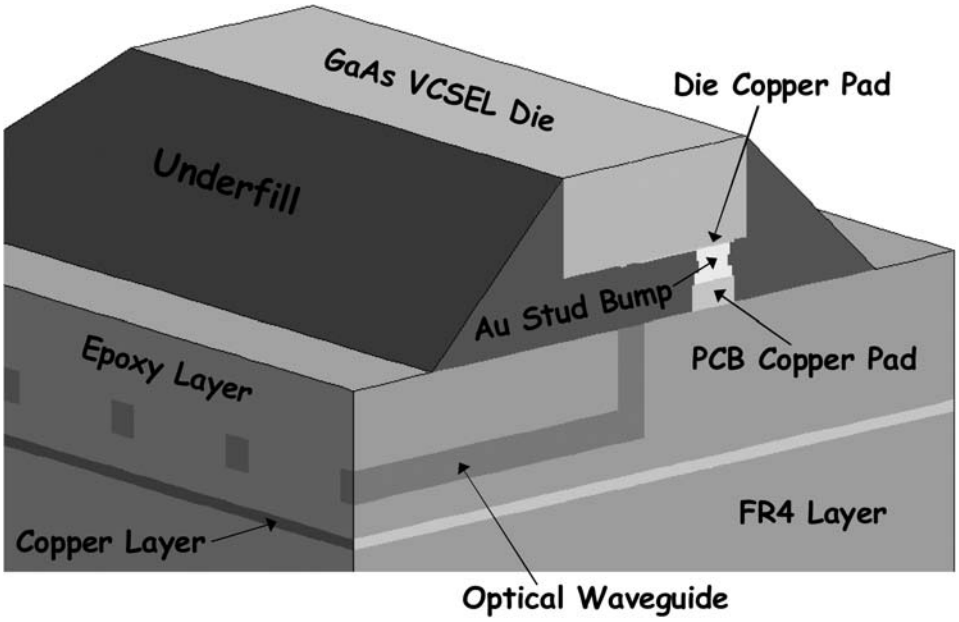


FIGURE 14.12. Material layout: the various materials and their locations on the model geometry.

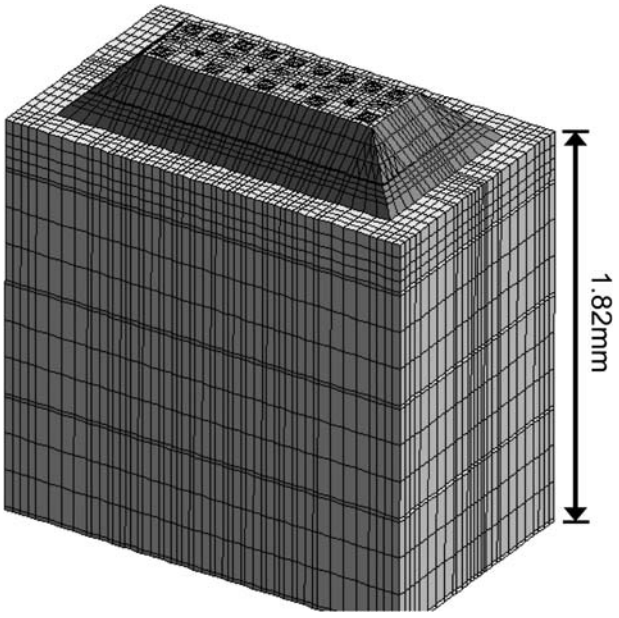
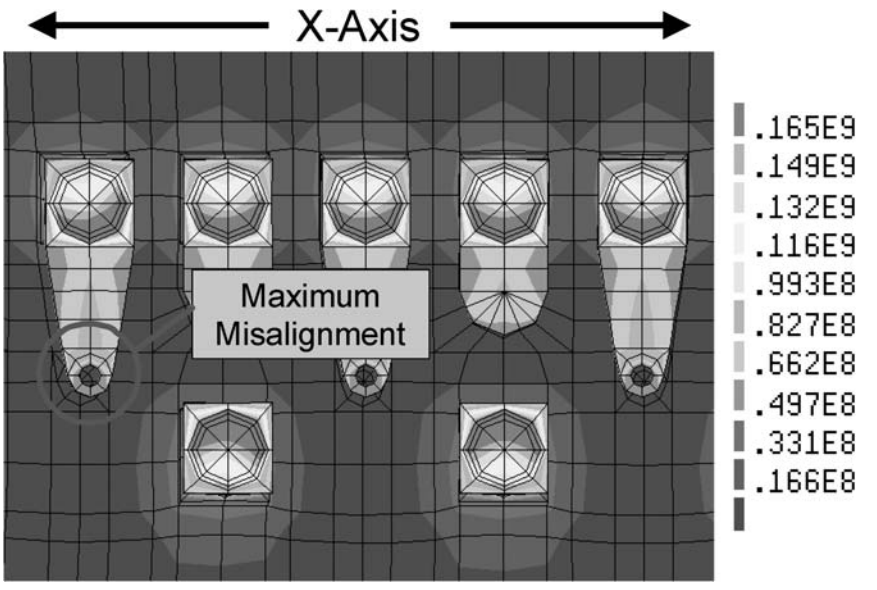
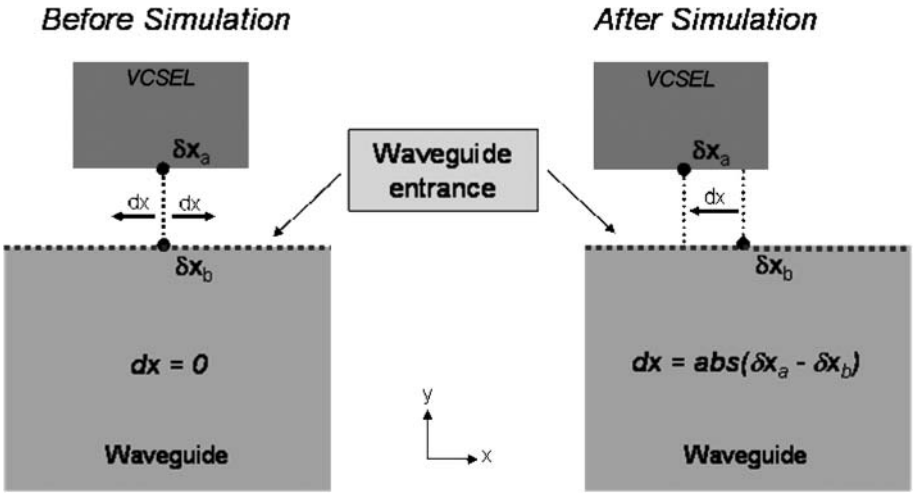


FIGURE 14.13. Computational model geometry: the VCSEL array embedded in underfill and mounted on an OECB. The mesh is shown.

During initial testing the thermo-mechanical model showed the greatest general deformation in a single direction, along the x-axis (Figure 14.14). Deformation in the other two axes directions movement was negligible. The misalignment, dx, is the measure of the



(a)



(b)

FIGURE 14.14. Misalignment results: (a) Maximum misalignment, in the x-direction, shown to be on the outer VCSEL of the array. GaAs die and underfill materials removed for clarity. (b) Method of calculating misalignment during simulations in the x-direction.

absolute difference between the center of the VCSEL aperture and center of the corresponding waveguide entrance. After simulation dx gives the amount of deformation occurring during the simulation.

Thermo-mechanical simulations induced stress in the VCSEL array by increasing the localized temperature surrounding each VCSEL and allowing the differences in the CTE values of the various materials to act. "Delta t" (Δt) is the temperature increase above the ambient starting temperature, which is 25°C at the beginning of each simulation. The

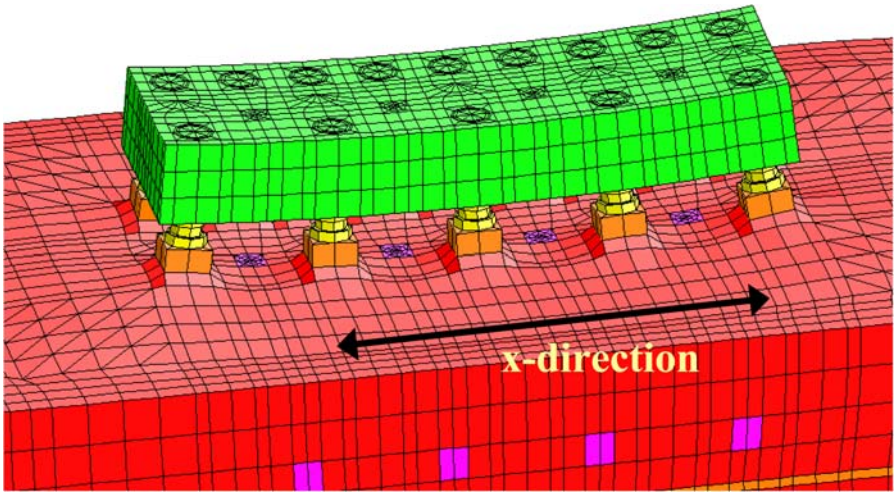


FIGURE 14.15. Mesh deformation: maximum misalignment is in the x-direction, and for the outer most VCSEL's in the array. GaAs die and underfill materials removed for clarity.

value of Δt, which is 60°C for the thermo-mechanical model, together with the CTE value for each material will give the maximum amount of displacement possible. For example, taking the CTE value for epoxy material to be 60 ppm (parts per million), the maximum linear displacement possible for $\Delta t = 60°\text{C}$ (in any axial direction) is given as follows:

$$\Delta L = \alpha \cdot L_0 \cdot \Delta t,$$

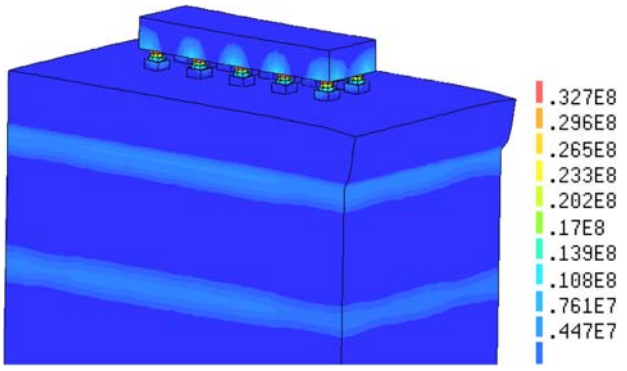
$$\Delta L = (60e - 6) \cdot (1e - 3) \cdot (60),$$

$$\Delta L = 3.6\ \mu\text{m}.$$

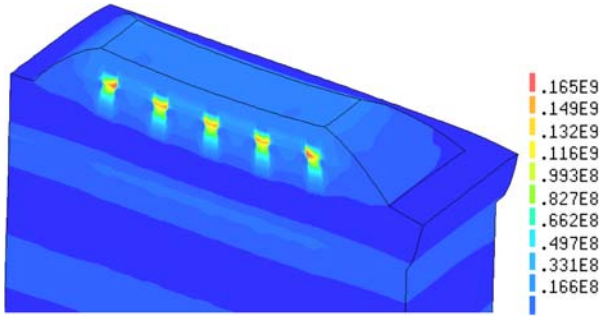
Taking the original length (L_0) to be 1 mm, which is the approximate length of the VCSEL array, we have the maximum possible linear expansion for epoxy material to be just 3.6 μm. With other materials having much smaller CTE values, the actual misalignment expected would be significantly less than for the calculated epoxy material alone. In fact, simulations results show sub micron misalignment for dx (Figure 14.15). This is consistent with expected results.

To fully understand the impact of optical underfill on the VCSEL array coupling to waveguide, simulations were run for two different underfill scenarios. One simulation had the encapsulant all the way around the sides of the GaAs die, and throughout the gaps between the stud bumps (see Figure 14.12). The other simulation scenario had no underfill present on the model, leaving the die and stud bumps effectively exposed [Figure 14.16(a)]. It was found that the higher CTE value for the optical underfill deforms the VCSEL array package significantly more than when the underfill is removed. This results in higher stress levels around the stud bump joints [Figures 14.16(a) (b)].

Also, further simulations were performed with optical and standard underfill with both traditional tin/lead solder (SnPb) joint material and Au stud bumps. For simulations where the Au stud bumps were used, an epoxy layer was present; and for the solder (SnPb) material, FR4 is used in place of the epoxy layer. From the results, simulations showed that standard underfill (when used with SnPb solder) acts as a support for the package and helps to reduce deformation and stress levels when compared to no underfill. This is an expected



(a)



(b)

FIGURE 14.16. Stress contour plots for the two underfill scenarios: (a) no underfill model; (b) complete underfill model (optical underfill), cutaway view to see the stress contours in the stud bumps.

Underfill Type	Au Bump Stress (MPa)	SnPb Stress (MPa)
Traditional	57.18	27.44
Optical	170.26	96.10
None	27.89	35.11

FIGURE 14.17. Stress results: comparison of the stress values for the Au stud bumps and solder stud bumps.

trend for traditional die, underfill, solder and FR4 materials. But for the simulations when using the optical underfill (with Au stud bumps) it was found that the presence of the optical underfill and epoxy layer acts as a stress enhancer, when compared to no underfill (Figure 14.17).

The optical model simulates the light coupling from each VCSEL aperture into the corresponding waveguide entrance. The waveguide is constructed of two straight sections

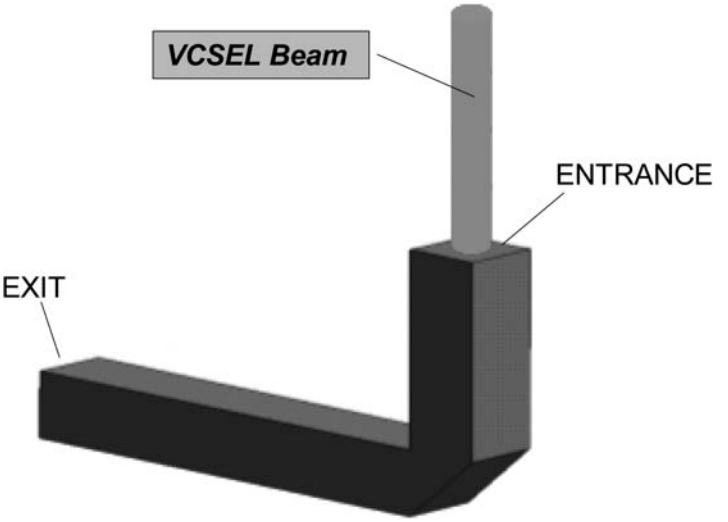


FIGURE 14.18. Optical waveguide: schematic diagram of the optical model setup.

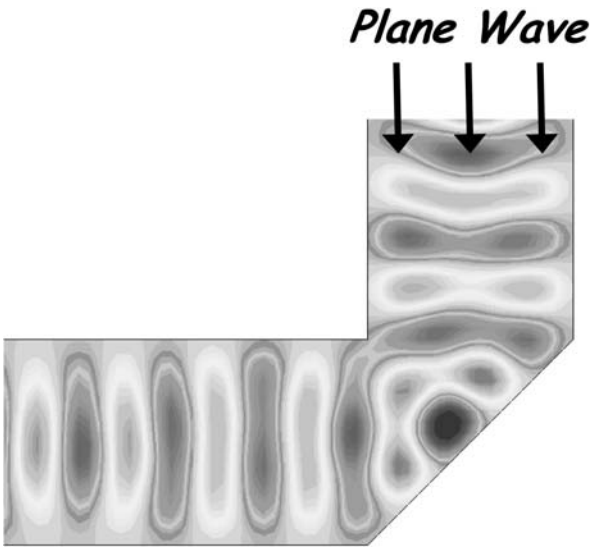


FIGURE 14.19. Optical model simulation: optical simulation results showing electromagnetic plane wave propagation through the waveguide.

connected at right angles via an integrated mirror (Figure 14.18). The governing equation solved in this analysis is the Helmholtz equation, which is a subset of Maxwell's equations, and assumes the light source is monochromatic and is propagated as a plane wave [39].

The aim of the optical simulations is to predict the coupling efficiency of the VCSEL beam. This is characterized by the attenuation value which is calculated by comparing the level of the optical signal at the VCSEL aperture to that at the waveguide exit. Any attenuation observed will be as a result of the misalignment between VCSEL and waveguide, geometry of the waveguide, and the polymer material properties used (Figure 14.19).

14.8. CONCLUSIONS

For decades optical interconnects have been replacing copper throughout the telecommunications network in order to provision for the continually increasing demand for greater bandwidth capacity. As this trend becomes to be pertinent even at the shortest link distances, the successful deployment of optical links becomes dependent on the technological advances necessary to create optical links with strong manufacturability and greater performance over established electrical interconnects whilst, most importantly, at a lower cost. Just as the development of low-cost VCSEL's allowed optical links to replace copper over VSR interconnects between co-located CO equipment, the development of hybrid OECB's, integrating both electrical interconnects and 3rd generation embedded waveguides into one substrate, are expected to unlock the path toward high speed optical backplanes and optically interconnected ICs.

Prior efforts to create OECB's have centered on laminating optical layers to FR4 and although they have yielded acceptable out-of-plane coupling losses of around 3 dB, the additional lamination process is not capable of achieving the required cost target. However, the authors have investigated a novel manufacturing method for creating low-cost OECB's by aligning the fabrication processes much closer to that of those already used inside PCB plants. With the research results to date suggesting that the resulting integrated mirrors are capable of coupling fibers and SMT components with attenuation losses of less than 3 dB, the future for OECB's looks promising. However, consistently achieving and improving this low-cost out-of-plane interconnection attenuation is only part way to seeing commercially viable OECB systems. Significant development is still required, specifically centered on the passive assembly of optical components to the OECB's and their long term reliability.

To aid the development of the OECB assemblies, computational modeling is being used. The computational models for the thermo-mechanical and optical simulations are both in place and producing positive results with close synergy to the experimental work. Initial results from the modeling work have shown that the temperature of the VCSEL will reach 85°C producing some misalignment of the VCSEL due to CTE mismatches. The optical model is now nearing completion and the aim here will be to predict attenuation for a particular set of design parameters. Once completed, a virtual DOE will be run to help understand the impact of each design parameter in an effort to reach the optimal conditions that minimize the signal attenuation.

Assuming the development of low-cost OECB substrates continues to progress and that there are no major component assembly and reliability issues to overcome. OECB enabled, high-speed passive optical backplanes should be a reality on CO telecommunications equipment within the near future. Future development in the assembly of MOEMSs to the OECB's will enable the backplanes to become reconfigurable, paving the way toward the low-cost all optical switch. OECB assemblies will also enable on-board optical clocks to time future generations of IC chips. Bringing ever closer the holy grail of photonic computing.

ACKNOWLEDGMENTS

The authors would like to acknowledge the financial support provided by both the EPSRC and Celestica in the form of two Industrial Case Awards. In addition they would

also like to acknowledge the assistance provided during the development of the OECB integrated mirrors by both Terahertz, in supplying the waveguides, and Ablestik, for the supply of the optical adhesives.

REFERENCES

1. MOLEX Website, www.ttieurope.com/microsites/molex/products_items/backplane/flexplane/overview.cfm (accessed 12th Dec. 2004).
2. BPA, Optical backplanes—A global market and technology review 2000–2005, BPA Consulting: Dorking, Surrey, UK, Report No. 762, 2001.
3. P. Van-Daele, Optics in PCB's: a Dream Coming True in Short Distance Optical Interconnects—From Backplane to Intra-chip Communication, CHACH, Chalmers University, Sweden, March 7th 2003.
4. D.K. Mynbaev and L.L. Scheiner, Fibre-Optic Communications Technology, Prentice-Hall, Inc., Upper Saddle River, New Jersey, USA, 2001.
5. G.P. Ryan, Ed., Dense Wavelength Division Multiplexing, ATG's Communications & Networking Technology Guide Series, The Applied Technologies Group: One Apple Hill, Suite 216, Natick, MA, USA, 1997.
6. D.A.B. Miller, Rationale and challenges for optical interconnects to electronic chips, Proc. IEEE, 88(6), pp. 728–749 (2000).
7. D.A.B. Miller, Optical interconnects to silicon, IEEE Journal: Selected Topics in Quantum Electronics, 6(6), pp. 1312–1317 (2000).
8. Cisco Systems Inc., Cisco Very Short Reach OC-192/STM-64 Interface: Optimizing for Network Intra-POP Interconnects, White Paper, 170 West Tasman Drive, San Jose, CA, USA, Cisco Systems, Inc., 2001.
9. M. Fuller, VSR optics acheives working concept and rough consensus, Lightwave, (February) (2001).
10. M.K. Dhodhi, S. Tariq, and K.A. Saleh, Bottlenecks in next generation DWDM-based optical networks, Computer Communications, 24, pp. 1726–1733 (2001).
11. Photonami, The Surface Emitting DFB M3 Laser, Photonami, 50 Mural Street, Richmond Hill, Onatario, Canada, June, 2003.
12. P.M. Henderson, Introduction to Optical Networks, January 29th, 2001, Mindspeed Technologies, Inc., Newport Beach, CA, USA, p. 36.
13. D. Welch, VCSELs: driving the cost out of high speed fibre optic data links, Compound Semiconductor Magazine, (July) (2000).
14. F. Mederer, et al., High performance selectively oxidized VCSELs and arrays for parrallel high-speed optical interconnects, IEEE Transactions on Advanced Packaging, 24(4), p. 442–449 (2001).
15. M. Peach, ULM refines VCSEL manufacturing technique, Lightwave Europe, June 2003.
16. Y.-C. Ju, Vertical-Cavity Surface Emitting Lasers (VCSEL), ETRI, KyungPook National University: Department of Physics Education, Teachers College, 2004.
17. Ulm Photonics, Product Datasheet: 850 nm multimode 2.5/3.125/5 Gbs N $\times$ M VCSEL array for direct FlipChip. Accessed (14th Dec. 2004), www.ulm-photomics.com.
18. E. Mohammed, et al., Optical interconnect system integration for ultra-short-reach applications, Intel Technology Journal: Optical Technologies and Applications, 8(2), pp. 115–128 (2004).
19. J. Theodoras, L. Paraschis, and A.C. Houle, Emerging Optical Technologies for Multi-Service Metro Networks, Advanced Technology Group, Optical Networking, Cisco Systems, 170 West Tasman Drive, San Jose, CA, USA.
20. Optical Internetworking Forum, www.oiforum.com. Accessed (8th Dec. 2004).
21. M.L. Hammond, Moore's law: the first 70 years, Semiconductor International, (April) (2004).
22. I. Young, Intel introduces chip-to-chip optical I/O prototype, Technology @ Intel Magazine, (April), p. 7 (2004).
23. N. Savage, Linking with light, IEEE Spectrum, (Aug.), pp. 32–36 (2002).
24. B. Layet and J.F. Snowdon, Comparison of two approaches for implementing free-space optical interconnection networks, Optics Communications, 189, pp. 39–46 (2001).
25. M. Chateaunenf, et al., 512-channel vertical-cavity surface-emitting laser based free-space optical link, Applied Optics, 41, pp. 5552–5561 (2002).
26. S. Mukherjee, The quest for an affordable replacement for inter-processor switched free-space interconnect, Short Distance Optical Interconnects—From Backplanes to Intrachip Communication, CHACH, Chalmers University, Sweden, March 7th 2003.

27. G.C. Boisset, B. Robertson, and H.S. Hinton, Design and construction of an active alignment demonstrator for a free-space optical interconnect, IEEE Photonics Technology Letters, 7(6), pp. 676–678 (1995).
28. K. Hirabayashi, et al., Optical beam direction compensating system for board-to-board free space optical interconnection in high-capacity ATM switch, Lightwave Technology, 14(5), pp. 874–882 (1997).
29. D.P. Resler, et al., High-efficiency liquid-crystal optical phase-array beam steering, Optics Letters, 21(9), pp. 689–691 (1996).
30. T.-Y. Yang, J. Gourlay, and A.C. Walker, Adaptive alignment packaging for 2-D arrays of free-space optical interconnected optoelectronic systems, IEEE Transactions on Advanced Packaging, 25(1), pp. 54–64 (2002).
31. T.D. Wilkinson and W.A. Crossland. Reconfigurable optical interconnects, MicroTech 2004, Moller Centre, Cambridge, IMAPS, UK, March 3rd–4th 2004.
32. Dickens, Optical backplanes—A global market and technology review 2000–2005, BPA Consulting, Dorking, Surrey, UK, 2001.
33. NTT-AT Website, www.ntt-at.com/products_e/opticalfiber/ (accessed 12th December 2004).
34. S. Agelis, et al., Modular interconnection system for optical PCB and backplane communication, Proc. International Parrallel and Distributed Processing Symposium, Fort Lauderdale, FL, USA, April 19th 2002.
35. Terahertz Photonics, Truemode Backplane: Product Information, Rosebank Park, Livingston, Scotland, UK, 2003.
36. OXL Website, www.opticalcrosslinks.com (accessed 14th December 2004).
37. J.J. Morikuni, P.V. Mena, A.V. Harton, and K.W. Wyatt, The mixed-technology modeling and simulation of opto-electronic microsystems, Journal of Modeling and Simulation of Microsystems, 1(1), pp. 9–18 (1999).
38. PHYSICA, Multi-physics software Limited, University of Greenwich, London. http://www.gre.ac.uk/~physica.
39. J. Piprek, Semiconductor Optoelectronic Devices—Introduction to Physics and Simulation, Academic Press, California, 2003.

15

Adhesives for Micro- and Opto-Electronics Application: Chemistry, Reliability and Mechanics

D.W. Dahringer

153 Hawthorne Avenue, Glen Ridge, New Jersey 07028, USA

15.1. INTRODUCTION

The performance and reliability of micro-and opto-electronic devices has been crucial to the recent advances in technology. The successful incorporation of adhesives into the design and manufacturing operations has been a major contributor to those advances. In many innovative optical and microelectronic applications, adhesives may be the only practical, cost-effective choice for producing such devices and components. Despite the ubiquiteness of adhesives in our daily lives, only sometimes obvious, there is a persistent concern and distrust for their use. Much of this can be attributed to causal personal experiences with so-called "miracle glues" and the associated outlandish performance claims. In reality, the overall aversion to adhesives can be attributed more to the misuse of them than to the actual material itself. This chapter will attempt to underscore the importance of proper joint design and assembly processing as well as the adhesive characteristics necessary for the ultimate success of a bonded joint.

15.1.1. Use of Adhesives in Micro and Opto-Electronic Assemblies

There are a number of reasons why adhesives are selected for assembly operations, even when there may be alternative assembly techniques available. Some of these reasons include performance, reliability, cost, ease of manufacturing and form factor.

15.1.1.1. Performance The high performance of adhesives in properly designed and manufactured devices has allowed designers to take maximum advantage of the precise part alignment and thermal and environmental stability achievable. Submicron assembly tolerances in many optical devices are commonplace, often achieving very high coupling efficiencies that are stable over broad temperature ranges. Adhesives are used to bond glass, ceramics and metals in configurations and dimensions frequently incompatible with other fastening techniques. For micro-electronic components, adhesives find use in enhancing the performance and reliability of its discreet chip packaging, BGA's, ferrite magnetics, etc.

15.1.1.2. Reliability Once established, a proper design and manufacturing process using adhesives as the attachment medium, can be both reproducible and reliable. Awareness of potential failure mechanisms and the conditions to which an assembled joint will be exposed, can aid a physical designer in developing the most reliable design and process. Examples can readily be seen in poor joint designs that fail just on cooling to room temperature after curing the adhesive; and other bonded structures that have survived without degradation in performance or integrity over many years of extreme environmental exposure.

15.1.1.3. Cost Adhesives are normally a small contribution to the overall cost of an assembly, despite what may seem like high prices for relatively small quantities of material. Part of this cost comes from the specialty nature of both optical and electro-mechanical bonding materials and partly from the small volume usage over which development costs must be recovered by the manufacturer.

As device dimensions get smaller and bondlines thinner, the quantity of adhesives per device and its cost, almost become irrelevant. Certainly, precious metal-filled (conductive) adhesives used on ground planes could be an exception and certainly the costs associated with the bonding and curing process are part of the total adhesive related costs, nevertheless, they are usually significantly below comparable costs for alternative assembly methods.

15.1.1.4. Ease of Manufacturing The ease of manufacturing a product with adhesives can only be appreciated when the assembly process is compared to a viable alternative. From an absolute perspective, the essence of a bonded joint is to prepare the substrates, fixture, apply the adhesive and cure the adhesive. Each of these steps can be simple or complicated depending on process and performance needs, but are usually less difficult than competing processes.

15.1.1.5. Form Factor Because adhesives will allow very thin bond lines and distribute stresses over relatively large areas, designers usually have wider options in deciding the final shape and size of a device compared to other assembly techniques. This is particularly true in the opto-electronic field where small size has become the industry standard.

15.1.2. Specific Applications

15.1.2.1. Micro-Electronics Most applications for adhesives in micro-electronics fall into the areas of die attach, BGA underfill, component attach (as in surface mount), and component assembly. For die attach, much but not all of the work is with conductive adhesives especially in the manufacture of discreet devices. Underfill applications are designed to improve the reliability of the solder interconnections which would otherwise be stressed by the thermal coefficient of expansion (TCE) difference between the BGA component and the printed circuit board substrate. Component assembly would include devices such as ferrite transformers and inductors, displays, switches and contact pads.

15.1.2.2. Opto-Electronics The opto-electronic field has many specific applications for adhesives, including the ubiquitous fiber to ferrule, V-groove arrays, splitters, gratings, optical bench component assemblies, integrated assemblies and protective packages.

15.2. ADHESIVE CHARACTERISTICS

15.2.1. General Properties of Adhesives

15.2.1.1. Data Sheets Most adhesive suppliers provide data sheets describing their products in an attempt to simplify the customer's selection process by indicating some measure of both application and performance related properties. Unfortunately, from a practical point of view, much of the data is of limited value. Material properties, such as filler content, color, viscosity of one or more parts, storage conditions and cure conditions can be important for the selection process, but in-depth comparisons between suppliers may be much more difficult to realize. Part of this is due to lack of standardization of test methods between suppliers; and part is due to insufficient description of the conditions of the material and the test process; and part to the lack of appreciation by the consumer of the effect different test conditions can have on the reported data. For example, a supplier may indicate that an adhesive is suitable for use to, let's say 250°C based on a TGA weight loss of 5 or 10% at that temperature. In reality, the test sample (whose shape is typically that of a pellet) is cured in bulk, and could behave totally different than a thin adhesive bondline between substrates. For example, volatiles, either in the uncured adhesive or the substrate, or even generated as part of the curing process, might be trapped in the bondline but not necessarily in the bulk specimens. Furthermore, a loss of 5–10% weight (as a gaseous degradation product) could have severe (mechanical) implications in a bondline, especially if the adhesive is likely above its glass transition temperature.

Another example of data sheet confusion is the glass transition, supposedly an indicator of the point where a material switches between a glassy solid and a rubber. In addition to the fact that it is usually a range instead of a "point," one can measure different T_g's on the same sample, depending on the test method, DSC, TMA or DMA. Even with a single test method, variations in the test parameters (e.g., sample size, scan rate) can result in different values. Interpretation of the collected data can also add variability, e.g., some DSC methods pick the inflection point midway between the change in measured heat capacity, while others may select the first deviation. With TMA analysis, either the initial point of change in TCE, or the intersection of the two slopes is considered the T_g.

Additionally, different sample preparation schemes can cause major variations in T_g even if all of the above considerations have been standardized. For example, if a test sample is cured dynamically in a DSC at some rate to some maximum temperature and then rescanned for T_g, the result could be significantly different depending on the cure scan rate, maximum temperature and any thermal dwells. Isothermal cures can add additional possibilities when both time and temperature factor into the results.

To further complicate this matter, the actual thermal profile of a specific adhesive in an actual joint is often quite different than a design process engineer's expectation. In oven cures, the bondline (adhesive) temperature can be very different than the oven temperature depending on thermal mass of the substrates and recovery time of the oven. This can be a major factor in determining the zero stress condition/temperature for a joint, especially for rapid curing adhesives.

Other typical data sheet properties, involving less controversy, are color, density, specific gravity, mix ratio (for multi-component system), storage requirements and cured electrical properties. Viscosity is a data sheet parameter having a major impact on process adaptability for some applications, especially those requiring capillary flow, such as chip underfill; or high thixotropy such as surface mount adhesives. For non-thixotropic (shear insensitive) adhesive systems, the viscosity reported is usually meaningful (at least at room

or reported temperature), however, for the systems with filler and the possibility of shear sensitivity, the actual reported viscosity may be significantly different from the real viscosity at the use condition. Thixotropic index (a ratio of viscosities at two different shear rates) can sometimes provide useful information (although mostly intuitive for the practitioner). Typically missing from the data sheet viscosity equation is the affect of temperature (either recommended application or cure) on this property.

Data sheet representation of moisture absorption is often misleading in that influential factors in specimen preparation, such as cure condition, size, shape, porosity and exposure conditions (such as time and temperature) are often missing. Generally, an individual supplier will be consistent with data sheet report of moisture pick-up values for different materials in their line, but comparing one supplier to another can be difficult at best. A further possible distortion of moisture data can come from materials that may have a water-leachable component. This can best be determined by a redried weight loss method, which is rarely reported and quite possibly be the explanation for some claims of low moisture absorption adhesives.

15.2.1.2. Physical Properties The physical properties of adhesives can and should vary with the specific process needs and the end use application, thus largely preempting the possibility of a "universal adhesive." Typical bulk material mechanical properties, like tensile strength, elongation and modulus, unfortunately provide little in the way of practical help in choosing an adhesive, although a low modulus product is sometimes selected in an attempt to reduce joint stresses where the substrates differ in TCE.

Tensile strength of an adhesive is unusually difficult to measure, mainly because of sample preparation problems. Sometimes a butt tensile test can be used, but rarely would one expect true, cohesive failure within a bondline. On the other hand, a non-axial loaded butt tensile failure is probably a better simulation of a cleavage test and can relate to some of the common component attachment applications in optical assemblies. Many companies have developed QC tests based on cleavage stress using a simple glass rod to glass slide test joint. Fixed length rods pushed at the top, parallel to the bondline, can be a convenient test specimen for not only measuring initial properties of an adhesive but also the effect of environmental conditions such as high temperature and/or high humidity aging or thermal cycling.

The property of elongation is somewhat useful in guessing the performance of a large fillet, calculating the edge stresses of a thin bondline or figuring the likelihood of a mismatched TCE joint failing. Elongation does come into play in joints with significantly thick bondlines and situations where a high elongation (and low modulus) adhesive is used for a potting type application. Factors often overlooked in these applications are the necking affect perpendicular to stress induced elongation and the usually high effective volume TCE of low modulus materials. For most well designed, thin bondline joints, elongation is of little practical importance, primarily due to the physical constraints imposed on the z-axis by the substrates' influence on the x and y-axis and, of course, the small "h" of the adhesive.

Specific gravity or sometimes density is useful in calculating cost per unit volume and volume ratios of multi-component adhesives when only weight ratios are provided. In general, unfilled adhesives average close to 1.0 g/cm^3 but can reach much higher (or even lower) densities with the addition of fillers and/or air. Air can be unintentionally added during hand or machine mixing without appropriate degassing. A simple method of confirming the existence of entrapped air (even with heavily filled adhesives) is to squeeze a drop between two glass microscope slides and inspect under magnification for typical

spherical (usually appearing dark) voids. Shims or large particles (spanning the bondline) can cause shrinkage voids on cure which differ in appearance from entrapped air. These voids often have a "crowfoot" appearance, a halo effect around a discreet, large particle or sometimes cracks between particles or columns as can occur in underfill encapsulants.

Refractive index typically averages 1.56 for epoxies and 1.42 for acrylics and perhaps even lower for silicones. What may need clarification in refractive index applications is that numbers reported are often obtained on an uncured specimen (for easy cleanup of the instrument), at a wavelength and temperature different from the final use conditions. Each of these factors, although quite small, can have an impact on some in the light path applications.

The change in volume of an adhesive as it cures is called cure shrinkage and results from the fact the curing agent and resins in a typical adhesive have a lower density (take up more volume) than the final cured material. A typical data sheet value can be determined by using density ratio (before and after cure) however, this may not reflect accurately on an application requirement or on other shrinkage test methods which may report lineal instead of volume shrinkage.

In a real life application, where an epoxy is placed between two substrates and heated to some curing temperature, a very complex scenario is played out. First, the liquid adhesive expands as the temperature rises and excess material exudes from the joint (or the substrates move slightly apart if they are not constrained). Second, as the cure mechanism begins, volume shrinkage causes the still liquid exudate to backfill up to the point of gelation. Gelation occurs roughly in the 30–40% of full cure range. At that point, no more backfill can occur and further shrinkage must develop stress in the z-axis of a constrained thickness bondline. What is even more significant is that after gelation, the x and y-axis of the adhesive in the bondline are constrained by the substrates and therefore all the remaining volume shrinkage is forced into the z-axis. The final step of the process is cooling the assembly from cure temperature to the use temperature which further adds to the z-axis shrinkage and the stress on a constrained assembly. In the case of an unconstrained assembly of two planer substrates of the same TCE (visualize two glass microscope slides), the bondline thickness will vary to compensate for each of the steps and will provide an essentially stress-free joint. Unfortunately, many practical bonding applications do not consist of both planer and same TCE substrates.

15.2.1.3. Thermal Properties Thermal properties of adhesives can be examined from both a chemical and a mechanical viewpoint as well as the short term and long term effects. The chemical effects tend to be relatively straightforward whereas the thermomechanical effects can be quite complex. The appreciation and understanding of these thermomechanical effects can dramatically improve a practitioner's success in designing high performance and reliable adhesive joints.

In general, most thermosetting epoxy adhesives undergo few chemical changes below the temperature at which they were cured. Exceptions to this include adhesives that have been incompletely cured at the "cure" temperature, adhesives cured at very high temperatures (sometimes called snap cure), adhesives that have been mis-formulated (i.e., too much or too little curing agent) and adhesives that contain unreacted or unreactive components. Subsequent heating of an incompletely cured adhesive can cause additional cross-linking which will generally raise T_g, cause additional shrinkage and improve chemical and moisture resistance. In some adhesives, additional heating causes slight chemical rearrangement where strained bonds are broken and new ones form. At higher temperatures several different forms of degradation can take place. Acrylics can undergo a depolymerization lead-

ing to volatilization and weight loss. Epoxies can develop color followed by charring and weight loss or gain possibly due to bond breakage or oxidation. Small amounts of even the most aggressive degradation associated with very brief high temperature exposure can, in some cases, be tolerated from a chemical perspective, if the mechanical aspect of the joint does not fail or degrade significantly.

15.2.1.4. Thermomechanical Properties One of the most interesting aspects of adhesive engineering and probably the most significant in terms of joint design for reliability in non-ideal assemblies (as opposed to the "ideal" joint having an unconstrained planer bondline between substrates of identical TCE) involve the interaction of thermomechanical properties of an adhesive and its substrates during thermal excursions. The TCE of an adhesive is typically defined as the change in length of a material, per degrees C change in temperature, and usually expressed as parts per million per degree C, or ppm.

TCE's are typically measured in a dilatometer where a specimen height (length) is recorded as a function of or programmed temperature ramp and the slopes of the curve are used to calculate $\alpha 1$ and $\alpha 2$. At the T_g of thermoset epoxies, a change in expansion coefficient takes place, with a new slope $\alpha 2$ and value approximately three times that below the T_g. The intersection of the two slopes provides the typical TMA value for T_g and from a practical point of view, the more useful data point for determining joint thermomechanical properties compared to the value obtained in a DSC. Typical TCE's of unfilled epoxies normally run around 60–70 ppm/°C below T_g and 180–210 ppm/°C above T_g.

The incorporation of fillers into adhesive formulations can be used to modify the TCE's of the adhesives. In general, the amount of TCE reduction tracks the volume fraction and individual TCE of the components. When lowering the adhesive's TCE is a prime concern, silica is generally the filler of choice because of its very low TCE. There is a practical limit to how low in TCE one can go and still maintain a fully "wet out" filler and void-free space between the filler particles. The use of spherical silica particles with precise (multi-modal size distribution) has enabled some adhesive formulators to manufacture adhesives with TCE's of 12–15 ppm/°C. This range can be very helpful in some applications where joint design is less than ideal. There are, however, trade-offs in the use of such materials, mainly in the handling where these heavily filled adhesives are likely to be stiff pastes, making adhesive dispensing and placement difficult. For a balance of handling and low TCE, values in the range of 30 ppm/°C seem to be popular, especially in the BGA underfill market.

The use of low TCE materials to control unconstrained z-axis movement in optical devices may actually be counter-productive. This is because the paste-like nature results in a thick bondline versus the extremely thin bondline achievable with some unfilled materials. For example, one might be able to squeeze a 15 ppm adhesive to 100 microns, but an unfilled version of the same adhesive could be capable of forming a joint with a thickness of a micron or less. In this example, the low TCE adhesive could cause a joint movement (z-axis) approximately 20 times that of the unfilled adhesive. It is also important to realize that a TCE number obtained from a TMA will probably not calculate the same as the z-axis expansion of the adhesive in a joint. The reason for this is that the substrates will constrain the adhesive in the x and y axes, forcing the "unused" portion of the volumetric expansion into the z-axis. In the case of two quartz substrates with TCE's close to zero, virtually all of the volumetric expansion of the adhesive will go into the z-axis. It should also be noted/repeated that this TCE "accommodation" does not cause temperature cycle stress problems for unconstrained planer joints with matching TCE substrates.

Thermal conductivity is another characteristic of an adhesive that can be important to assemblies requiring heat sinking or heat dissipation. Unfilled epoxies typically have very low thermal conductivity, roughly 1 to 5% of copper. Thermal conductivity can be enhanced by incorporating select fillers, such as aluminum oxide or boron nitride if electrical insulation is important; or metal powders if electrical conductivity is desired or not a concern. Even with the highest conductivity fillers, it is rare to achieve bulk thermal conductivity as high as 15% of copper. One can again make an analogy similar to that of TCE above where higher thermal flux may be achievable with a thin unfilled bondline versus a much thicker, higher TC material.

Thermal conductivity, like electrical conductivity, does not appear to follow volume fraction averages, but seems to require some minimum concentration, where the filler particles are sufficiently numerous so that each particle has a statistical probability of touching at least two other particles. This, like any filler addition, can raise handling and application issues.

15.2.1.5. Rheology From a practical point of view, the rheological properties of an adhesive define the way it can or needs to be used, viscosity and thixotropy being the most common and generally useful properties. A low viscosity will permit easy mixing if necessary, syringe dispensing, flow into confined spaces and fast wicking, however, it can also be the cause of unwanted spreading, dripping, sagging and drooling. High viscosity may reverse the advantages mentioned above but not necessarily prevent spread, drooling, etc., especially if a high viscosity material is applied at room temperature and subsequently heated for cure. As indicated earlier, very few data sheets mention the affect of temperature on viscosity. Perhaps the closest practical indicator is for underfill type materials where a flow temperature is suggested along with some measure of time to fill a simulated BGA cavity.

Viscosity can be controlled (up or down) in a number of ways; molecular weight of the resins and curing agents (which can include adducting), both reactive and non-reactive additives, inert fillers (usually only up) and of course temperature. Another technique used to raise viscosity by some end users is to allow a room temperature reactive mixed adhesive to advance (pre-react). Barring historical reasons, the interdependency of time and temperature can cause this method to be technologically challenging for both product consistency and process operating window.

Unlike viscosity, where one can easily develop a sense of a material from a single number, thixotropy is a more difficult concept to grasp. Thixotropy refers to a material's apparent viscosity that becomes lower under the influence of shear. Data sheet references to thixotropic index (TI) are typically based on the ratio of viscosities at two different viscometer spindle speeds. Little standardization exists in setting the parameters for TI, either between suppliers or within adhesive categories. These numbers rarely conjure up a usable mental picture of a material unless at least one of the actual viscosity and test speed conditions are indicated. The TI, however, can be a very useful QC tool for suppliers as it can be significantly influenced by formulation precision, the manufacturing process and air incorporation, as well as storage conditions. For the novice, one of the best ways to describe thixotropy is by a familiar example such as mayonnaise, where it spreads easily and doesn't run or drip. Another example could be the latest in decorator paints where the material will brush or roll easily yet not drip between the bucket and wall or ceiling.

Thixotropy can be a major contribution to the successful processing of electronic and optical assemblies. And perhaps one of the most obvious illustrations of thixotropy in electronics is the typical surface mount adhesive. This is dispensed onto a board, often between electrical contact pads to hold a surface mount component in place during a solder

wave step (often in defiance of gravity). The material is formulated to dispense through very fine needles and to retain its shape until component placement without slumping or contaminating nearby solder pads. The adhesive must also not trail (a like tail) "strings" as the needle is moved from place to place on the board. Other versions of the functional type of adhesive may have slightly different levels of thixotropy to allow pin transfer or stencil application techniques.

15.2.2. Adhesive Chemistry

15.2.2.1. General Many, perhaps many hundreds, of attempts have been made to define adhesives, for the obvious reason that the perfect definition has yet to be recorded. From a practical standpoint, let us consider an adhesive to be "a material that can be applied between two substrates and which will then change to a solid material having desirable (functional) properties." Interestingly, this "definition" would include solders and even water for sub-freezing applications but would exclude a significant portion of the pressure-sensitive adhesive field.

15.2.2.2. Classifications Adhesive can be classified by; one of their reactive groups (e.g., epoxy, acrylic, cyanoacrylate); by reaction product (e.g., urethane, silicone, polyamide); by end use (e.g., construction, paper, medical); by application technique (e.g., spray, screen, hot melt); and by cure characteristics (e.g., thermosetting, UV, anaerobic) to name a few. These casual classifications can lead to confusion outside the specific area of use and especially when used incorrectly as in the case of "epoxy" being used to describe many non-epoxy adhesives or "UV epoxy" frequently used to describe a UV curable acrylic.

In the micro and opto-electronic fields, the most prominent categories of interest are the thermosetting epoxy and UV curable acrylic and epoxy formulations. In addition, some amount of silicone, anaerobic, thermoplastic and cyanoacrylate can be found along with pressure-sensitive types.

15.2.2.3. Cure Specifics A number of scholarly texts exist on the detailed chemistry of adhesives (1–4) and will not be part of this chapter. However, we will try to cover the more practical aspects of how one gets from the pure chemistry to the usable adhesive product. The most common forms of adhesives are generally based on a resin and a curing agent whose primary purpose is to change the resin from a dispensable state (usually fluid) to a solid, capable of holding the two substrates together.

Epoxy resins can vary between low viscosity fluids to meltable solids and can contain two or more reactive epoxy groups per molecule. The most popular resin is the diglycidyl ether of bisphenol A with two reactive groups, and a viscosity similar to heavy motor oil or honey. Other resins include; cycloaliphatic, aliphatic, novolacs and other multifunctional epoxies. In the curing or cross-linking process, a curing agent activates the reactive groups, causing them to combine with the curing agent or themselves, leading to cross-linked networks or higher molecular weight resins.

Much of the sophistication in epoxy formulation lies in the selection of the curing agent, for that mostly determines shelf life, pot life, cure condition (time and temperature), packaging options and to some degree, dispensing options. Resin selection can also influence the same properties but usually to a lesser degree. The most common curing agents for the typical epoxy resin adhesives are based on amines and their relatives including aliphatic, aromatic, cycloaliphatic, tertiary, adducts, amides, imidazoles, etc. On occasion, more than one curing agent can be used to achieve desired handling or performance characteristics.

Another class of epoxy curing agents include the acid anhydrides, but these are used more for potting and casting applications than adhesives, mainly because of longer cure times. The (acid) anhydride curing systems usually require an accelerator such as a tertiary amine. Lewis acid complexes and mercaptans are additional categories of curing agents for epoxy resins. Several of these curing agents can be considered latent by virtue of no (or almost no) reactivity at room temperature, but activate to cure the resin at some elevated temperature. The ways that this may be achieved include; chemical blockage with a sharp disassociative temperature; mechanical blockage as with a meltable wax coating; a solid curing agent with a sharp melting point and a very slow reactivity at room temperature. Mechanical blocking systems intuitively result in a coarser dispersion than can be achieved with liquid curing agents or very finely powdered curing agents dispersed under high shear. The mix ratio of resin to curing agent is usually based on chemical stoichiometry, but can be modified to optimize certain properties and performance.

Another class of curing agents for epoxy resins (mostly cycloaliphatic) are the UV cationic and more recently the thermal cationic agents. These curing agents can be relatively stable (protected from UV) at room temperature, yet solidify rapidly when exposed to intense sources of UV, or elevated temperature for the thermal cationics. Certain chemical compounds (e.g., amines and materials containing sulfur) can inhibit the cure of adhesives based on this cure chemistry. The popularity of these UV systems in the opto-electronics assembly area can be attributed to the fast cure, which sometimes can then be supplemented with additional thermal cure to improve properties such as T_g and chemical resistance.

UV acrylics have been used as optical adhesives for a relatively long time, but have been significantly replaced with thermal or UV epoxies due to better performance. The acrylate monomers (resin) are most often cured by a free radical mechanism which, depending on the functionality of the resin, can yield a linear (thermoplastic) polymer or cross-linked (thermoset) polymer. The source of free radicals could be thermal (e.g., peroxide); UV activators or electron beam. Disadvantages of the early acrylates included high cure shrinkage, outgassing, limited environment resistance, poor high temperature performance and possible air inhibition. Newer generations of acrylate formulations have enhanced the ease of use and the performance, but generally not to the level of thermal epoxies.

Silicones are usually rubbery materials mostly used as sealants and shock and vibration mounting materials, although occasionally used as an adhesive. They are available in a two part, room temperature and elevated curing systems as well as one part moisture, thermal and UV curable systems. The cured silicone rubber typically has very high TCE and low modulus over an extended temperature range, and is resistant to water and many chemicals, especially acids and bases. Exposure to many organic chemicals (e.g., solvents) will most likely cause swelling, sometimes severe. As a sealant, silicones provide excellent water barriers but allow moisture to pass quite readily.

Anaerobic adhesive/sealants are close relatives of the acrylate adhesives that are formulated to cure in the absence of oxygen (e.g., when placed between two substrates) and quite rapidly with an initiator (usually on an at least one of the surfaces). Originally employed as thread locking materials, extensive development efforts have led to other types of adhesive applications. Most notable use in the optics industry is in the field connectorization of optical fiber (inside a ferrule), where a no-mix adhesives and rapid room temperature cures are of interest; oxygen inhibition at the air interface can negatively impact the polishing process.

Thermoplastics or hot melt adhesives, that are applied hot and form a bond when they cool and solidify, are sometimes used for non-critical or temporary "hold down during assembly" applications but they usually suffer from creep/flow under mildly elevated

temperatures. The traditional hot melt adhesive is based on a linear polymer chain with additives for stability and tack. Advanced hot melt systems can incorporated some cure capability that proceeds, usually slowly, after application. The most common mechanism seems to be moisture induced cross-linking.

Cyanoacrylate adhesives, the so-called "super glues," are one part, with or without a primer/activator) that cures very rapidly when pressed between catalytically active surfaces. Issues of performance, such as heat and moisture resistance and poor peel and impact resistance, have limited the usefulness of this class of adhesive in the optical and electronic industries, except for temporary fixturing or holding.

15.2.2.4. Additives Adhesives are rarely just combinations of pure chemicals (e.g., resins and curing agents) as suggested above. Adhesive suppliers devote much of their development energy to modifying the "two chemicals" to enhance either the ease of use or the performance of the cured material and assembly. Some of the additives used include; fillers, flexibilizers, toughenersand adhesion promoters.

15.2.2.4.1. Fillers. Fillers are used for a variety of purposes in adhesives, including density control, viscosity, thixotropy, TCE, mix ratio modification, cost, opacity, color, compressive strength, heat distortion temperature, electrical conductivity and thermal conductivity. Most of the fillers used in the industry are mineral or metal but organic fillers such as ground rubber or wood flour have been used. Silica in numerous variants is probably the most common filler material in high technology epoxy systems. Calcium carbonate, talc, aluminum powder, silver flake, boron nitride and chopped glass fiber are also used for specific purposes. The usual form is a fine particle with a fairly narrow size distribution (achieved by screening, air flotation or other methods). Particle size and surface area will affect the amount of filler that can be added to a specific resin or curing agent without exceeding a rheology requirement. Incorporation of a filler into a liquid adhesive component becomes more difficult as the average particle size of the filler decreases and the viscosity of the liquid increases. Most adhesive manufacturers will incorporate fillers under vacuum and high shear conditions, primarily to eliminate voids (air) and particle agglomerates.

15.2.2.4.2. Flexibilizers. Flexibilizers can be added to resins or curing agents if the adhesive is deemed to be too brittle. Choices for flexibilizing agents include higher molecular weight epoxy resins, non-aromatic backbone epoxy resins (both mono and difunctional), compatible but non-reactive plasticizers and high molecular weight curing agents. Most of the approaches to increasing flexibility will lower T_g of the cured adhesive compared to the non-flex version. In addition, most of the flexibilizing additives will change the pre-cure viscosity of the adhesive, some significantly lower, others higher. The addition of most flexibilizing agents to traditional epoxy systems generally causes the cure rate to slow.

15.2.2.4.3. Tougheners. In some respects, toughening agents appear similar to flexibilizers but in reality, their function is to act as crack stoppers. This is usually accomplished by the formation of discreet particles of a "rubber" within the adhesive matrix. Two approaches have been used; in situ formation of rubber particles separated from a compatible mixture as the resin begins its cure process, and direct incorporation of fine rubber particles into the adhesive matrix as a filler.

15.2.2.4.4. Adhesion Promoters. Adhesion promoters can be added to an adhesive to enhance joint strength, durability and environmental resistance. They can also act to enhance or modify the surface of fillers to change the amount of reinforcement realized or even to modify the interaction of the filler surface with the environment (such as moisture). The most common adhesion promoters are based in silanes but other chemistries, such as titanates and zirconates have also been explored.

In theory, the silane is capable of reacting with the surface of a substrate and also provide a "tail" capable of reacting with the adhesive matrix. The result is a substrate interface that theoretically should have more resistance to attack by aggressive environments, such as high temperature and humidity. The best utilization of an adhesion promoter is to apply it directly to the substrate before applying the adhesive, however, it can also be incorporated into the adhesive so that upon adhesive application, the silanes have a chance to get to and react with the substrate surface. Care must be exercised in selecting the silane reactive "tail" to avoid premature reaction with the adhesive and causing shortened shelf life.

15.3. DESIGN OBJECTIVE

15.3.1. Adhesive Joint Design

15.3.1.1. Stress Sources Whenever possible, it should be the goal of a physical designer to minimize inadvertent stresses on a bonded joint. In the real world, more often than not, the design of a joint is fixed by system requirements and the adhesive must try to make the accommodation. This can, in fact, create situations in which no adhesive could survive and lead to extensive (and expensive) screening of materials (and perhaps processes) from many suppliers; and in the end contribute to the questionable reputation of adhesives.

Stress on a joint can emanate from two distinct sources; external mechanical and "internal," resulting from the interaction of substrate properties, adhesive properties and the changes in the adhesive during cure. Externally induced stresses are usually used to measure a joint's performance, which is in essence the "true" joint resistance to stress minus the preexisting internal stress.

Once a joint is assembled and the adhesive cured, stress in the joint can develop as a result of:

1. Adhesive post gel cure shrinkage.
2. Non-planar adhesive bondline.
3. Substrate TCE differences.
4. Static thermal gradients.
5. Dynamic thermal gradients.
6. Moisture or chemical induced adhesive swelling.
7. Chemical changes at the substrate interface.
8. Presence of bondline shims.

Some of the causes of stress can be reversible (usually on temperature change), others irreversible and even a combination of the two as when a partial delamination or disbond occurs.

15.3.1.2. Substrate Properties The properties of each substrate in an adhesive joint are in many respects as or more important than those of the adhesive, certainly the substrates are the means through which external stress is applied to the joint, and the means which cause or allow internal stresses to be created.

The internal stress can be thought of as composed of two major components; those created by the combination of adhesive cure shrinkage and dimensional constraints of the substrate design; and those created after joint formation (adhesive cure) by the relative di-

mensional changes of the substrates (and adhesive) as the temperature of the joint deviates from the zero stress condition*.

The TCE of each substrate will determine its actual size (adjusted for stress) at any given temperature, so that substrates with different TCE's will change size relative to each other as the temperature changes. This differential will increase proportionally to the change in temperature (in either direction) from the zero stress condition/temperature. Depending on the rigidity of the substrates (section modulus), the bonded assembly may simply warp (as in a bimetallic element) or place shear stress on the adhesive if the substrates are incapable of bending.

In the so-called ideal joint, where the bondline is planer and the TCE's of the substrate are identical, there is essentially no internal stress on the assembly. In this type of a joint (where the substrates are *unconstrained* by fixtures or shims, etc.) the adhesive post gel shrinkage will simply move the substrates closer together and not generate stresses. Since the post gel adhesive is effectively attached to the substrates in the x and y planes of the adhesive, the total volume shrinkage (less the substrate x and y change) must occur in the z-axis or normal to the adhesive plane.

One can demonstrate both the effects of constrained and essentially unconstrained adhesive shrinkage in a simple experiment. By assembling two glass microscope slides on a hot plate with two 5 mil shims near the ends (broken cover slips work well) and wicking an adhesive (e.g., Epo-tec 353 ND or Zymet F-711) into the pre-heated ($\sim$100–125°C) assembly gap. Just before the adhesive begins to gel, voids will form around the shims during the brief period when the material is still fluid by too viscous to backfill, followed by gelation and cure. After cure, one can actually measure the difference in bondline thickness at the center of the slide compared to the ends over the shims. The elasticity of the microscope slides allows the bending to occur near the shims leaving the central portion of the assembly relatively stress free.

In the non-ideal joint configuration, where the bondline is planer, but the two substrates differ in TCE and have a high section modulus, there will be shear stress at the adhesive layer/substrate interface as the assembly temperature deviates from the zero stress temperature. Depending on the thickness and modulus of the adhesive layer, some of that stress may be dissipated, however, since most bondline (especially for optical assemblies) are usually quite thin, the adhesives' ability to absorb the differential expansion is quite limited. This situation can lead to progressive delaminating of the adhesive from one of the substrates, and is most evident upon repeated low temperature exposure such as during temperature cycling. High temperature exposure is less likely to cause delamination in this type of joint design because the higher temperatures are usually closer to the zero stress temperature (therefore minimizing ΔT) and also closer to T_g which significantly lowers the modulus of the adhesive.

More complex joint designs, for example, those with non-planer bondlines and those with the equivalent of shims can become more vulnerable to failure during temperature cycling, elevated temperature exposure and even moisture uptake. Two common examples of poor design and materials choices in the fiber optic arena include the V-groove fiber array with a flat glass plate cover, and a typical fiber ferrule with a funnel end mainly

*The zero stress condition is normally thought of as the temperature at which adhesive cure takes place, however, this is not exactly correct. Firstly, the major reference point for zero stress condition is the temperature at which gelation occurs and then modified upward to compensate for post gel shrinkage in the adhesive. In cases where the substrates differ in TCE or the bondline is non-planer, it may be impossible to attain a true zero stress condition.

used for fiber guidance and after assembly strain relief. In the v-groove array assembly, the etched silicon array is used to provide highly accurate fiber positioning in the x, y and θ orientations. The flat plate makes the third point contact on each fiber (the first two from the sides of each v-groove); and is sometimes used to allow UV light to cure a UV adhesive and to inspect for voids which could compromise end polishing of the fibers. The relatively large volumes of adhesive at the apex of the v-groove and between the fibers, cover plate and wafer enhance the possibility of voids forming as a result of cure shrinkage; as well as disbonds between the mechanically constrained (by the fibers) wafer and plate. The reverse will occur at temperatures above the zero stress temperature where the layer of adhesive will expand and actually lift the plate, causing disbonds around the fibers. This last mechanism can be exacerbated by moisture uptake (whose effect can simulate much higher temperatures in the adhesive).

The second example is the widely used fiber ferrule, where a short length of capillary tube (made from glass, ceramic or metal) is used to mechanically support/protect the end of a fiber and to allow end polishing, handling and sometimes passive alignment. In the most common ferrules, the tube has a center bore with micron or sub micron level clearance for a fiber and a "funnel" like end which is used for a variety of purposes including ease of adhesive injection, fiber entry guidance, fiber buffer termination and fiber strain relief. Depending on the adhesive and the application technique, the annulus (between fiber and ferrule) can be from 50–95% full, which generally prevents movement of the fiber within the ferrule. The incomplete fill of the capillary can be attributed to poor application technique, air entrapment in the adhesive and shrinkage of the adhesive upon cure and cool down. Problems can arise if the funnel area is filled with a rigid adhesive and/or one that has a high moisture uptake. Firstly, if the assembly exceeds the zero stress temperature, the expansion of the cone of adhesive can put a strong tensile force on the fiber and actually cause a fracture in the capillary very near the apex of the cone. This, of course, assumes that the cone adhesive is strongly bonded to the fiber and can be explained by the high TCE of the typical unfilled adhesive in the cone and the resultant movement of cone bulk away from the apex as it expands. Again, moisture uptake can provide the same effect or exacerbate the temperature effect. Another problem related to this assembly is the that the rigid "strain relief" is in fact a stress concentration that can cause fiber fracture under mild side loading of the buffered fiber relative to the ferrule.

15.3.1.3. Non-Planar Bondlines In adhesive joint designs where the adhesive bondline is not planar, both the TCE and cure shrinkage of the adhesive becomes significant reliability factors. Depending on the actual geometry of the joint design, a non-planar bondline will usually cause a partial disbond to occur during excursions from the zero stress temperature. Exceptions can arise from compliant substrates and some cases of bondline fold symmetry (visualize two pieces of angle iron bonded together).

In a special case of two flat substrates bonded with a wedge of adhesive (substrate tilt), the joint would likely resist temperature cycling but will cause temperature dependent relative movement of the substrates and account for poor optical assembly performance. To illustrate a different effect of non-planer joints, imagine a block of, let's say aluminum, bonded into the apex of an uneven angle of the same material. The longer leg of the angle will constrain movement along that leg, preventing the short leg from following the normal shrinkage of the adhesive between it and the block. The result is that the shorter bondline will experience severe normal stresses, not only just as a result of cure, but also the added TCE shrinkage as the assembly cools from the zero stress temperature.

15.3.1.4. Thermal Gradients The designer of an adhesive joint needs to be aware of the potential for thermal gradients to raise stress levels in bonded assemblies. Even ideal adhesive joints assembled at a uniform temperature can experience stresses if exposed to thermal gradients where one substrate is at a different temperature than the other substrate. In reality, this situation can mimic that of a substrate TCE mismatch. An even more potentially damaging scenario can be encountered under thermal shock conditions where a severe dynamic gradient can sometimes damage both bondline and substrate.

15.3.1.5. Moisture or Chemical Absorption Most adhesives (and other organic materials) will absorb moisture (and other chemicals) when environmentally exposed. Conditions of exposure, such as time, temperature, pressure and concentration can influence the total amount of absorption. In addition to those factors, geometry and diffusion rate will affect equilibration time. The absorption of moisture or solvents usually results in an increase in volume or swelling and can also affect other properties such as modulus and T_g through a plasticization mechanism. Most of these changes are reversible, although sometimes T_g can be permanently lowered. In the case where exposure to moisture actually leaches material from the adhesive (as demonstrated by lower re-dry weight than the original pre exposure weight), reversibility is less likely and T_g might even increase.

Stress on a joint can change due to moisture absorption in a fashion similar to raising the temperature. Another way of looking at this is to expect that the zero stress temperature of a moisture saturated bondline would be considerably lower than that of a dry sample. If one then adds elevated temperature to the assembly, stress conditions can increase dramatically over a dry sample. Some studies in the past have shown that diffusion into a bondline (even an ideal joint) can cause a swelling gradient apparently related to the diffusion characteristics of the adhesive. This "wave front" separating the dry from the damp (swollen) adhesive can add to joint stresses.

15.3.1.6. Substrate Surface Chemical Changes Many substrates are bonded with the surface in a relatively active state. This provides the opportunity for post bond reaction with chemical species present in the environment. Probably the most common example would be moisture-induced oxidation (corrosion) of the substrate at the adhesive/substrate interface, similar to rusting of steel under a layer of paint. Most pre-bond surface preparations (at least of metals) try to stabilize the surface and then try to prevent corrosion, even if it means replacing a native oxide layer with a more tenacious and less reactive oxide or other chemical species.

When the chemistry of a substrate surface at an adhesive interface changes, the original bond strength will usually be compromised. One can easily envision the case of oxide growth, but there is also the situation where the adhesive may be capable of reducing substrate oxides to which the adhesive was originally bonded. Either case can cause weakness and/or failed joints.

15.3.1.7. Performance Qualifications During the early development stage of a new optical device or component and occasionally a microelectronic component, it is often necessary to build prototype or even proof of concept designs to gage performance and ultimately to prepare performance specifications for the commercial product. This process is generally interactive in both design and assembly as its environmental window (e.g., temperature, humidity, shock, etc.) migrates from a room temperature operating prototype to a viable commercial range. Typically, the change in performance of a device measured over a temperature range can be useful in analyzing the design factors responsible and thus the

remedy. Subtleties in the repeatability, or lack thereof, of the measured performance on temperature cycling can further aid the design iteration process.

Another part of the design process involves the commercial viability of the assembly steps and of course the materials selection process. Ideally, timely expertise in all of these areas would be available for any new design project, but unfortunately, this is usually not the case, often resulting in significant project and development cost overruns.

15.3.2. Manufacturing Issues

15.3.2.1. General The goals of manufacturing engineers are on occasion somewhat diverse from those of the design, quality and sales portions of the organization. Typically, a cost driven position in manufacturing can lead to compromises in a products' performance and reliability. It is very important that each of those organizational areas work together to achieve a product that will meet or surpass a customers needs. This may mean the use of a more costly or longer manufacturing step to achieve better performance and reliability. There are, however, cases where a lower manufacturing costs process can still provide a product that will meet the reliability needs of a customer and makes economical sense. An example might be the use of a non hermetic package for a central office type application. Many times the use of extra time or care in an adhesive bonding process can lead to yield improvements and perhaps even lower manufacturing costs.

The earlier on in a new project that physical designers and manufacturing engineers can work together, usually the more successful the result, especially if each group gains a broader perspective of the others challenges. Early and frequent design reviews can be a productive approach, perhaps using a facilitator with an interdisciplinary background.

The following sections will address manufacturing issues specific to the practice of using adhesives in the assembly of optical and electronic devices.

15.3.2.2. Fixturing The design of bonding fixtures can be a challenge, especially in the optical assembly field where submicron positioning of components is often required. Additionally, an adhesive must be applied to the joint and then the assembly must undergo a cure step before unfixturing. Consideration must also be paid to the affect of adhesive shrinkage on the joint stress as well as thermal movement of the fixturing during the cure step.

Fixturing tools can be quite expensive and have a major impact in a volume manufacturing environment where cycle time requirements and cost influence the number of fixtures needed.

Fixturing in micro-electronic systems is usually not as elaborate as with optical components and sometimes even rely on the self-fixturing afforded by the rheology of adhesives; as in the placement of surface mount components prior to cure. In the case of chips/BGA underfill, the component is already fixtured by the solder interconnections.

In the design of fixturing for optical components assembly, some of the following may be considered important:

1. Gripping interface (part holders).
2. Degrees of movement.
3. Robustness.
4. Precision of movement.
5. Repeatability.
6. Backlash.

7. Thermal stability.
8. Ease of loading.
9. Ease of alignment.
10. Mechanical stability.

15.3.2.3. Curing there are several approaches to curing adhesives in a manufacturing environment. Some of the factors influencing the cure selection will include those associated with the adhesive, the fixturing, the components and the production schedule.

Thermally cured adhesives, be they two component room temperature curable or one component (premixed), will generally cure faster at higher temperatures. If, for some reason, a low temperature cure is necessary, relatively long cure times will be encountered (e.g., 20 minutes to 24 hrs), obviously depending on the formulation. For some applications where fixturing is quite simple (as for instance a small clip), many units can be assembled and allowed to cure overnight. The shorter low temperature cure time comes with the typical trade off of short pot life, frequent mixing and usually significant material waste. By using a higher cure temperature, adhesives with longer pot lives can be selected. Again, with simple fixturing, this approach allows both batch oven and continuous or tunnel oven curing. Cure times of relatively long pot life premixed adhesive compositions can be as short as several seconds to an hour or so, again, depending on the formulation and temperature. Some of the means used to heat assemblies include focused IR, forced hot air, resistance wire "ovens," inductive heating, magnetic (eddy current) heating, embedded heating elements, as well as the simple hot plate. Depending on the heating method, attempts to achieve very short cure times may cause thermomechanical problems in the assembly due to the possibility of severe thermal gradients in both the substrates and the adhesive during the pre and post gel stages of the adhesive.

Adhesives designed for UV/visible activation generally appeal to the process engineers because of no mixing, long pot life and rapid cure and usually no heating.

Some formulations that have dual cure capability allow gel of the adhesive with UV exposure and subsequent thermal post cure after removal from fixturing so that the adhesive may develop better properties. Implicit in the use of this type of adhesive is the need for at least one substrate to be transparent to the radiation that activates the adhesive. An issue of concern for these adhesives is the possibility of cure inhibition due to the presence of certain chemical species such as amines and sulfur compounds. One must also be aware of the possibility of thick bondlines causing significant part movement due to shrinkage and the possibility of variable properties from run to run or even within a single bondline if the UV source varies in intensity, duration and even temperature.

Other curing mechanisms such as surface activation/catalysis, moisture reaction and anaerobic are less common in the optical bonding field and will be bypassed here.

15.3.2.4. Bondline Thickness Thick bondlines (e.g., over 3 mils) are usually less desirable than thin ones. Part of the reason is the control of that thickness becomes difficult along with possibility of tilted substrates and voids in the adhesive. With a thick bondline in an optical assembly, one can expect shrinkage movement in the z-axis (normal to the adhesive bondline). After cure, temperature and humidity will affect z-axis as well as pitch and roll if there is a wedge of adhesive. Obviously, the thinner the bondline, the smaller these effects.

15.3.2.5. Repeatability In any manufacturing operation, once a process/procedure has been optimized, it is usually desirable to be able to repeat that on a day to day basis. Some

of the adhesive related variables that could lead to poor repeatability include; adhesive consistency (formulations), mixing accuracy; property variation within the adhesives' pot life. Other factors that can influence manufacturing repeatability include; size of gap (positional uniformity), cure environment consistency, substrate surface contamination, excessively moist or dry humidity, static electricity, voids and trapped air.

15.3.2.6. Voids There are five sources of voids in an adhesive bondline; air mixed into the adhesive, air trapped in the bondline by the assembly process, boiling of an adhesive component, voids created by pre gel shrinkage of a constrained bondline and delamination caused by post gel shrinkage of a constrained bondline. The first three sources of voids are process related and usually easily corrected, while the latter two are related to the physical design of the joint, and can usually only be fixed by modifying that design. The exception being a change in the filler content of the adhesive or gel temperature to reduce overall shrinkage. Air voids are generally less damaging than shrinkage voids, unless they represent a significant portion of the bondline and compromise the joint strength. One must also be concerned about he possibility of water transpiration for seriously porous adhesive bondlines.

15.3.2.7. Surface Preparation Generally considered to be the cornerstone of good bonding practice, surface preparation is usually costly and time consuming from a manufacturing point of view. And it is frequently a subject of debate between proponents of quality, reliability and manufacturing costs, especially when the development and qualification process time constraints limit the collection of proof test data. Unlike the aerospace and automotive industries, where resources are more apt to be available and time to market is not as critical; the micro and opto-electronic fields are often plagued with judgment calls and trade-offs. That said, the obvious approaches to surface preparation are (in order of ascended cost):

1. None to minimal.
2. Just enough to pass requirements.
3. Best possible.

Some of the surface preparation techniques used include; air blow, sanding/sandblast, solvent wipe/rinse, detergent wash, chemical etch, firing, plasma etch, silane treatment, surface oxidation, conversion coatings, plating, priming, etc. The choice or appropriateness of techniques or combination of techniques will also depend on the nature of the substrate.

15.3.2.8. Adhesive Mixing Most adhesives are mixtures of at least two chemicals, generically called resin and curing agent (plus additives). Depending on the stability or pot life of the mixed adhesive, a manufacturer may supply the adhesive as two separate components (each of which could contain at least one active and none to many inactive materials). The separate component systems are usually (but not exclusively) those with short pot lives and that can cure under fairly mild conditions. The main disadvantages in using two-part adhesives are that proportioning (weighing), mixing and air entrapment can lead to suboptimum results. For example, many room temperature adhesives have mixing ratios in the 10–1 range, where weighing accuracy of small batches can be challenging, especially in a manufacturing environment. Add to that the problem of trying to get a homogeneous mixture with tools and containers all wetted by the components, as well as, the problem of air mixed into the adhesive.

The use of pre-weighed plastic packages can eliminate some of the proportioning inaccuracies and offer a convenient mix container. Issues of air bubbles and homogeneity

can still be a concern depending on component viscosity, package wall thickness/stiffness and the operators dedication. In volume operations, the use of a mix-metering machine, that can proportion and mix two component systems from reservoirs then dispense precise amount of adhesive for each part or group of parts, can be cost effective. If mechanical mixers are used, maintenance of the mixer can be a problem, otherwise, static mixers can be used and replaced as needed. Although the initial cost of these machines can be high, the wide variety of material properties and mix ratios capable of being handled provides a great deal of application flexibility for large production runs.

Although usually more costly than two (reactive) component systems, pre-mixed and pre-mixed frozen versions may actually be very cost effective. Many manufacturers will mix their adhesives under vacuum to prevent air incorporation, prepare large batches (thus increasing proportion accuracy) and load multiple syringes or other containers, provide a batch quality control for all the containers and quick-freeze them to preserve pot life. Some of the more latent curing agents can be used to make adhesive formulations that are stable at room temperature for long periods of time, and therefore not require low temperature storage. Variations of the latter include the pre-form, sometimes available as a shaped pressed powder that melts and flows (controlled) at cure temperature; and the thin bonding film with or without reinforcement.

15.3.2.9. Application Techniques The placement of the correct amount of adhesive in the correct place and in the correct (optimum) fashion can involve an interesting number of options. Some of the techniques for adhesive placement include; dab, spatula, trowel, syringe, screen, stencil, preform, pin transfer, platen transfer, ink jet. The first four are generally used in low volume applications while the rest fit better into multiple patterned deposits as in a surface mount adhesive. The selection of technique may be limited by the properties of the adhesive and vice versa. For example, a low viscosity adhesive, which might work well in a syringe application, might also work with a fine screen but probably not with a stencil or an ink jet.

When a manual drop/dab/blob of adhesive is placed on a substrate and the second substrate pressed into place, several, less than ideal, bondline characteristics may result. First, the odds strongly favor either an inadequate amount or an over-abundance of adhesive leaving peripheral voids or messy fillets. Second, depending on viscosity and the aspect ratio (bond area over bondline thickness), pressure on the substrate may be insufficient to achieve the desired thickness. Third, the substrate pressure may be off normal, yielding a wedge of adhesive. Fourth, voids due to trapped air or mixed-in air can expand on elevated temperature cure and result in partial bondlines. And fifth, absorbed moisture and other gas on the substrates can influence the integrity of the adhesive substrate interface.

In the optical component field, a particularly attractive application technique exists that eliminates most of the above problems, and has at lest three other significant advantages. It is the "hot wicking" technique. It works extremely well in critical alignment light path joints with micron or sub-micron bondlines and with adhesives that cure rapidly at a temperature where the viscosity drops precipitously before gel. The process is quite simple; requiring pre-heating of the substrate (and presumably the immediate fixturing) so that uniform and stable elevated temperatures exist at the bondline (usually 100–120°C); the substrates are brought into close contact (co-planar); alignment achieved and then an edge of the bondline is touched with a drop of adhesive. The adhesive will wick across and fill the entire joint without voids, flushing out volatile contaminants from the surface and then cure rapidly (assuming the correct choice of adhesive) usually within 1–5 minutes and without relative substrate movement except in the direction normal to the plane of the

adhesive to compensate for post-gel volume shrinkage. The extra advantages of the hot wicking process are rapid gradient-free cure, the bondline is extremely thin and coplanar, the alignment process is set at the cure temperature and post cure is usually not required.

15.4. FAILURE MECHANISM

15.4.1. General

Much of the discussion here will be a consolidation of information detailed in other parts of the chapter. Basically an adhesive joint fails or degrades because:

1. The adhesive changes.
2. The interface changes.
3. Interfacial stresses exceed the adhesives' capability.
4. Externally applied stresses exceed the adhesives' capability.
5. Combination of the above.

Some of these mechanisms are strictly adhesive related, others are a function of the environment and still others are related to an inappropriate joint design. An experienced practitioner can often predict a failure mode just from viewing an assembly design, observing performance data under temperature and environmental exposure or dissecting failed parts.

15.4.2. Adhesive Changes

Changes in an adhesive can be four basic types; thermal degradation due to temperatures that cause irreversible changes in the chemistry of the adhesive, changes in the volume/density of the adhesive as a result of crosslinking (also irreversible); changes due to TCE after cure, which are generally reversible; and changes caused by exposure to chemical environments (e.g., water or solvents) which can be both reversible or non-reversible if, for example, a solvent extracts some material from the adhesive. All of these changes can affect interfacial stresses as a function of the joint design.

15.4.3. Interfacial Changes

Changes at an interface can be caused by chemical reaction with moisture, oxygen, reactive materials in the adhesive or even with the substrate. The most common effect is one of surface oxidation similar to rusting, however, the opposite effect of reduction has also been observed. Physical change in the structure of plated surfaces is another possibility for joint failure. All of these changes can lead to diminished joint strength and ultimate joint failure simply because the adhesive is essentially being "pushed off" the substrate to which it was originally attached at the time the bond was made.

15.4.4. Interfacial Stress

Reasons noted in the preceding paragraphs can cause interfacial stress, but in addition to those noted factors, there is the inherent joint design contribution that can exacerbate both adhesive and interface stress. Examples of joint design stresses include shims two step bonding, adhesive wedges and cross joint thermal gradients.

15.4.5. External Stress

Adding external applied stress to a joint simply enhances the effect of all the other sources of stress present. This is why, for example, that the joint strength of a non ideal joint (different TCE substrates and/or non planer bondline) will decrease as the test temperature gets farther away from the zero stress temperature. Shock and vibration testing is essentially another form of external stress that can add to the "internal" stresses.

REFERENCES

1. S.R. Hartshorn, Structural Adhesives, Chemistry and Technology, Plenum Press, New York, 1986.
2. L.-H. Lee, Adhesive Chemistry Developments and Trends, Plenum Press, New York, 1984.
3. C.A. May and Y. Tanaka, Epoxy Resins Chemistry and Technology, Marcel Dekker, New York, 1973.
4. A. Pizzi and K.L. Mittal, Handbook of Adhesives Technology, Marcel Dekker, New York, 1996.

16

Multi-Stage Peel Tests and Evaluation of Interfacial Adhesion Strength for Micro- and Opto-Electronic Materials

Masaki Omiya, Kikuo Kishimoto, and Wei Yang

Tokyo Institute of Technology, Graduate School of Science and Engineering, Department of Mechanical and Control Engineering 2-12-1, O-okayama, Meguro-ku, Tokyo 152-8552, Japan

16.1. INTRODUCTION

Rapidly development of the electronic products requires small, high density and functional devices. Many layers deposited on silicon substrate can accomplish several kinds of functions in one package. Moreover, in these days, 3-dimensional packaging and assembling technologies have been developing all over the world (e.g., [1]). Multi-layer technology is a key for developing electronic products in future.

The ensuring of reliability is the one of the critical issues in micro- and opto-electronic devices. Those electronic devices contain several kinds of metal or polymer thin films. Due to intrinsic/thermal residual stresses in films or substrates, or elastic/lattice mismatch between film and substrate, the delaminations between layers sometimes occur. The debonding between layers brings about the failure of devices and it might be the source of a tragic accident, since, nowadays, the electronic devices are closely related to human life. One need to design those devices to work well through its entire life time and the information of interfacial strength is essential to designing the reliable devices. Therefore, it is important to evaluate the interfacial strength precisely and the development of reliable testing methods for evaluating the interfacial strength is needed.

Meanwhile, attention has been directed to an electrically conductive ceramic film, which is deposited on a polymeric substrate [2–4]. The applications of those conductive films are used for the display of mobile computers, cellular phones or the flexible paper type display. Those conductive films have the advantage of low power consumption or flexibility to deformation. The popular components of those films especially for the display use are ITO (indium tin oxide) and PET (poly(ethylene terephthalate)). It is well known that the mechanical properties of polymers, such as tensile strength or rupture strain, are degraded by the irradiation of ultraviolet (UV) rays [5]. When the polymer-based conductive films are

used in the open air, it is necessary to consider the degradation of the mechanical properties in designing of such products. The bonding mechanism between ceramic and polymer is mainly intermolecular force. When UV rays irradiate polymeric materials, the principal chains are broken down and oxidized. Those reactions may affect the bonding strength between ceramics and polymers.

In view of interface mechanics, interfacial strength depends on the phase angle of loading. To assess the interface strength, one need to conduct the interfacial fracture tests under a wide range of phase angle. A lot of researches related to the interfacial fracture tests have been published and a detailed review of the appropriate mechanics has been given by Hutchinson and Suo [6], Evans and Hutchinson [7], Evans et al. [8].

Especially for the adhesion strength of thin films on substrates, the conventional methods can not be applied. Thin film structures are widely used in nano-machines or electronic devices. Therefore, the development of the testing method to evaluate the adhesion strength of thin film and multilayer thin films is needed. Russel et al. [9] have used tape testing and scratch testing methods to measure the interfacial strength of Cu/SiO_2. The tape test is qualitative and often not a reliable test, useful only for testing weakly adhering film. The scratch test is semi-qualitative, in that the normal load at which a predefined failure event or morphology occurs is defined as a measure of adhesion. While these semi-qualitative tests are simple and informative, they are incapable of incorporating all the relevant parameters.

One such quantitative test that retains the same ease of preparation and test conduction as the scratch test is the indentation-induced delamination test. Marshall and Evans and Marshall et al. [10,11] have presented the fracture mechanics analysis of indentation-induced delamination of thin films. Evans and Hutchinson [12] have analyzed the mechanics of the delamination and spalling of compressed films or coatings by using a combination of fracture mechanics and post-buckling theory. Rosenfield et al. [13] have also conducted the indentation-induced delamination tests. They have compared the indentation tests with double-cantilever-beam technique, four-point flexure-beam technique and finite element analyses. Bahr et al. [14] have conducted nano-scale indentation-induced delamination tests. They have used an acoustic emission in conjunction with nanoindentation tests to monitor a delamination or cracking event. Indentation methods generally rely on the formation of a dilated plastic zone in the film to cause the film to blister [15]. Values of G_c, which is fracture toughness of the interface, are related to the indentation volume (or plastic zone size) and extent of debonding. Therefore, this test method suffers from a limitation similar to that the tape test in that it is limited to a very weakly adhered film. In well-adhered films, indentation fails to produce a delamination unless in ordinarily, high loads or depths are used, in which case the substrate deformation renders deconvolution of the adhesion energy from test parameters impossible. However, a modification of this test method has been developed by Kriese et al. [16,17], Gerberich et al. [18] in which the use of a thin hard coating film on the original film to constrain the plastic deformation and brings about the delamination between the original film and substrate.

Other quantitative method for interfacial strength is the pressurized blister test (Jensen [19], Jensen and Thouless [20]). This testing method has been successfully developed and analyzed for thin polymer films, but is often compromised by the inherently compliant loading system, chemical interactions between the debond and the pressurized environment (stress-corrosion cracking), and the etching or machining procedures are needed to produce the cavity.

Some novel testing methods have also proposed. Bagchi et al. [21], Bagchi and Evans [22], Zhuk et al. [23] have developed “superlayer tests,” to measure the debond-

ing energy for thin metalization lines on dielectrics. They fabricated Cu films on various dielectric substrates and evaporated superlayers on Cu films to constrain the plastic deformation of Cu layer. Kitamura et al. [24–26] have used a sandwich specimen, where the deformation of thin films is tightly constraint by substrates for preventing the plastic deformation and fracture of thin films and measured the interfacial strength of nanoscale thickness film. Nakasa et al. [27], Zhang et al. [28] have developed the edge-indentation method to measure the delamination strength of thermally sprayed coatings.

The most well known and straightforward method to delaminate the film from the substrate is the peel test. Peel test is a simple mechanical test to measure the adhesion strength, especially for the case of a thin film deposited on a substrate. Many experimental efforts and analyses have been devoted and a comprehensive survey on the earlier development of the subject can be found elsewhere, see Kim and Aravas [29]. For purely elastic case, all earlier works [30–40] identified the following relation among the adhesion strength Γ, the peeling force P and the peeling angle ϕ formed between the interface and the peeling force:

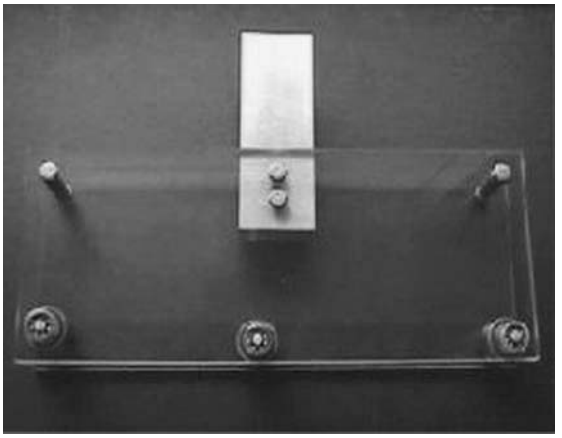
$$\Gamma = P(1 - \cos\phi), \tag{16.1}$$

where P is the forces per unit width of film. When evaluated according to (16.1), the symbol Γ contained a contribution from the residual stress. Moreover, the possible dependence of the adhesion strength on the peeling angle ϕ was not clear at that time. Extensive works [41–47] were devoted to the plastic deformation of the peel as it detached from the substrate, and bend through the moment-curvature hysteresis loop (including plastic loading, elastic unloading, plastic reverse loading, and elastic reverse unloading). A cohesive strip along the leading portion of the interface gives a new twist to the problem. The work by Wei and Hutchinson [48,49] emphasized the influence of cohesive strength improving the nearby plastic dissipation in film and substrate. Their cohesive law [50], however, is normalized in such a way that only an isotropic response with respect to decohesion direction can be accommodated. Other interface cohesive laws (e.g., [51–53]) elaborated several delicate issues of interface debonding. The anisotropic cohesive law by Ma and Kishimoto [51] has the potential to predict a concave adhesive strength versus phase angle curve.

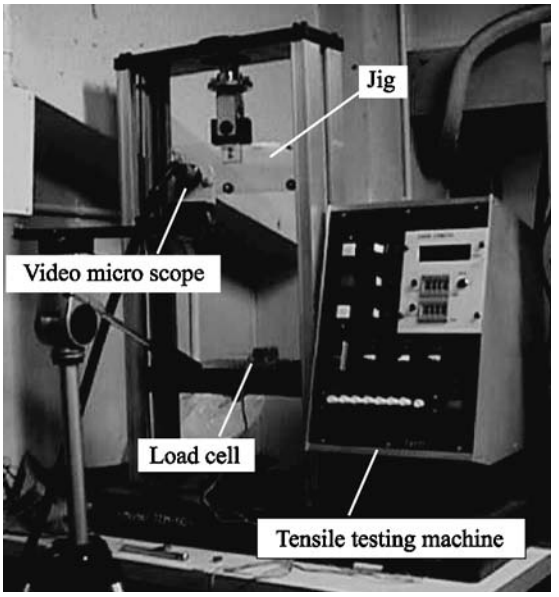
For the interfacial strength of materials used in micro- and opto-electronic devices, Park et al. [54–56] measured the interfacial fracture energy of Cu/Cr/polyimide system by 90° peel test. However, they conducted only one peel angle and it is not enough to discuss the interfacial strength, which is depend on the phase angle. A method of multi-stage peel test (MPT) is proposed in this paper to tackle the issue of measuring adhesion strength as a curve of phase angles. A testing fixture is presented that allows the application of different lateral loads. Balance between the lateral loading and that projected by the peeling load gives rise to a specific peeling angle. Different peeling angles may result for a single film-substrate specimen if one deliberately varies the lateral loads. A steady state peeling load can be achieved after certain amount of decohesion under a prescribed lateral load. MPT involves the measurements on steady state peeling loads for an incremental sequence of lateral loads. These steady state peeling loads are used to correlate the adhesion strength versus phase angle curve.

The plan of this paper is outlined as follows. Testing scheme for MPT will be described in the next section. In Section 16.3, peel tests for copper/chromium/polyimide/silicon substrate will be presented. Those structures are widely used in chip scale packages

(CSP). We will discuss the effect of copper thickness on the interfacial strength and the multi-layer delamination will be considered. The interfacial strength between the conductive thin ceramic film and polymer substrate will be presented in Section 16.4. Those polymer-based films are damaged by ultraviolet rays. We will discuss the degradation of mechanical properties of polymer film and interfacial strength between ceramics and polymer substrates. We also carried out the *in situ* observation of surface cracks on the ceramic layer during the tensile test. The interfacial strength affects the crack formation on the ceramic layer. The concluding remarks with the limitation of multi-stage peel test and future studies will be shown in Section 16.5.



(a)



(b)

FIGURE 16.1. The apparatus of multi-stage peel test. (a) Special jig for multi-stage peel test. (b) Setup of the testing machine.

16.2. MULTI-STAGE PEEL TEST (MPT)

Peel tests have been developed for the evaluation of interfacial strength in practical usage, for example, adhesive strength of thin films, coating films. Many papers related to the peel test have also been published (e.g., [29,46]). The most attractive feature of multi-stage peel test is that it is possible to evaluate the interfacial strength of thin films under various phase angles for only one specimen.

16.2.1. Testing Setup

A specially designed apparatus shown in Figure 16.1(a) facilitates the MPT. The special jig is attached to the upper cross head [Figure 16.1(b)] that moves upward at a controlled peeling rate. The movement is recorded by an extensometer. The peeled film is calmed to the lower cross-head that is immobile during a test. A load cell is installed adjacent to the peeling end that records the history of peeling force P. The schematic representation of the MPT is drawn in Figure 16.2. The specimen is put on two roller bearings and by pulling the film, the film delaminates from the substrate. A film/substrate assembly can move horizontally on the bearings with suppressed friction. The friction between rollers and specimen is controlled within less than 0.1 N. During the MPT, the peel front keeps staying near the central bearing and it makes it possible to measure the peel angle continuously. The peel angle was measured and recorded by using the video-microscope and digital video recorder.

As a departure from the conventional peel test, the film/substrate assembly is stretched horizontally by a dead weight P_h through a pulley system that is controlled during the test (in Figure 16.3). Under quasi-static peeling, balance of forces in the horizontal direction predicts the following peel angle:

$$\phi = \cos^{-1}\left(\frac{P_h}{P}\right). \tag{16.2}$$

Vertical component of the peel force is countered by roller supports. To conduct peeling, the force P should be larger that the horizontal stretching force, P_h. The peel angle

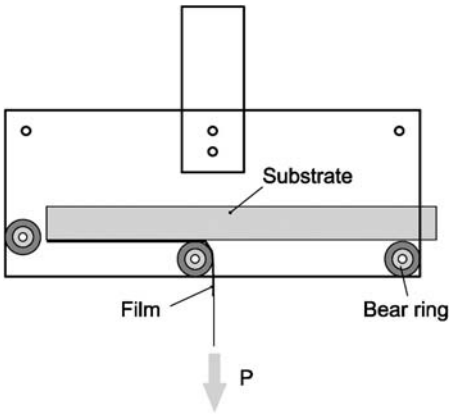


FIGURE 16.2. The schematic representation of peel test.

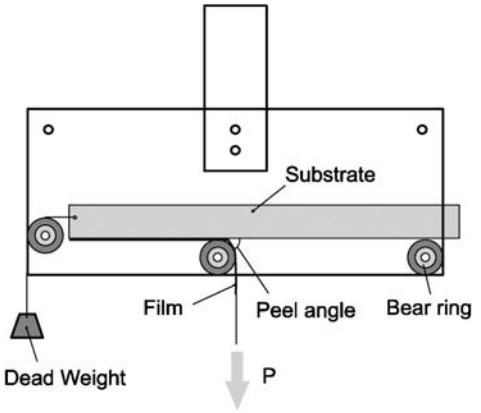
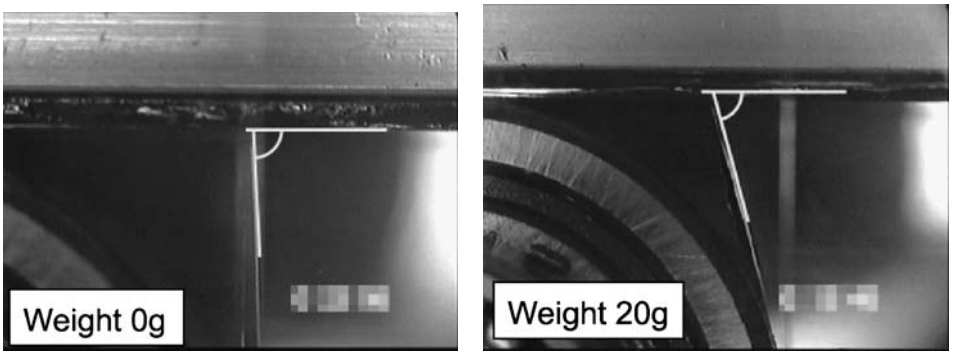


FIGURE 16.3. The dead weight attached on the specimen for changing the peel angle.



(a) (b)

FIGURE 16.4. The effect of the dead weight on the peel angle. (a) No weight. The peel angle is 89°. (b) The dead weight is 20 g. The peel angle is 75°.

decreases monotonically with the increase of P_h. By varying P_h, a wide range of peel angles (between 30 degrees under a large horizontal force to almost 90 degrees for negligible horizontal force) can be achieved. The peel angle can be measured independently and *in situ* by a microscope horizontally mounted on the side facing the test machine, see Figure 16.1(b). A typical image is shown in Figure 16.4. When the dead weight is attached to the specimen, the peel angle changed due to the horizontal force. The relationship between the horizontal force and the peel force decide the peel angle. This effect causes the phase angle shift at the peel front. Therefore, the mixed mode delamination tests under the wide range can be possible by the MPT.

16.2.2. Multi-Stage Peel Test

Figure 16.5 describes a typical curve for the evolution of peel force when the peeling rate maintains at 5 mm/min. A steady state emerges after about 5 mm of peeling length. Samples peeled at other rates (from 1 mm to 5 mm/min) deliver similar results of steady state peeling convened at a peeling length about 5 mm. If the dead load (that controls

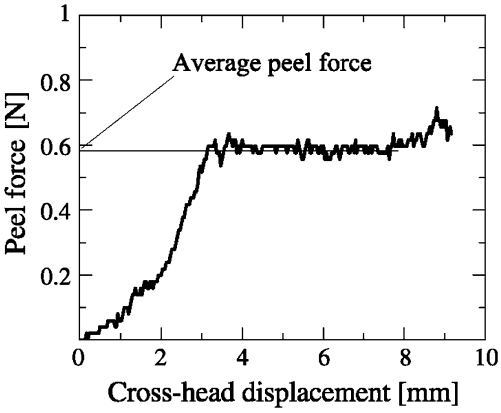


FIGURE 16.5. Typical example of peel force evolution during the MPT.

the horizontal stretching force) is applied incrementally, and a minimum peeling length of 5 mm is imposed on each peeling increment, one can have a multi-stage peeling test in one pass that records the steady state peeling forces at different peeling angles. The apparatus has a lateral span of 220 mm, and the range available for peeling is about 80 mm. Accordingly, over 10 peeling stages can be accommodated in one testing block. Those steady state peeling forces will correlate to the phase angle dependence of adhesion strength in the next section.

16.2.3. Energy Variation in Steady State Peeling

After the attainment of a steady state, the peeling configuration stabilizes. The load point displacement can be chosen as the time variable. Possible rate dependences of the film and the adherent become implicit. Analysis for steady state peeling is further simplified by considering the energetic aspect of the system, as schematically shown in Figure 16.6. During each peeling increment of Δl in a steady state, neither the peeling configuration nor the energy storage and dissipation change for most portion of the system for an observer fixed spatially, say, to the central roller. The only difference in energy exchange consists of a segment of length Δl far behind the peeling edge converting to a segment of the same length far ahead of the peeling edge. The corresponding change in the substrate is negligible since the substrate parts of both segments are essentially stress-free.

Stored in the film far behind the peeling edge is the elastic energy caused by the residual stresses generated during film deposition, the density W_{res} of this energy can be computed as:

$$W_{res} = \frac{\alpha(\nu)}{E} \int_0^h \sigma_{res}^2(y)dy, \tag{16.3}$$

where E and ν denote the Young's modulus and the Poisson's ratio of the film, and σ_{res} denotes the horizontal residual stress stored in the film during deposition. If σ_{res} is uniform across the film thickness h, then

$$W_{res} = \frac{\alpha(\nu)}{E}\sigma_{res}^2 h. \tag{16.4}$$

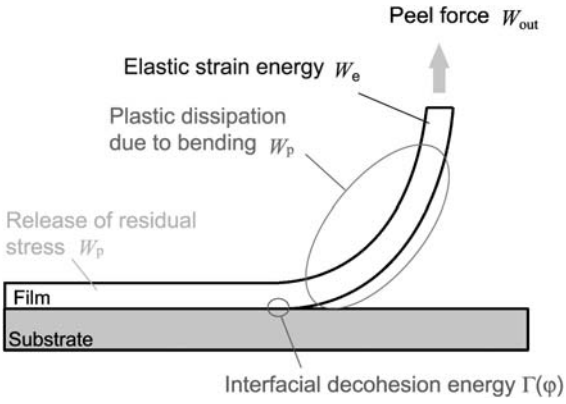


FIGURE 16.6. Energy balance under steady state peeling.

The coefficient $\alpha(\nu)$ depends on the stress state in the film and the Poisson's ratio. It equals to $(1-\nu^2)/2$ for a plane strain film, i.e., see Yu et al. [57] and equals to $1-\nu$ for an equal biaxial stress state caused by thermal mismatch, see Yang and Freund [58].

Near the film end far away from the debonding edge, the energy density comes from the following sources: (1) the decohesion energy, $\Gamma(\varphi)$; (2) the plastic dissipation, W_p, of a certain film moment-curvature hysteresis terminated at the film state near the pulling end; (3) the strain energy due to the residual stress in the film; and (4) the elastic strain energy W_e of the film near the pulling end. For a plane strain situation, the last contribution includes a stretching part of $[(1-\nu^2)/(2Eh)]P^2$ and a coiling part of

$$\frac{Eh^3}{24(1-\nu^2)K_{coil}^2},$$

where K_{coil} is the coiling curvature at the end of the film when it is released from the peeling grip. Contribution from the elastic strain energy W_e is usually small when compared with the others.

Beside the variation of the energies stored or dissipated in the system, the work of the peel force P and that of the dead weight P_h contribute to the potential energy of the system. The rate of those works is denoted as W_{out}, and is given by:

$$W_{out} = P - P_h, \tag{16.5}$$

where, θ is the peel angle measured in the MPT. Apart from the steady state assumption, Equation (16.5) is derived under the ignorance on the work done by all frictional forces. The absence of friction also leads to the equivalence between the raising rate of the dead load and the pulling rate of the lower cross-head when a steady state prevails. Conservation of the potential energy gives:

$$W_{out} = \Gamma(\varphi) + W_p + W_e + W_{res}, \tag{16.6}$$

that is,

$$P - P_h = \Gamma(\varphi) + W_p + \frac{Eh^3}{24(1-\nu^2)K_{coil}^2} + \frac{1-\nu^2}{2E}\left[\frac{P^2}{h} - \int_0^h \sigma_{res}^2(y)dy\right], \tag{16.7}$$

when subjected to a plane strain condition. In Equation (16.7), the material properties such as the Young's modulus E, the Poisson's ratio ν, and those needed in determining W_P can be measured independently. The film thickness h and the dead weight P_h are known *a priori*. The steady state peeling force P and the residual coil curvature K_{coil} can be measured during the test. Appropriate analyses [29,47,48] can be invoked to evaluate W_p. However, one still has the difficulty to distinct the adhesion $\Gamma(\varphi)$ from the contribution of the unknown deposition stress σ_{res} and this value should be measured by another experiments.

The difficulty to obtain the interfacial strength from the above equation is how to estimate the plastic dissipation in detached film. Kim and co-workers [29,46] have proposed a generalized elastic-plastic slender beam theory for the analysis of the detached part of the film in a peel test. They have taken account of elastic unloading and reverse plastic bending of the film and given a closed form solution for the maximum curvature (root curvature) and hence the plastic dissipation, attained by an elastic-perfectly plastic film. Kinloch et al. [47] have studied the peeling of bilinear work hardening materials and found a good agreement with experiment. Since copper thin films are well approximated by bilinear work hardening constitutive equation, we followed the Kinloch et al. [47] approach to estimate the plastic dissipation.

The stress–strain curve at a point on the film cross-section and the moment-curvature diagram are shown in Figures 16.7 and 16.8, respectively. When the loading and unloading of the peeling film both involve plastic deformation, the plastic dissipation energy is correspond to the total energy loss in the loading and unloading cycle, the area [OABC] in Figure 16.8. The plastic dissipation W_p is,

$$W_p = \frac{\Delta \mathrm{OABC}}{b}. \tag{16.8}$$

Thus,

$$W_p = G^e_{\max} f_1(k_0), \tag{16.9}$$

$$W_{out} = G^e_{\max} \frac{1-\cos\theta}{1-\cos(\theta-\theta_0)} f_2(k_0), \tag{16.10}$$

where,

$$f_1(k_0) = \frac{4}{3}\alpha(1-\alpha)^2 k_0^2 + 2(1-\alpha)^2(1-2\alpha)k_0 + \frac{2(1-\alpha)}{3(1-2\alpha)k_0}\left[1+4(1-\alpha)^3\right] - (1-\alpha)\left[1+4(1-\alpha)^2\right], \tag{16.11}$$

$$f_2(k_0) = \frac{1}{3}\alpha\left[1+4(1-\alpha)^2\right] + 2(1-\alpha)^2(1-2\alpha)k_0 + \frac{8}{3}\frac{(1-\alpha)^4}{(1-2\alpha)k_0} - 4(1-\alpha)^3. \tag{16.12}$$

$G^e_{\max}$ is the maximum stored elastic energy as defined by,

$$G^e_{\max} = \frac{1}{2} E \varepsilon_y^2 h, \tag{16.13}$$

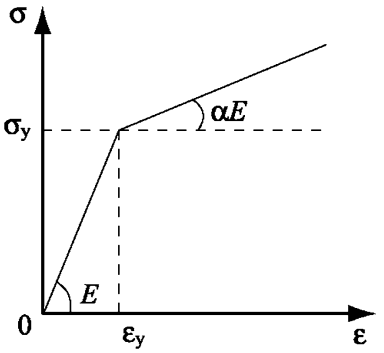


FIGURE 16.7. The stress–strain curve for bilinear, work-hardening material.

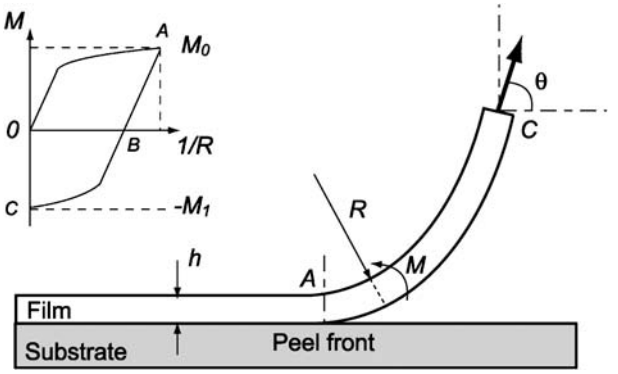


FIGURE 16.8. Deformation of the peeling film and the moment-curvature diagram undergone by peeling film.

and α is the strain hardening coefficient and ε_y is the yield strain. The term k_0 is given by,

$$k_0 = \frac{R_1}{R_0}, \tag{16.14}$$

where R_0 is the actual radius of curvature at the peel front and R_1 is the radius of curvature at the onset of plastic yielding and is given by,

$$R_1 = \frac{h}{2\varepsilon_y}. \tag{16.15}$$

It is noted that the actual radius of curvature at the peel front is difficult to determine, but it is an important parameter. The attached part of the film has been modeled as an elastic beam on an elastic foundation of thickness $h/2$. Applying a beam theory, the relationship of the root angle and k_0 has been obtained as,

$$\theta_0 = \frac{1}{3}(4\varepsilon_y)k_0. \tag{16.16}$$

From Equations (16.10) and (16.16), the root angle and k_0 can be obtained. Then, the plastic dissipation can be calculated from Equation (16.8). The detail of derivation of those equations can be seen in Kinloch et al. [47].

16.3. INTERFACIAL ADHESION STRENGTH OF COPPER THIN FILM

16.3.1. Preparation of Specimen

The specimen used in this paper is composed of copper, chromium, polyimide layers deposited on silicon substrate which is fabricated by the standard method of CSP packages. The cross-sectional view of the specimen is shown in Figure 16.9. The thickness of silicon wafer is about 1 mm, polyimide layer which is coated on Si substrate by spin-coating method is 11 μm and chromium layer which is spattered on polyimide layer is 0.2 μm. The copper layers those are plated on chromium layer are changed as 5, 10 and 20 μm to investigate the effect of copper film thickness on the interfacial strength. The copper spattering layers at the edge of specimen are installed as the scarified layer for the crack initiation. On the multi-stage peel test, the specimen is attached on an aluminum bar (which cross-section is 10 mm × 10 mm) with an epoxy adhesive to prevent the bending of silicon wafer during the peel test.

During the manufacturing process, the residual stresses would be induced in the specimen. Therefore, before peel tests, the residual stresses in copper films were measured by X-ray analysis. The measured residual stresses are shown in Figure 16.10. From this figure, the residual stresses in copper films are less than 3 MPa. The stress relaxation would be occurred in polyimide layer and reduce the residual stresses in copper film. When the residual stresses are assumed to be constant over the film thickness, the dissipation energy needed to release the residual stress can be calculated from Equation (16.4) as shown in Figure 16.10. The dissipation energy is so small that one can neglect the effect of residual stresses in this case.

The stress–strain curve of copper film is necessary to estimate the plastic dissipation as described in previous section. The tensile tests of copper film which thickness is 15 μm were carried out and obtained Young's modulus and 0.2% yield stress. The properties of each material are shown in Table 16.1. Poisson's ratio of copper film and other properties are referred from those of bulk materials.

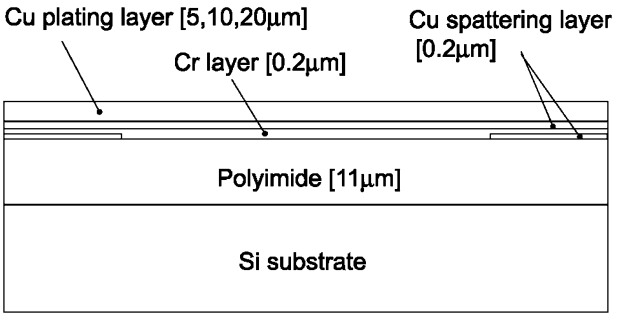


FIGURE 16.9. The cross-sectional view of the specimen.

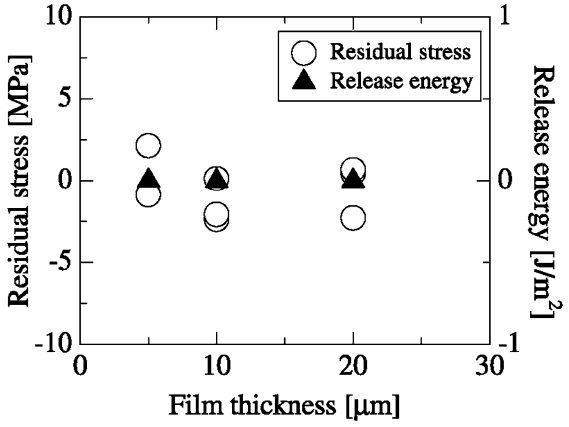


FIGURE 16.10. The residual stresses and the dissipation energy needed to release the residual stresses.

TABLE 16.1.
Material properties of the specimen.

Material	Young's modulus (GPa)	Poisson's ratio	Yield stress $\sigma_{0.2}$ (MPa)
Copper	30	0.3	440
Chromium	115	0.3	—
Polyimide	3	0.45	—

16.3.2. Measurement of Adhesion Strength by the MPT

The characteristic load–displacement curve obtained by the multi-stage peel test is shown in Figure 16.11. After the onset of the peeling, the peel force becomes constant and the delamination is in steady-state condition. Even in steady-state condition, the peel force slightly scatters since the adhesion strength is not locally constant. Hence, we averaged the peel force on some time period for the calculation of the decohesion energy. Figure 16.12 shows the top view of the specimen after the peel test. From the observation of the peeled specimen, the delamination mostly occurred at copper/chromium interface. Therefore, we discuss the interfacial strength of the copper/chromium interface in this paper.

From the averaged peel force and the peel angle, the work done by peel force can be calculated by Equation (16.5). Figure 16.13 shows the work done by peel force, i.e., the energy from outside into the peel front, in no dead weight case. As the film is thinner, the work needed to delaminate becomes larger. From the energy balance of Equation (16.6) and considering the dissipations due to the plastic deformation and the residual stresses, the decohesion energy can be calculated as shown in Figure 16.13. Comparing to the results of the work done by peel force, the effect of film thickness becomes small. When the film thickness is over 10 μm, the decohesion energy becomes constant and is approximately 20 J/m^2. This value is considered to be the interfacial strength between copper and chromium layers. By changing the dead weight attached on the specimen, the dependence of the decohesion energy on the peel angle was obtained in 10, 20 μm cases as shown in Figure 16.14. The peel angles varied from about 45 degrees to 90 degrees and the decohesion energy increase with the peel angles.

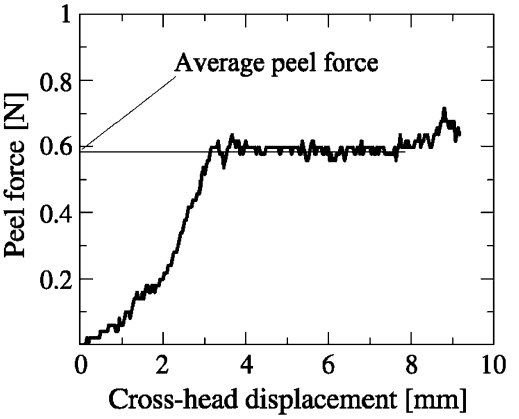


FIGURE 16.11. The characteristic peel force and displacement curve.

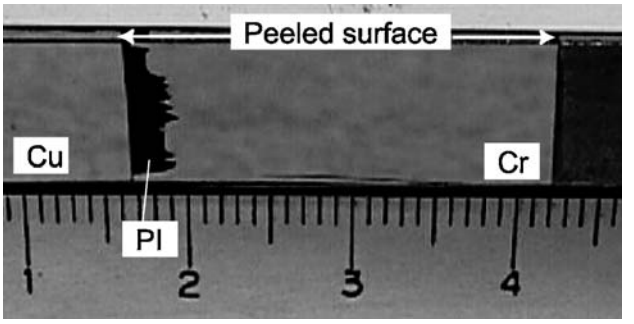


FIGURE 16.12. The top view of the specimen after the peel test. The thickness of Cu layer is 20 μm.

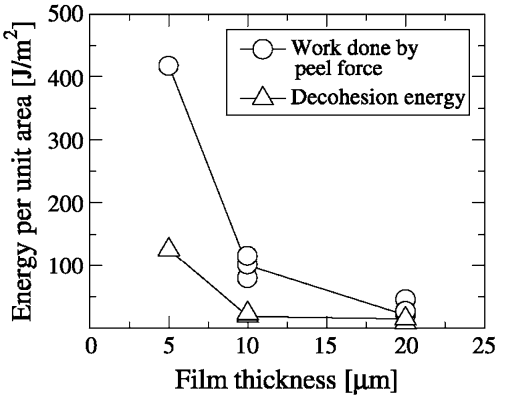


FIGURE 16.13. The work done by peel force and the decohesion energy of Cu/Cr layer.

16.3.3. Discussions

In previous section, we obtained the decohesion energy of copper and chromium interface by the multi-stage peel test. However, it should be considered whether this obtained

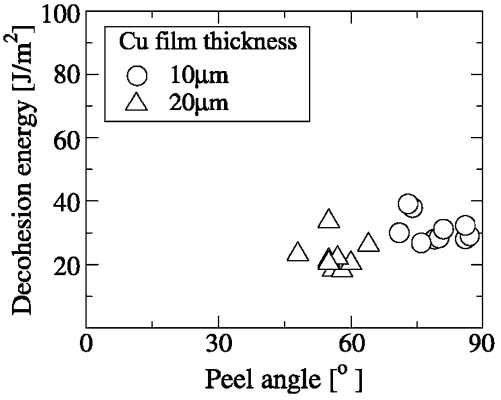


FIGURE 16.14. The peel angle effects on the decohesion energy.

value is reasonable or not. To confirm it, we have carried out finite element analysis and compare experimental results with numerical results. During the steady-state condition, the peel problem can be divided to the detached part and attached part as shown in Figure 16.15. Our interest is the energy flow around the peel front. Then, we can only consider the attached part with appropriate boundary conditions. From the equilibrium condition of force and moment between detached and attached part, the boundary conditions can be obtained as:

$$\left.\begin{aligned} N &= F\cos\theta - F_h \\ Q &= F\sin\theta \\ M &= F\sin\theta \cdot d \end{aligned}\right\}, \tag{16.17}$$

where, F is the averaged peel force, F_h is the dead weight, θ is the peel angle and d is the reference length from the peel front at which the peel angle is measured. Then, the numerical model of the multi-stage peel test is reduced to the equivalent interface crack problem as shown in Figure 16.16. In this case, silicon substrate is assumed to be rigid for simplicity.

From experiments, the averaged peel forces and peel angles during the steady-state condition were measured. Those values are used as the boundary conditions for numerical simulations. The interfacial fracture energy, which is correspond to the decohesion energy, was evaluated by J-integral in numerical simulations. Figure 16.17 shows the comparison between the results of the MPT method and J-integral calculation. In elastic analyses, the results of the J-integral calculation agree well with the MPT results, but the obtained results depend on film thickness. On the other hand, in elastic-plastic analyses, the J-integral values become constant, about 20 J/m^2, and are independent of film thickness. J-integral value for elastic-plastic material still has the meaning of energy release rate. Hence, this value is considered to be the interfacial strength of Cu/Cr interface. The difference between the MPT results and J-integral values stems from the formation of plastic zone around the peel front. Not only bending of the film but also stress concentration at the peel front induced the plastic deformation at the peel front. The plastic zone ahead of peel front is shown in Figure 16.18. In 20 μm case, the plastic zone is formed at the vicinity of the peel front and can be neglected like the small-scale yielding condition. On the contrary, in 5 μm case, the copper film is largely bended and the plastic zone is formed around the peel front

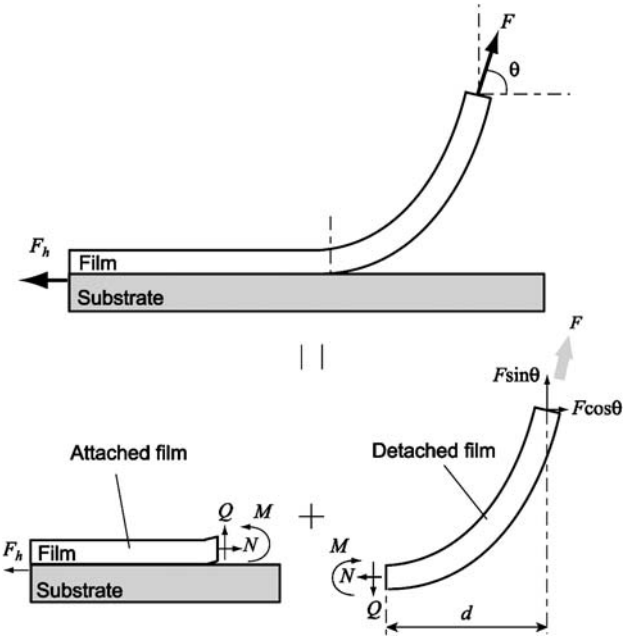


FIGURE 16.15. The reduction of solved problem into the equivalent interface crack problem.

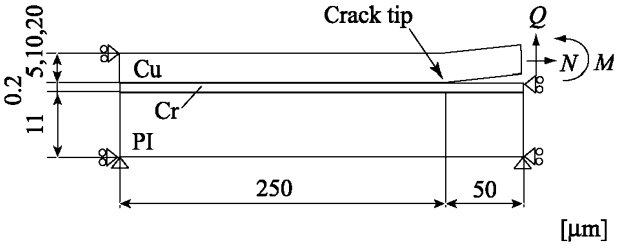


FIGURE 16.16. Numerical model for multi-stage peel test.

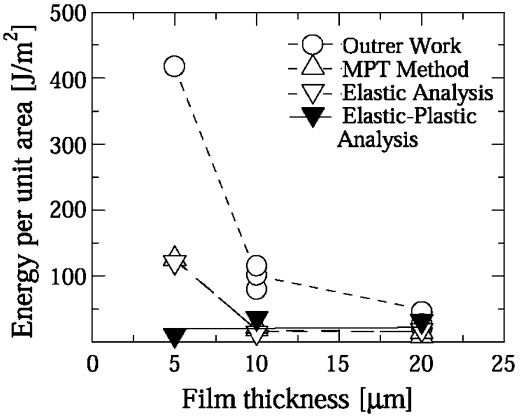
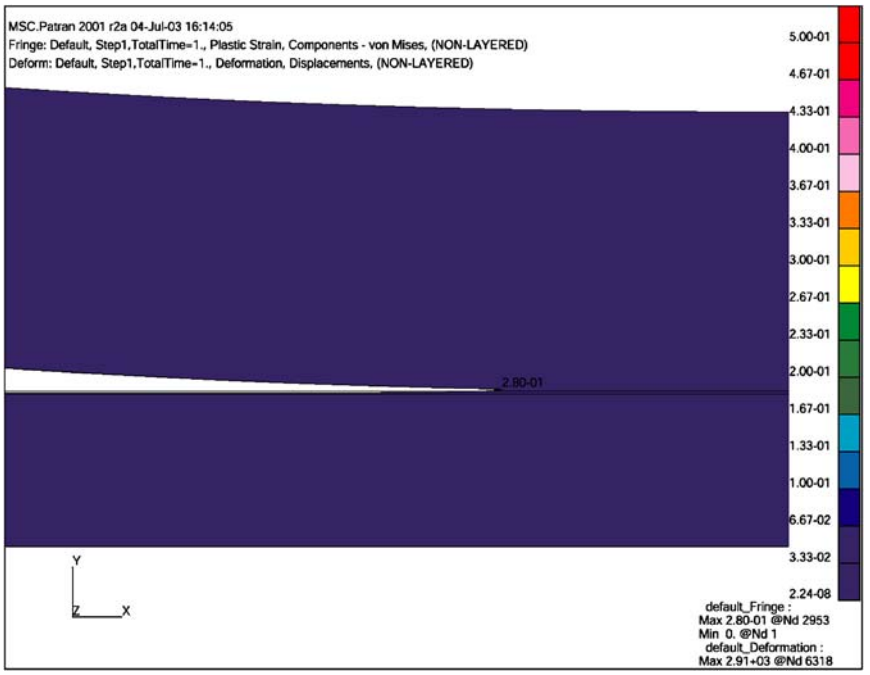
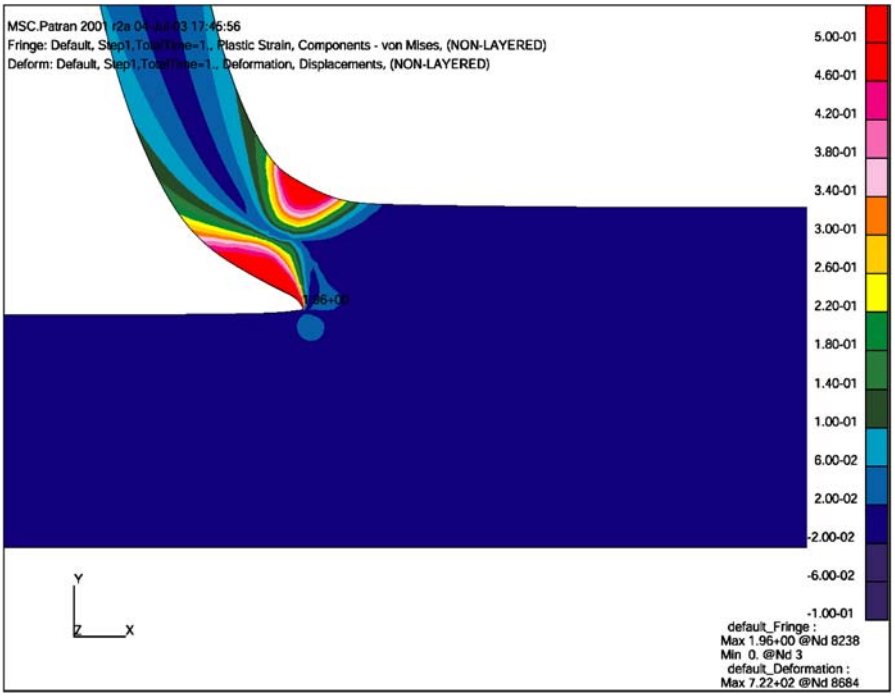


FIGURE 16.17. Comparison with the results of MPT and numerical simulation.



(a)



(b)

FIGURE 16.18. The plastic strain distribution around the peel front. (a) Copper film thickness is 20 μm. (b) Copper film thickness is 5 μm.

widely. This case is corresponding to the large-scale yielding condition. The energy evaluation in the MPT method does not cover the large-scale yielding condition. Therefore, the MPT evaluation includes the energy dissipation due to the plastic deformation.

The J-integral evaluation in 5 μm case is slightly smaller than the other cases. This means that the decohesion energy between copper and chromium is smaller than the other cases. To consider this difference, the surface observation of peeled specimen was carried out by Atomic Force Microscope (AFM). Figure 16.19 shows the results of surface observation after the multi-stage peel test. The chromium surface is very flat and the roughness is less than a few nano-meter in 20 μm case. On the other hand, in 5 μm case, small hills those heights were about 40 nm were observed. These results imply that the delamination of Cr/PI and Cu/Cr occurred simultaneously during the peel test. Figure 16.20 shows the mean stress distribution on Cr/PI interface and high mean stress concentration occurred ahead of the peel front. This mean stress may expand micro or nano voids on Cr/PI interface and induce the delamination on Cr/PI interface. The decohesion energy measured by the MPT method for 5 μm case includes the decohesion energy both Cu/Cr and Cr/PI interfaces. The delamination occurred between Cr and PI layers may influence the stress distribution and constrain the plastic deformation ahead of the peel front. The finite element analyses did not consider the multiple delamination and it overestimated the plastic dissipation than the realistic case. Hence, the J-integral evaluation in 5 μm case is slightly smaller than the other cases.

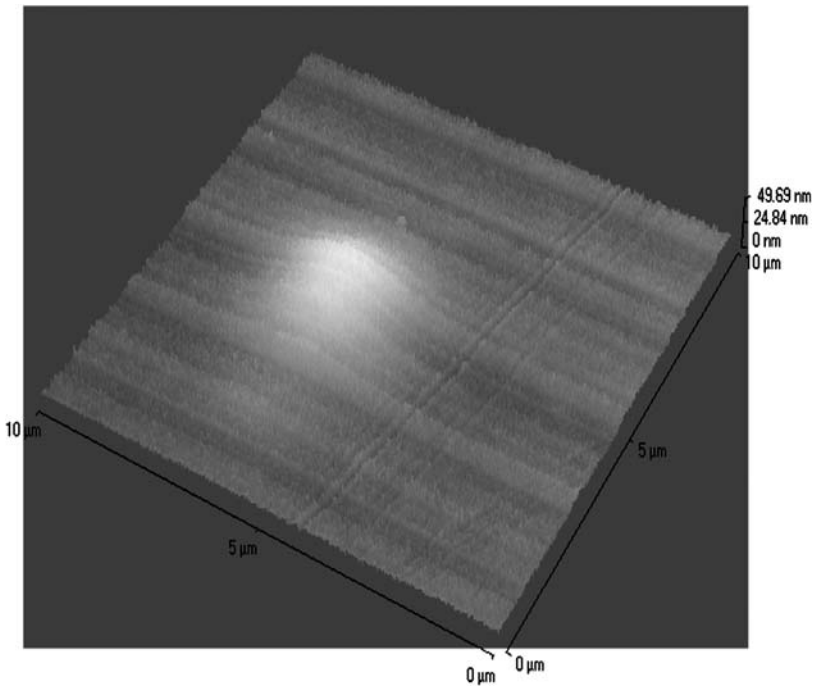
The MPT evaluation gives us only the energy flow into the peel front and does not eliminate the energy dissipation around the peel front, such as plastic dissipation. Therefore, the MPT method can be applied to the peeling under small-scale yielding condition. Nowadays, the film thickness becomes thinner and thinner going down to nano or sub-nano thickness in electronic devices. In those situations, the multiple delamination would be one of critical issues. The development of precious evaluation methods for multi-layer systems are needed and left in future works.

16.4. UV-IRRADIATION EFFECT ON CERAMIC/POLYMER INTERFACIAL STRENGTH

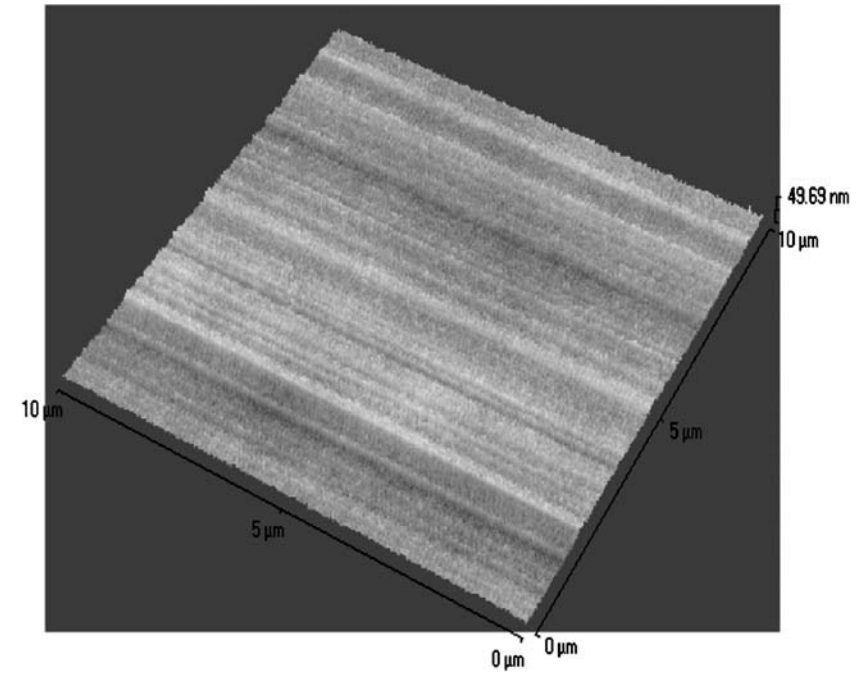
It is well known that the mechanical properties of polymers, such as tensile strength or rupture strain, are degraded by the irradiation of ultraviolet (UV) rays. When the polymer-based conductive films are used in the open air, it is necessary to consider the degradation of the mechanical properties in design of products. The bonding mechanism between ceramic and polymer is mainly intermolecular force. When UV rays irradiate polymeric materials, the principal chains are cut and oxidized. Those reactions may affect the bonding strength between ceramics and polymers. That is the motivation of this research and we investigated the effects of UV irradiation on the interfacial adhesion strength between ITO (indium tin oxide) coating layer and PET (poly(ethylene terephthalate)) substrate by multi-stage peel test.

16.4.1. Preparation of PET/ITO Specimen

Two types of specimens were prepared. One type is composed of ITO and PET and the other is composed of ITO, TiO_2 and PET. All specimens were fabricated by spattering ITO or TiO_2 on PET substrate. The thickness of each layer is summarized in Table 16.2.



(a)



(b)

FIGURE 16.19. The Cr surface configuration after the peel test. (a) The thickness of Cu layer is 5 µm. (b) The thickness of Cu layer is 20 μm.

To investigate the effect of UV irradiation, half of specimens were kept in a fade meter for 120 hours. The UV ray is exposed on ITO face and the intensity of UV ray is about 30 W/m^2. The cross-sectional view of the specimen is shown in Figure 16.21. The specimen was attached on aluminum bar with an epoxy adhesive, since the specimen was too thin to conduct the peel test directly. The stress–strain curves of PET film needed to estimate the plastic dissipation are shown in Figure 16.22. The stress–strain relation depends on the direction of loading. Young's modulus and yield stresses are summarized in Table 16.3. After UV irradiation, the rupture strain gradually decreased. The relation between rapture strain and irradiation period is represented in Figure 16.23. The rupture strain becomes smaller and smaller, as the irradiation time become long. After 120 hour irradiation, PET film is ruptured almost within elastic region.

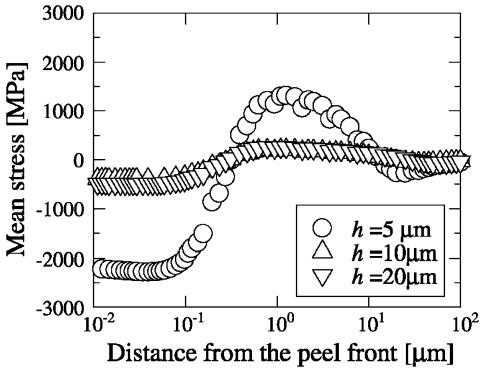


FIGURE 16.20. The mean stress distribution on chromium and polyimide interface.

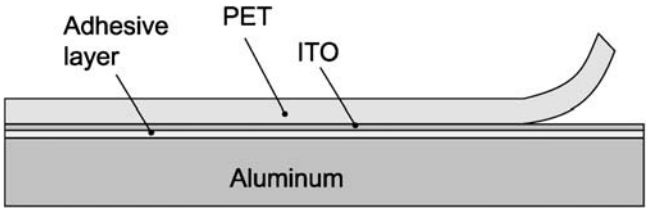


FIGURE 16.21. The cross-sectional view of test specimen.

TABLE 16.2.
The condition of test specimen.

Name	PI	PIUV	PIT	PITUV
Substrate	PET 100 μm	PET 100 μm	PET 100 μm	PET 100 μm
Coating layer	ITO 108 nm	ITO 108 nm	ITO/TiO_2 100/20 nm	ITO/TiO_2 100/20 nm
UV	—	120 hours	—	120 hours

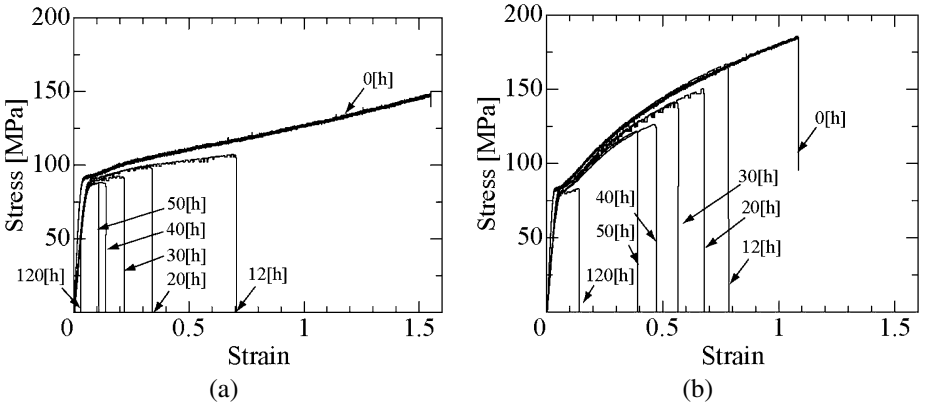


FIGURE 16.22. The stress–strain relation of PET film. (a) Parallel to the rolling direction. (b) Vertical to the rolling direction.

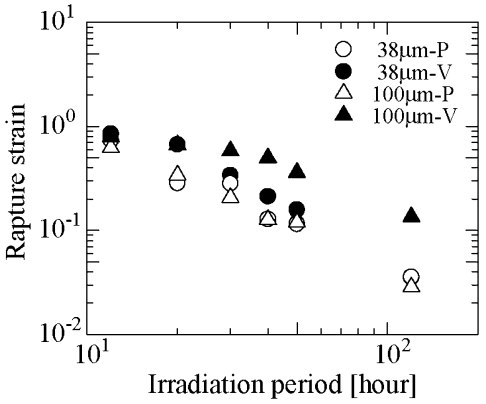


FIGURE 16.23. The irradiation effects on rupture strain of PET film.

TABLE 16.3.
Mechanical properties of PET film.

Name	LT	TL	Name	LT	TL
Young's modulus	3.8 GPa	4.1 GPa	Strain hardening coefficient, α	0.105	0.194
Poisson's ratio	0.41	0.49	Yield strain, ε_y	0.028	0.02

16.4.2. Measurement of Interfacial Strength by MPT

The load history during the peel test for PET/ITO specimen is shown in Figure 16.24. and the peel angle at the peel front during the peel test is shown in Figure 16.25. When the dead weight is about 0.83 N, the peel angle is about 55°. Without the dead weight, the peel angle is about 71°. From the load history (Figure 16.24) and the peel angle at that time, the

work done by peel force can be calculated from Equation (16.5). Also, the plastic dissipation in PET film can be estimated by Equation (16.8). When neglecting the residual stress in PET film because of the relaxation of polymeric material, the decohesion energy, then, can be evaluated from Equation (16.7). The decohesion energy by MPT method is shown in Figure 16.26 and 16.27. Figure 16.26 represents the decohesion energy of PET/ITO interface. For virgin films, the interfacial strength is about 20 J/m^2 and slightly depends on the peel angle. On the other hand, after UV irradiation, the interfacial strength is drastically decreased about 1 to 10 J/m^2 and clearly depends on the peel angle. Figure 16.27 represents the decohesion energy for PET/TiO_2/ITO specimen. The delamination occurred between PET and TiO_2 interface and the decohesion energy correspond to the interfacial strength between PET and TiO_2. In this case, for virgin films, the decohesion energy is smaller than that of PET/ITO interface. However, the degradation of interfacial strength due to UV irradiation is smaller than that of PET/ITO interface. It is considered that TiO_2 layer works as a filter and it makes the intensity of UV ray transmitted to PET layer smaller. Therefore, inserting TiO_2 layer between ITO and PET is useful to prevent the degradation of interfacial strength.

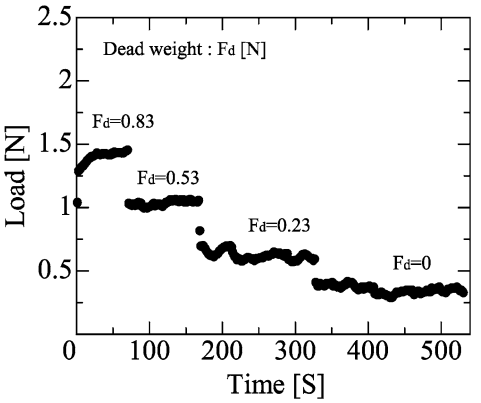


FIGURE 16.24. Load history of multi-stage peel test (PET/ITO_{LT}).



(a) (b)

FIGURE 16.25. Measurement of peel angle: (a) dead weight is 0.83 N, (b) no dead weight.

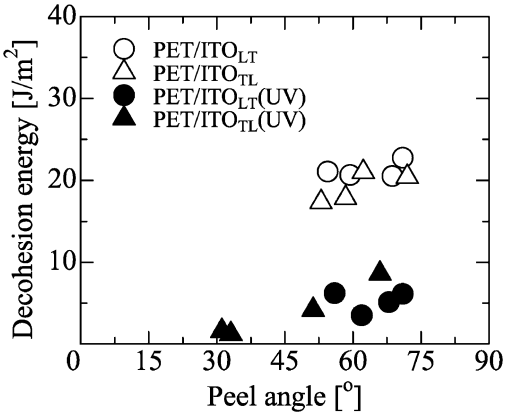


FIGURE 16.26. Decohesion energy of PET/ITO interface and effect of UV irradiation.

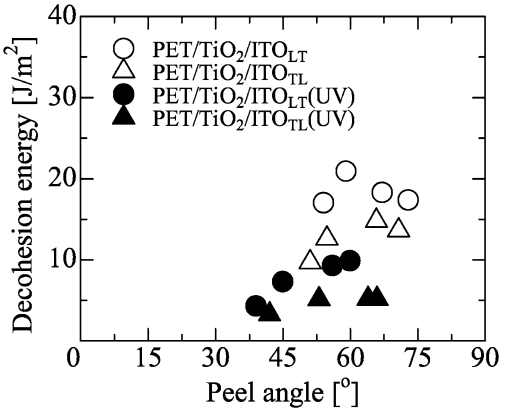


FIGURE 16.27. Decohesion energy of PET/TiO_2 interface and effect of UV irradiation.

16.4.3. Surface Crack Formation on ITO Layer under Tensile Loading

Cracks easily formed on ITO layer under tensile loading since ITO is a brittle ceramic. Once the surface crack is formed, the functionality of the complex film will be lost. Our interest is how the interfacial strength affects the crack formation on ITO layer. Therefore, in this sub-section, we carried out the tensile tests for PET/ITO complex film and the *in situ* observation of the ITO surface.

The small tensile testing machine was used under the confocal laser scanning microscope. The rectangular specimen was prepared which width is 1 mm and length is 50 mm. The stress–strain relation of the complex film is shown in Figure 16.28. It is noted that the strain is measured from the cross head displacement divided by the initial length between cramps. The stress–strain relation is almost same with that of PET film and the mechanical properties of these complex films are decided by that of PET film. However, the rupture strain is larger than that of PET film. This is because ITO layer make the intensity of UV ray reached to PET film weaker.

The *in situ* observation of surface crack formation is shown in Figure 16.29. When the strain reached 2.8%, "vertical cracks" can be observed on ITO layer. The distance between

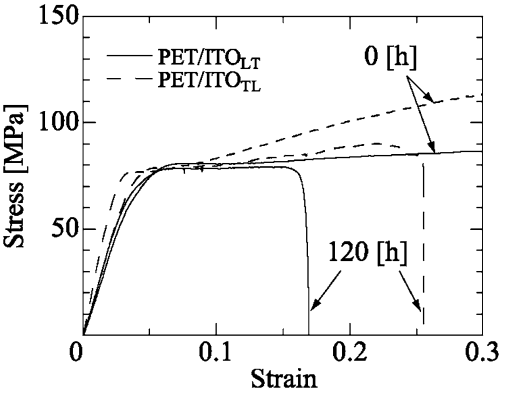


FIGURE 16.28. The stress–strain relation of PET/ITO specimen.

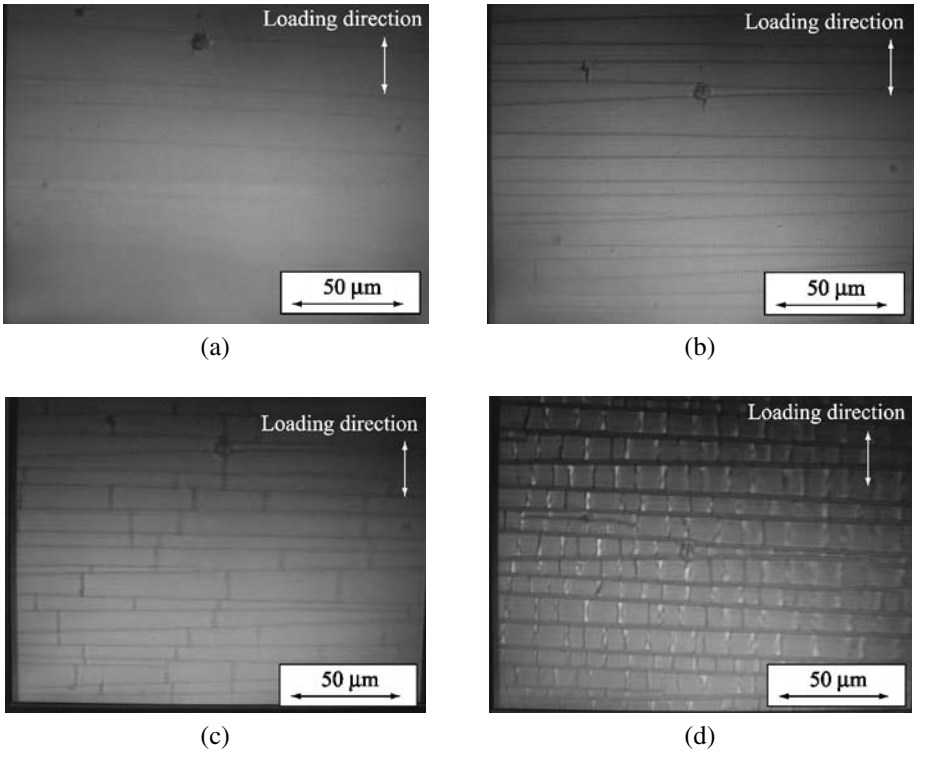
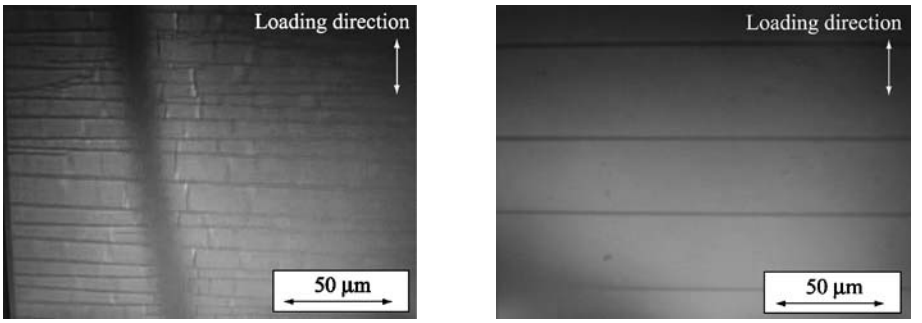


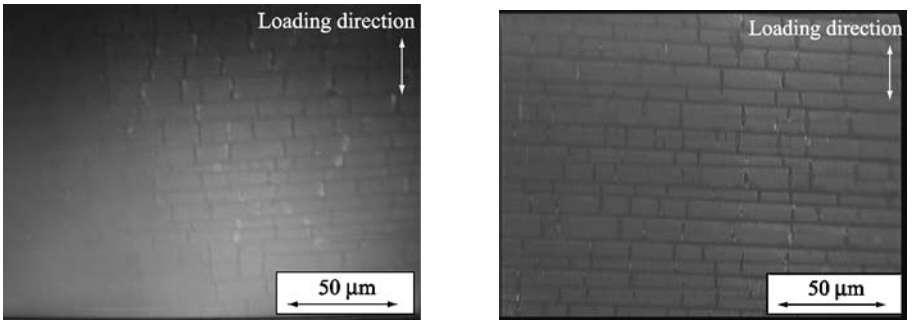
FIGURE 16.29. *In situ* observation of PET/ITO$_{LT}$ surface during tensile test. (a) $\varepsilon = 2.8\%$, (b) $\varepsilon = 6.4\%$, (c) $\varepsilon = 12\%$, (d) $\varepsilon = 30\%$.

cracks is almost constant. The density of cracks increased with applied strain as shown in Figure 16.29(b). The driving force of crack formation is shear force between ITO and PET layers [59,60]. When the interfacial strength is large, the interfacial shear rigidity is stiff and the same strain within PET film will be induced in ITO layer. That brings about high tensile stress in ITO layer and cracks are formed even in small strain range. After the strain reached about 12%, "parallel cracks" to the loading direction were observed. Due



(a) (b)

FIGURE 16.30. The effect of UV irradiation on the fracture pattern of PET/ITO_{TL} surface: $\varepsilon = 10\%$. (a) Virgin specimen. (b) UV-irradiated specimen (120 hours).



(a) (b)

FIGURE 16.31. The effect of UV irradiation on the fracture pattern of PET/TiO_2/ITO_{TL} surface: $\varepsilon = 10\%$. (a) Virgin specimen. (b) UV-irradiated specimen (120 hours).

to the Poisson's effect, the compressed stress vertical to the loading direction induced and buckling cracks occurred in ITO layer. After the initiation of buckling crack, the density of "vertical cracks" is constant and only the density of "parallel cracks" increased.

The comparison of UV irradiation effects on crack formation is shown in Figure 16.30 and 16.31. Figure 16.30 represents of PET/ITO_{TL} specimen at the tensile strain is about 10%. After UV irradiation for 120 hours, only "vertical cracks" can be observed. This is because the interfacial strength between PET and ITO decreased by UV irradiation as shown in Figure 16.26 and then, the interfacial shear rigidity also decreased. Even when PET layer deformed largely, the strain induced in ITO layer can not be enough to form "parallel cracks." On the other hand, the crack formation of PET/TiO_2/ITO specimen is almost same as that of virgin specimen as shown in Figure 16.31. In this case, the degradation of the interfacial strength is smaller than that of PET/ITO case. From these results, the interfacial strength closely related the crack formation behavior on ITO layer.

16.5. CONCLUDING REMARKS

In this chapter, the multi-stage peel test method is introduced and is applied to the measurement of the adhesion strength for copper thin film and conductive ceramics film.

During the steady-state peeling, the interfacial strength can be evaluated by the energy balance of peeling system. When the film thickness is thick enough to neglect the plastic zone size at the peeling tip, i.e., under small-scale yielding condition, MPT method can be valid to evaluate the interfacial strength precisely. However, when the film thickness is thinner and the plastic zone size can not be neglected, i.e., under large-scale yielding condition, the MPT method can not be applied directly and needs the help of some numerical calculations. More, in multi-layer system, the delamination on other layers would be occurred during peel tests. In those cases, the delamination of other layers will affect the stress condition at the peeling tip and the conventional evaluation techniques can not be applied. Those multiple delamination would be important issues in nano- and subnano-thickness film structures and the solution of those problems has been left in future works.

We also applied MPT method to the measuring of the interfacial strength between conductive ceramics thin film (ITO) and polymer substrate (PET). The stress–strain relations of PET film are significantly degraded by UV irradiation. The interfacial strength between ITO and PET was also degraded. When TiO_2 layer inserted between ITO and PET layers, the degradation of interfacial strength became smaller than that of PET/ITO interface. TiO_2 layer works as a filter and protect the interfacial bonding between ITO and PET from UV attack. The interfacial strength is closely related to the crack formation of ITO layer under tensile loadings. However, it has not been clear how much molecules on the interface or under the interface are damaged by UV irradiation and how to connect the number of damaged molecules and the degradation of interfacial strength quantitatively. The bottom-up approaches, such as molecular dynamics, are necessary to solve these problems. Moreover, when supplying the power voltage on ITO layer, the interfacial strength and crack formation would be different from obtained results. Therefore, it is necessary to conduct those experiments under more realistic situations.

ACKNOWLEDGMENT

The authors would like to acknowledge Dr. Amagai of Texas Instruments Japan and Dr. Yanaka of Toppan Printing Co. Ltd., for their preparation of test specimen.

REFERENCES

1. S.F. Al-sarawi, D. Abbott, and P.D. Franzon, A review of 3-D packaging technology, IEEE Transactions on Components, Packaging and Manufacturing Technology, Part B, 21, pp. 2–14 (1998).
2. E. Harlev, T. Gulakhmedova, I. Rubinovich, and G. Aizenshtein, New method for the preparation of conductive polyaniline solutions: Application to liquid crystal devices, Advanced Materials, 8, pp. 994–997 (1996).
3. G. Gu, P.E. Buroows, S. Venkatesh, S.R. Forrest, and M.E. Thompson, Vacuum-deposited, nonpolymeric f lexible organic light-emitting devices, Optics Letters, 22, pp. 172–174 (1997).
4. Z. Chen, B. Cotterell, W. Wang, E. Guenther, and S.J. Chua, A mechanical assessment of flexible optoelectronic devices, Thin Solid Films, 394, pp. 201–205 (2001).
5. M. Notomi, T. Gotoh, K. Kishimoto, and T. Shibuya, Evaluation of ultra-violet degradation on PP and PC by layering-films exposure test, Transactions of the Japan Society of Mechanical Engineers, Part A, 63, pp. 437–444 (1997).
6. J.W. Hutchinson and Z. Suo, Mixed mode cracking in layered materials, Advances in Applied Mechanics, 29, pp. 63–191 (1992).
7. A.G. Evans and J.W. Hutchinson, The thermomechanical integrity of thin films and multilayers, Acta Metallurgica Materialia, 43, pp. 2507–2530 (1995).

8. A.G. Evans, J.W. Hutchinson, and Y. Wei, Interface adhesion: effects of plasticity and segragation, Acta Metallurgica Materialia, 47, pp. 4093–4113 (1999).
9. S.W. Russel, S.A. Rafalski, R.L. Spreitzer, J. Li, M. Moinpour, F. Moghadam, and T.L. Alford, Enhanced adhesion of copper to dielectrics via titanium and chromium additions and sacrificial reactions, Thin Solid Films, 262, pp. 154–167 (1995).
10. D.B. Marshall and A.G. Evans, Measurement of adherence of residually stressed thin films by indentation. I. Mechanics of interfaced delamination, Journal of Applied Physics, 56, pp. 2632–2638 (1984).
11. C. Rossington, A.G. Evans, D.B. Marshall, and B.T. Khuri-Yakub, Measurement of adherence of residually stressed thin films by indentaion. II. Experiments with ZnO/Si, Journal of Applied Physics, 56, pp. 2639–2644 (1984).
12. A.G. Evans and J.W. Hutchinson, On the mechanics of delamination and spalling in compressed films, International Journal of Solids and Structures, 20, pp. 455–466 (1984).
13. L.G. Rosenfeld, J.E. Ritter, T.J. Lardner, and M.R. Lin, Use of the microindentation technique for determining interfacial fracture energy, Journal of Applied Physics, 67, pp. 3291–3298 (1990).
14. D.F. Bahr, J.W. Hoehn, N.R. Moody, and W.W. Gerberich, Adhesion and acoustic emission analysis of failures in nitride films with a metal interlayer, Acta Materialia, 45, pp. 5163–5175 (1997).
15. M.D. Drory and J.W. Hutchinson, Measurement of the adhesion of a brittle film on a ductile substrate by indentation, Proceedings of the Royal Society of London Series A—Mathematical Physical and Engineering Sciences, 452, pp. 2319–2341 (1996).
16. M.D. Kriese, N.R. Moody, and W.W. Gerberich, Effects of annealing and interlayers on the adhesion energy of copper thin films to SiO2/Si substrate, Acta materialia, 46, pp. 6623–6630 (1998).
17. M.D. Kriese, D.A. Boismier, N.R. Moody, and W.W. Gerberich, Nonomechanical fracture-testing of thin films, Engineering Fracture Mechanics, 61, pp. 1–20 (1998).
18. W.W. Gerberich, D.E. Kramer, N.I. Tymiak, A.A. Volinsky, D.F. Bahr, and M.D. Kriese, Nanoindentation-induced defect-interface interactions: phenomena. Methods and limitations, Acta Materialia, 47, pp. 4115–4123 (1999).
19. H.M. Jensen, The blister test for interface toughness measurement, Engineering Fracture Mechanics, 40, pp. 475–486 (1991).
20. H.M. Jensen and M.D. Thouless, Effects of residual stresses in the blister test, International Journal of Solids and Structures, 30, pp. 779–795 (1993).
21. A. Bagchi, G.E. Lucas, Z. Suo, and A.G. Evans, A new procedure for measuring the decohesion energy for thin ductile films on substrates, Journal of Material Research, 9, pp. 1734–1741 (1994).
22. A. Bagchi and A.G. Evans, Measurements of the debond energy for thin metallization lines on dielectrics, Thin Solid Films, 286, pp. 203–212 (1996).
23. A.V. Zhuk, A.G. Evans, J.W. Hutchinson, and G.M. Whitesides, The adhesion energy between polymer thin films and self-assembled monolayers, Journal of Material Research, 13, pp. 3555–3565 (1998).
24. T. Kitamura, T. Shibutani, and T. Ueno, Development of evaluation method for interface strength between thin films and its application on delamination of Cu/TaN in an advanced LSI, Transactions of the Japan Society of Mechanical Engineers, 66, pp. 1568–1573 (2000).
25. T. Kitamura, H. Hirakata, and T. Itsuji, Delamination strength of Cu thin film characterized by nanoscale stress field near interface edge, Transactions of the Japan Society of Mechanical Engineers, Part A, 68, pp. 119–125 (2002).
26. T. Kitamura, H. Hirakata, and Y. Yamamoto, Interface strength of tungsten micro-component on silicon substrate by means of AFM, Transactions of the Japan Society of Mechanical Engineers, Part A, 69, pp. 1216–1221 (2003).
27. K. Nakasa, M. Kato, D. Zhang, and K. Tasaka, Evaluation of delamination strength of thermally sprayed coating by edge-indentation method, Journal of the Society of Material Science, Japan, 47, pp. 413–419 (1998).
28. D. Zhang, M. Kato, and K. Nakasa, Fracture mechanics analysis of edge-indentation method for evaluation of delamination strength of coating, Journal of the Society of Material Science, Japan, 49, pp. 572–578 (2000).
29. K.S. Kim and N. Aravas, Elasto-plastic analysis of the peel test, International Journal of Solids and Structures, 24, pp. 417–435 (1988).
30. G.J. Spies, The peeling test on redux-bonded joints, Journal of Aircraft Engineering, 25, pp. 64–70 (1953).
31. J.J. Bickerman, Theory of peeling through a hookean solid, Journal of Applied Physics, 28, pp. 1484–1485 (1957).
32. D.H. Kaeble, Theory and analysis of peel adhesions: mechanism and mechanics, Transaction of Society of Rheology, 3, pp. 161–180 (1959).

33. D.H. Kaeble, Theory and analysis of peel adhesions: bond stresses and distributions, Transaction of Society of Rheology, 4, pp. 45–73 (1960).
34. C. Jouwersma, On the theory of peeling, Journal of Polymer Sciences, 45, pp. 253–255 (1960).
35. S. Yurenka, Peel testing of adhesive bonded metal, Journal of Applied Polymer Science, 6, pp. 136–144 (1962).
36. J.L. Gardon, Peel adhesion. II. A theoretical analysis, Journal of Applied Polymer Sciences, 7, pp. 643–664 (1963).
37. E.B. Saubestre, L.J. Durney, J. Haidu, and E. Bastenbeck, The adhesion of electrodeposits to plastics, Plating, 52, pp. 982–1000 (1965).
38. K. Kendall, The shapes of peeling solid films, Journal of Adhesion, 5, pp. 105–117 (1973).
39. A.N. Gent and G.R. Hamed, Peel mechanics, Journal of Adhesion, 7, pp. 91–95 (1975).
40. D.W. Nicholson, Peel mechanics with large bending, International Journal of Fracture, 13, pp. 279–287 (1977).
41. M.D. Chang, K.L. Devries, and M.L. Williams, The effects of plasticity in adhesive fracture, Journal of Adhesion, 4, pp. 221–231 (1972).
42. A.N. Gent and G.R. Hamed, Peel mechanics for an elastic-plastic adherent, Journal of Applied Polymer Sciences, 21, pp. 2817–2831 (1977).
43. A.D. Crocombe and R.D. Adams, Peel analysis using the finite element method, Journal of Adhesion, 12, pp. 127–139 (1981).
44. A.D. Crocombe and R.D. Adams, An elasto-plastic investigation of the peel test, Journal of Adhesion, 13, pp. 241–267 (1982).
45. A.G. Atkins and Y.W. Mai, Residual strain energy in elastoplastic adhesive and cohesive fracture, International Journal of Fracture, 30, pp. 203–221 (1986).
46. K.-S. Kim and J. Kim, Elasto-plastic analysis of the peel test for thin film adhesion, Journal of Engineering Materials and Technology, 110, pp. 266–273 (1988).
47. A.J. Kinloch, C.C. Lau, and J.G. Williams, The peeling of flexible laminates, International Journal of Fracture, 66, pp. 45–70 (1994).
48. Y. Wei and J.W. Hutchinson, Interface strength, work of adhesion and plasticity in the peel test, International Journal of Fracture, 93, pp. 315–333 (1998).
49. Y. Wei and J.W. Hutchinson, Peel test and interfacial toughness, in W. Gerberich and W. Yang, Eds., Interfacial and Nanoscale Failure, Volume 8, Comprehensive Structural Integrity, I. Milne, R.O. Ritchie, and B. Karihaloo, Editors-in-Chief, Elsevier, Amsterdam, 2003, pp. 181–217.
50. V. Tvergaard and J.W. Hutchins, The influence of plasticity on mixed mode interface toughness, Journal of the Mechanics and Physics of Solids, 41, pp. 1119–1135 (1993).
51. F. Ma and K. Kishimoto, A continuum interface debonding model and application to matrix cracking of composite, JSME International Journal, Series A, 39, pp. 496–507 (1996).
52. C. Zhou, W. Yang, and D. Fang, Damage of short-fiber-reinforced metal matrix composite considering cooling and thermal cycling, Journal of Engineering Materials and Technology, 122, pp. 203–209 (2000).
53. M. Omiya, K. Kishimoto, and W.M. Yang, Interface debonding model and it's application to the mixed mode interface fracture toughness, International Journal of Damage Mechanics, 11, pp. 263–286 (2002).
54. I.S. Park and J. Yu, An X-ray study on the mechanical effects of the peel test in a Cu/Cr/polyimide system, Acta Materialia, 46, pp. 2947–2953 (1988).
55. Y.B. Park, I.S. Park, and J. Yu, Interfacial fracture energy measurements in the Cu/Cr/polyimide system, Materials Science and Engineering A: Structural Materials: Properties, Microstructure and Processing, 266, pp. 261–266 (1999).
56. Y.B. Park, I.S. Park, and J. Yu, Phase angle in the Cu/polyimide/alumina system, Materials Science and Engineering A: Structural Materials: Properties, Microstructure and Processing, 266, pp. 109–114 (1999).
57. H.-H. Yu, M.Y. He, and J.W. Hutchinson, Edge effects in thin films, Acta Materialia, 49, pp. 93–107 (2001).
58. W. Yang and L.B. Freund, Shear stress concentration near the edge of a thin film deposited in a substrate, Brown Technical Report, November 1984.
59. M. Yanaka, Y. Kato, Y. Tsukahara, and N. Takeda, Effects of temperature on the multiple cracking progress of sub-micron thick glass films deposited on a polymer substrate, Thin Solid Films, 355-356, pp. 337–342 (1999).
60. B.F. Chen, J. Hwang, G.P. Yu, and J.H. Huang, In situ observation of the cracking behavior of TiN coating on 304 stainless steel subjected to tensile strain, Thin Solid Films, 352, pp. 173–178 (1999).

17

The Effect of Moisture on the Adhesion and Fracture of Interfaces in Microelectronic Packaging

Timothy P. Ferguson[a] and Jianmin Qu[b]

[a]*Southern Research Institute, 757 Tom Martin Drive, Birmingham, AL 35211, USA*

[b]*G.W. Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0405, USA*

Abstract

A significant problem in the microelectronic packaging industry is the presence of moisture-induced failure mechanisms. Moisture is a multi-dimensional concern in packaging, having an adverse effect on package reliability by introducing corrosion, development of hygro-stresses, and degradation of polymers present in the package. Moisture can also accelerate delamination by deteriorating the polymer interfaces within the package. As the interfacial adhesion between the chip, underfill, and substrate decreases, the likelihood of delamination at each encapsulant interface increases. Once the package delaminates, the solder joints in the delaminated area are exposed to high stress concentrations, resulting in a reduction of overall package life.

Moisture can affect interfacial adhesion through two primary mechanisms. The first mechanism is the direct presence of moisture at the interface altering the interfacial integrity of the adhesive joint. The second mechanism is the absorbed moisture in either the adhesive and/or substrate altering the mechanical properties of those materials, which changes the response of the adhesive structure in the presence of an externally applied load. Inevitably, the effect of moisture on the adhesion and fracture of interfaces entails a multi-disciplinary study, and several aspects should be considered. From a global perspective, the primary aspects include moisture transport behavior, changes in bulk material properties from moisture absorption, effect of moisture on interfacial adhesion, and recovery from moisture upon fully drying, although several subsections within each major group occur due to the complexity of the problem.

In this chapter, a systematic and multi-disciplinary study is presented to address the fundamental science of moisture-induced degradation of interfacial adhesion. First, the

moisture transport behavior within underfill adhesives is experimentally characterized. The results are incorporated into a finite element model to depict the moisture ingress and interfacial moisture concentration after moisture preconditioning. Second, the effect of moisture on the variation of the adhesive elastic modulus is demonstrated and the physical mechanisms for the change identified. Third, the aggregate effect of moisture on the interfacial fracture toughness is determined. This includes the primary effect of moisture being physically present at the interface and the secondary effect of moisture changing the elastic modulus of the adhesive when absorbed. Both reversible and irreversible components of the interfacial moisture degradation are evaluated. Using adsorption theory in conjunction with fracture mechanics, an analytical model is developed that predicts the loss in interfacial fracture toughness as a function of moisture content. The model incorporates key parameters relevant to the problem of moisture in epoxy joints identified from the experimental portion of this research, including the interfacial hydrophobicity, active nanopore density, saturation concentration, and density of water.

17.1. INTRODUCTION

It is inevitable that an electronic device will be exposed to varying degrees of moisture. Since many microelectronic packages utilize epoxy based materials such as underfill and molding compounds, they are highly susceptible to moisture absorption. The moisture uptake can lead to undesirable changes in mechanical performance, interfacial adhesion, and reliability [46,50].

Long term reliability and life prediction of microelectronic assemblies requires a rooted understanding in the interfacial failure mechanisms and associated debonding behavior of adhesive joints within these assemblies. With the advent of flip-chip technology, the need for improved understanding of delamination in these assemblies has taken on added importance. One of the keys to the success of flip-chip technology lies in development of underfill, which is an epoxy-based encapsulant that mechanically couples the chip to the board. Underfill drastically enhances the reliability of microelectronic assembles when compared to unencapsulated devices [51], provided the structural integrity of the adhesive bond is maintained. Although delamination of the underfill in the microelectronic assembly tends to cause near immediate failure as soon as it reaches a solder joint, until recently the factors that affect the strength and durability of these interfaces have not been investigated and are the focal points of current studies in reliability research in microelectronic packaging. One of the most detrimental of these factors is moisture, which can significantly compromise the interfacial adhesion and accelerate the onset of delamination.

Another major area of concern in microelectronic packaging occurs at the interface between the copper alloy lead frame and the epoxy mold compound. Due to its relatively low cost in conjunction with its high electrical and thermal conductivity, copper alloys are widely used as a lead frame material. However, the epoxy/copper interface has poor interfacial adhesion strength and relatively high residual stress, which predisposes it to delamination. The copper surface is also highly susceptible to oxidation, which is an additional consideration when evaluating the interfacial adhesion of interfaces involving copper [9,28]. The delamination between the copper lead frame and the mold compound adversely affects the durability of these packages and is a common failure mode during the qualification process. In addition to further compromising the interfacial integrity of the interface, the delamination can also affect long term package reliability by yielding enhanced transport

of moisture along the interface resulting in corrosion. The corrosion process will be accelerated if the absorbed moisture is a carrier of ionic impurities from the surrounding external environment [62]. Consequently, the epoxy/copper interface is another significant area of concern in microelectronic packaging reliability. Several studies continue to investigate this topic to better understand the durability and failure mechanisms, including the loss of adhesion in the presence of moisture.

17.2. MOISTURE TRANSPORT BEHAVIOR

17.2.1. Background

Central to understanding the effect of moisture on interfacial adhesion is to first identify the rate at which moisture is delivered to the interface. The three primary parameters that have the greatest effect on diffusion rates are the size of the diffusing particles, temperature, and viscosity of the environment. Lighter particles have a higher velocity for the same kinetic energy as heavier particles, thus lighter particles diffuse faster than heavier particles. Similarly, an increase in temperature will produce a higher kinetic energy yielding an increase in velocity, thus particles will diffuse more rapidly at elevated temperatures. Last, diffusion is more rapid in a gas than in a solid as a result of less atomic interactions, which retard the diffusion process.

Since the vast majority of underfills are epoxy based, they are highly susceptible to moisture absorption. A standard epoxy formulation can absorb between 1 and 7 wt% moisture [48]. Additional considerations that apply specifically to moisture absorption in epoxies include the epoxy surface topology and resin polarity. Soles et al. [47] have found that water initially enters the epoxy network through the nanopores that are inherent in the epoxy surface topology. They have determined the average size of a nanopore diameter to vary from 5.0 to 6.1 Å and account for 3–7% of the total volume of the epoxy material. Since the approximate diameter of a kinetic water molecule is just 3.0 Å, moisture can easily traverse into the epoxy via the nanopores.

Although surface topology can influence moisture penetration into an epoxy, of primary importance is the resin polarity, with the high polarity of the water molecule being susceptible to specific epoxy–water interactions. Less polar resins such as non-amine resins have more enhanced moisture diffusion coefficients than amine-containing resins. Soles and Yee [48] have shown that polar sites, such as amine functional groups, provide low energy wells for the water molecules to attach. Consequently, polar hydroxyls and amines can regulate transport through the nanopores by either blocking or allowing moisture to traverse the epoxy resin depending on the orientation of the resin with respect to nanopore position. Conversely, the absence of hydroxyls and amines in a non-amine resin leads to an enhanced moisture diffusion coefficient. In addition, non-amine resins absorb very little water relative to more polar resins, such as amine resins. Soles and Yee [48] have shown that by increasing the cross-link density, the intrinsic hole volume fraction is increased, which yields an increase in the equilibrium moisture content. Steric hindrances located at cross-link junctions open the epoxy matrix to facilitate interactions of water with polar groups, thus increasing the moisture uptake. Depending on the various chemical conformations of the epoxy resin in association with the inherent nanopores present in the epoxy structure, water molecules will behave differently in various epoxy resins.

17.2.2. Diffusion Theory

Since the transfer of heat by conduction is also attributed to random molecular motions, it is clear that diffusion is analogous to heat conduction. Fick adopted Fourier's mathematical expression for heat conduction to quantify diffusion. Fick's first law states that the rate of transfer of diffusing particles per plane of unit area is proportional to the concentration gradient measured normal to the plane:

$$F_x = -D\frac{\partial C}{\partial x}, \tag{17.1}$$

where F_x is the diffusion flux in the x direction, D is the diffusion coefficient, and $\partial C/\partial x$ is the concentration gradient. The negative sign in the above expression accounts for the fact that diffusion occurs in the opposite direction of increasing concentration. In addition, the expression is only valid for an isotropic medium.

Fick's second law of diffusion describes the nonsteady state diffusion of a substance and can be derived using Equation (17.1). Utilizing Equation (17.1) and a differential volume element in Cartesian coordinates, Crank [11] has shown that the following expression can be obtained assuming a constant diffusion coefficient:

$$\frac{\partial C}{\partial t} = D\left(\frac{\partial^2 C}{\partial x^2} + \frac{\partial^2 C}{\partial y^2} + \frac{\partial^2 C}{\partial z^2}\right), \tag{17.2}$$

where C is the concentration of the diffusing substance and D is the diffusion coefficient. For one-dimensional diffusion along the x-axis, the previous relation reduces to the following form:

$$\frac{\partial C}{\partial t} = D\left(\frac{\partial^2 C}{\partial x^2}\right). \tag{17.3}$$

The solution of Equation (17.3) for the concentration of a diffusing substance in an isotropic plane sheet of finite thickness as a function of time and space is given by [11]:

$$\frac{C(x,t)}{C_1} = 1 - \frac{4}{\pi}\sum_{n=0}^{\infty}\frac{(-1)^n}{2n+1}\exp\left[\frac{-D(2n+1)^2\pi^2 t}{4\ell^2}\right]\cos\frac{(2n+1)\pi x}{2\ell}, \tag{17.4}$$

where D is the diffusion coefficient, ℓ is the half-thickness of the sheet ($-\ell < x < \ell$), C is concentration of the diffusing substance absorbed by the sample at position x and time t, and C_1 is the saturation concentration of the absorbed substance. The application of Equation (17.4) assumes that immediately after the sheet is placed in the vapor both surfaces obtain a concentration that is equivalent to the equilibrium uptake, remaining constant. In addition the equation assumes that the diffusion coefficient remains constant throughout the diffusion process and that the initial concentration of the diffusing substance in the specimen is zero. An analogous expression given on a mass basis has been shown by Crank [11] to be the following:

$$\frac{M_t}{M_\infty} = 1 - \frac{8}{\pi^2}\sum_{m=0}^{\infty}\frac{1}{(2m+1)^2}\exp\left[\frac{-D(2m+1)^2\pi^2 t}{h^2}\right], \tag{17.5}$$

where D is the diffusion coefficient, h is the total sheet thickness, M_t is total mass of the diffusing substance absorbed by the sample at time t, and M_∞ is the equilibrium mass of the absorbed substance. In the initial stages of absorption where $M_t/M_\infty < 1/2$ and assuming a constant diffusion coefficient, D, Equation (17.5) can be shown to be approximated by the following:

$$\frac{M_t}{M_\infty} = \frac{4}{h}\sqrt{\frac{Dt}{\pi}}. \tag{17.6}$$

If absorption data is plotted with M_t/M_∞ as a function of $(t/h^2)^{1/2}$ and exhibits linear behavior for $M_t/M_\infty < 1/2$, the diffusion coefficient can be determined by rearranging Equation (17.6) to the following form:

$$D = \frac{\pi}{16}\left(\frac{M_t/M_\infty}{\sqrt{t}/h}\right)^2. \tag{17.7}$$

The diffusivity, D, can now be experimentally determined using absorption data with Equation (17.7). Again, Equations (17.4), (17.5), (17.6), and (17.7) all assume that the one-dimensional absorption occurs on both sides of the plane sheet with a concentration-independent, constant diffusivity. If absorption results in a diffusion coefficient that is variable rather than constant, explicit analytical solutions are no longer available.

17.2.3. Underfill Moisture Absorption Characteristics

Being epoxy-based, underfill resins are highly susceptible to moisture ingress. It is important to ascertain the fundamental moisture absorption behavior of each underfill when evaluating reliability performance in moist environments. In the case of adhesion testing, it is essential to accurately quantify the change in adhesion for a particular interfacial moisture concentration. This provides insight into the constitutive behavior of adhesion in the presence of moisture.

Two no-flow underfills were evaluated to determine their absorption behavior to select an ideal candidate for a fundamental study in the effect of moisture on interfacial adhesion. Underfill resin A (UR-A) was developed at the Georgia Institute of Technology. Underfill resin B (UR-B) was supplied by a commercial manufacturer. It should be noted that since both underfills were formulated for no-flow assembly, neither contained any filler content. Test samples were made approximately 60 mm in diameter and 2 mm thick, hence promoting predominately one-dimensional diffusion through the thickness of the sample. The samples were baked at 115°C for at least 12 hours to remove moisture prior to being placed into a humidity chamber for moisture preconditioning. The atmosphere within the humidity chamber was maintained at a constant temperature (85 ± 1°C), humidity (85 ± 1%RH), and pressure (P_{atm}). Moisture uptake profiles for each underfill are shown in Figures 17.1 and 17.2.

It is evident from Figures 17.1 and 17.2 that UR-A had not reached saturation after 168 hours of exposure, whereas UR-B had approached a saturated state within the same timeframe. In fact, samples constructed from UR-A did not reach saturation even after 725 hours of exposure. This absorption behavior is not uncommon, with Vanlandingham et al. [54] noting that some of the epoxies evaluated in their study had not reached saturation even after 3000–4000 hours of exposure at 50°C/85%RH. Similarly, Ardebili et al. [1]

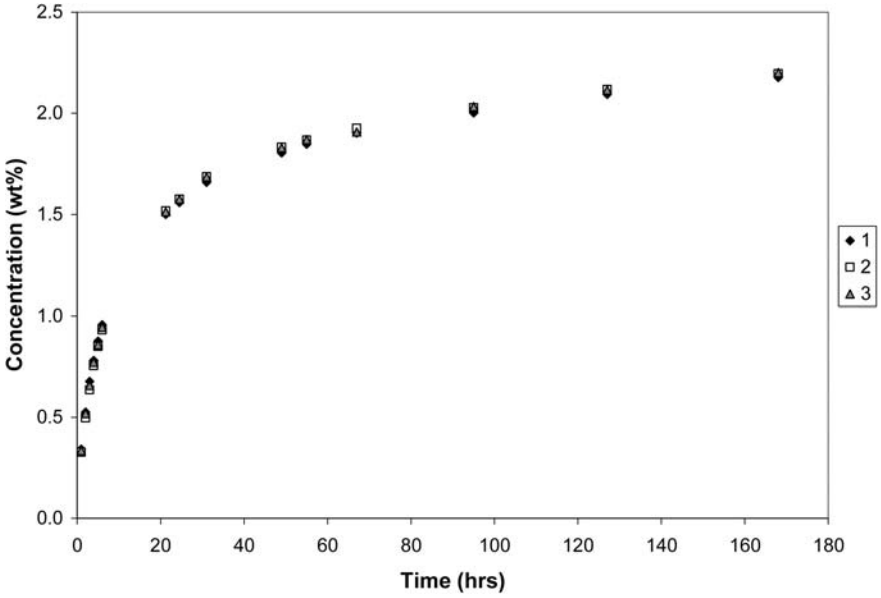


FIGURE 17.1. Moisture uptake profile for UR-A at 85°C/85%RH. (Ferguson and Qu [20], reprinted with permission of ASME.)

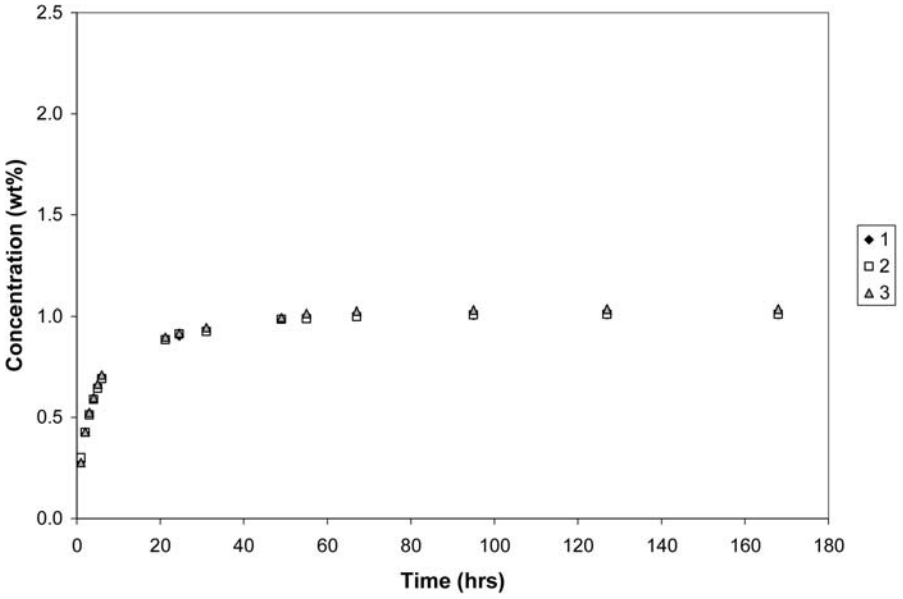


FIGURE 17.2. Moisture uptake profiles for UR-B at 85°C/85%RH. (Ferguson and Qu [20], reprinted with permission of ASME.)

found some of their epoxies to exhibit a gradual increase in moisture content with time, attributing this increase to void growth in the epoxy network caused by swelling.

The diffusivity of moisture through the thickness of the underfill resin is needed to apply an analytical, Fickian solution for modeling the moisture diffusion into the in-

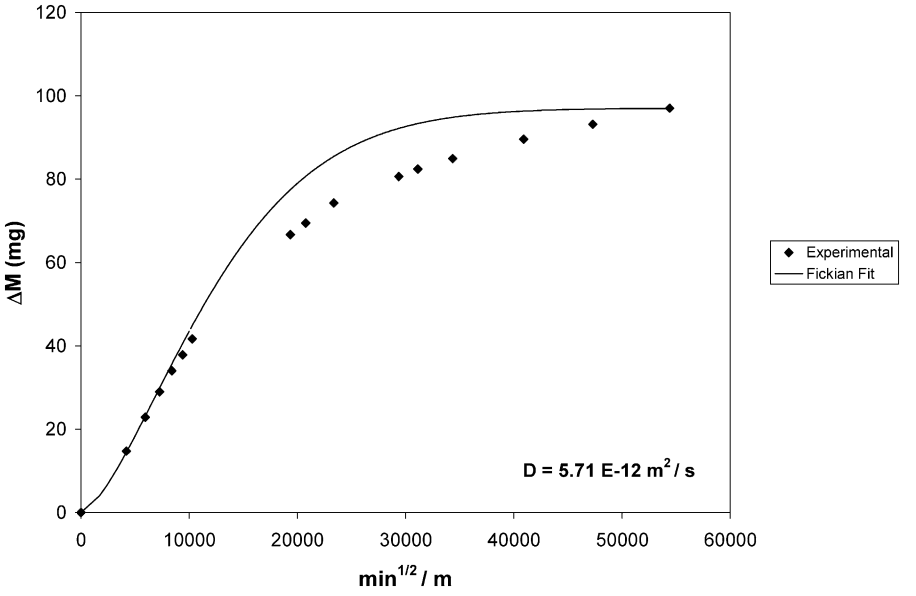


FIGURE 17.3. Fickian curve fit at 85°C/85%RH for UR-A. (Ferguson and Qu [20], reprinted with permission of ASME.)

terfacial fracture test specimens. The diffusion coefficient, D, can be experimentally determined using a test specimen that promotes predominantly one-dimensional diffusion into the test specimen. Using Equation (17.7) with absorption data, the diffusion coefficients were experimentally determined for both UR-A ($D = 5.71E{-}12$ m^2/s) and UR-B ($D = 1.47E{-}11$ m^2/s). Since the diffusion coefficient is a measure of how quickly a material responds to mass concentration changes in its environment, the larger value of diffusivity for UR-B indicates it will respond more quickly to those changes. As a result, UR-B test specimens will approach saturation more rapidly than UR-A test specimens, which quantitatively supports what was already qualitatively observed (Figures 17.1 and 17.2).

A Fickian curve was generated for each underfill to examine the extent that the moisture uptake of the specimens demonstrated Fickian behavior at conditions of 85°C/85%RH. The following relation developed by Shen and Springer [44] was implemented to generate the Fickian profile since it simplifies the infinite series of Equation (17.5):

$$\frac{M_t}{M_\infty} = 1 - \exp\left[-7.3\left(\frac{Dt}{h^2}\right)^{0.75}\right]. \tag{17.8}$$

A Fickian curve for each data set at 85°C/85%RH is shown in Figures 17.3 and 17.4.

It is clear from Figures 17.3 and 17.4 that neither UR-A nor UR-B exhibited true Fickian behavior at 85°C/85%RH, although UR-B appeared to obtain a better curve fit than UR-A. Since test specimens promoted predominately one-dimensional diffusion and exhibited non-Fickian absorption behavior, it is evident that the diffusion coefficients of both UR-A and UR-B were dependent on moisture concentration rather than being constant throughout the entire diffusion process at 85°C/85%RH.

Wong et al. [56] found varied diffusion behavior in the epoxy resins they evaluated at 85°C/85%RH, with some resins exhibiting Fickian diffusion while others did not. They

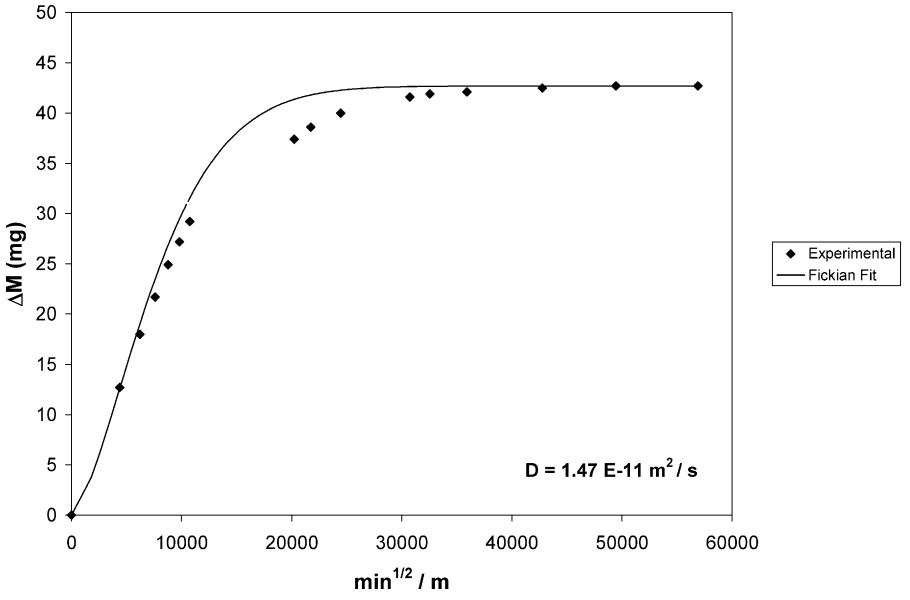


FIGURE 17.4. Fickian curve fit at 85°C/85%RH for UR-B. (Ferguson and Qu [20], reprinted with permission of ASME.)

postulated that diffusivity is constant and moisture diffusion exhibits Fickian behavior for epoxy resins at lower temperature and humidity levels such as 30°C/60%RH. Increasing the humidity level results in a corresponding amplification of the saturation level, while increasing the temperature level produces more prominent non-Fickian behavior [54]. Although test specimens will absorb more moisture in less time at higher temperatures and relative humidity levels, the trade-off is that the specimens will also exhibit an increased likelihood of non-Fickian diffusion behavior. The concentration dependence of the diffusivity in non-Fickian diffusion behavior complicates the modeling of the moisture ingress; however, numerical algorithms have been published that demonstrate how to model the non-Fickian diffusion process [56].

17.2.4. Moisture Absorption Modeling

To illustrate the moisture distribution graphically in interfacial fracture test specimens, a transient, finite element analysis was implemented to model the associated moisture concentration distribution in test specimens for small times of exposure. Since the substrates of the interfacial fracture test specimens were metallic and impermeable to moisture, it should be noted that only the moisture distribution in each underfill was modeled. Results of the finite element model illustrating the transient moisture distribution in the underfill resins are shown in Figures 17.5 and 17.6. Both figures refer to the interfacial fracture test specimens as unmodified, which indicates that this is the moisture absorption behavior exhibited by the test specimens if placed in 85°C/85%RH conditions immediately after test specimen manufacture without consideration to how the moisture uptake could influence fracture results.

It is apparent from the model of the transient moisture ingress that edge effects are significant. This can be clearly seen by examining the interface of the test specimens in

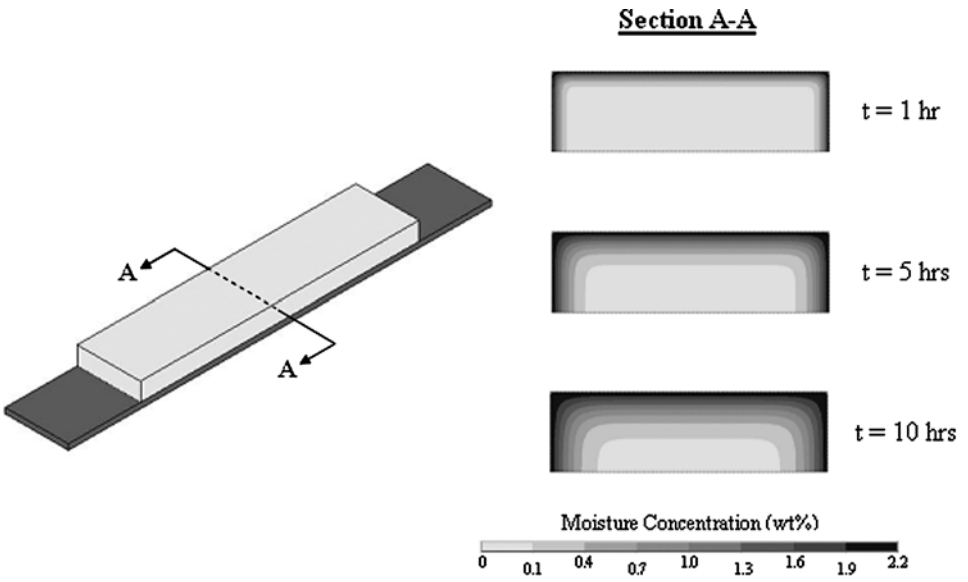


FIGURE 17.5. Moisture concentration distribution for unmodified UR-A interfacial fracture test specimen at 85°C/85%RH after 1, 5, and 10 hours of exposure. (Ferguson and Qu [20], reprinted with permission of ASME.)

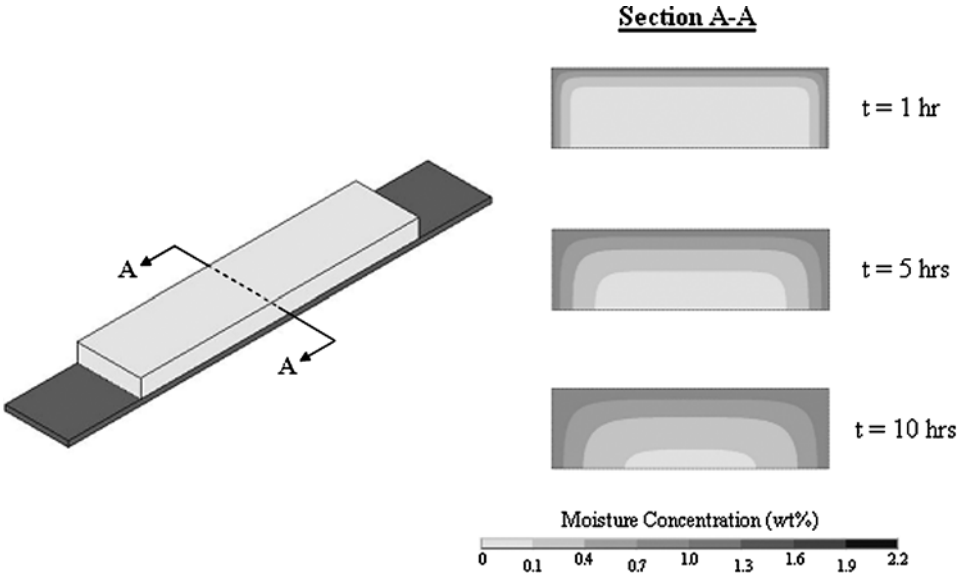


FIGURE 17.6. Moisture concentration distribution for unmodified UR-B interfacial fracture test specimen at 85°C/85%RH after 1, 5, and 10 hours of exposure. (Ferguson and Qu [20], reprinted with permission of ASME.)

Figures 17.5 and 17.6 (bottom of each cross section A–A), where it is evident that a gradient of moisture exists at the interface until saturation is reached. This is undesirable since the interface will experience different levels of moisture spatially relative to the exposure time, which will not allow a fracture toughness measurement to be identified with a par-

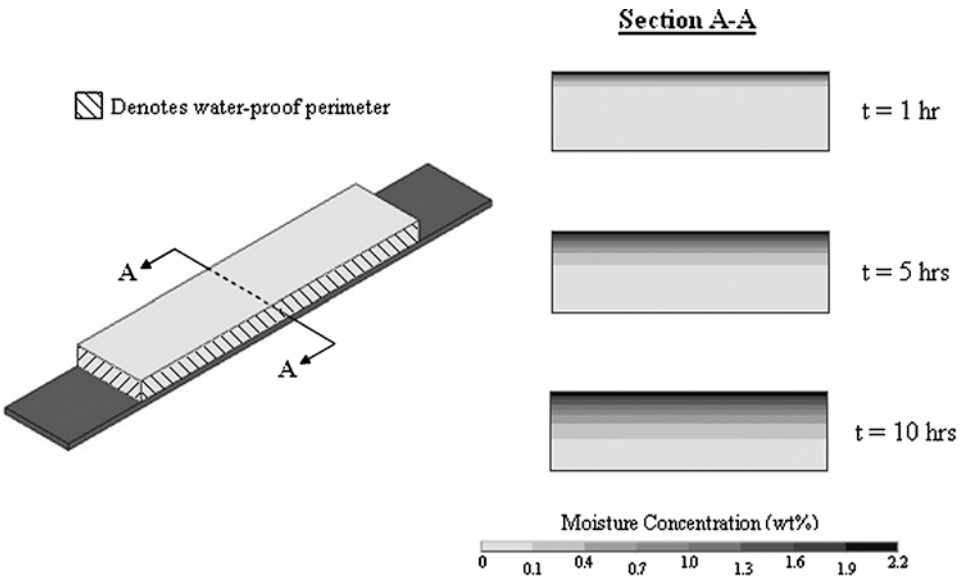


FIGURE 17.7. Moisture concentration distribution for modified UR-A interfacial fracture test specimen at 85°C/85%RH after 1, 5, and 10 hours of exposure.

ticular level of interfacial moisture concentration until saturation is reached. Furthermore, it is also possible that the non-uniform moisture gradient at the interface could influence interfacial fracture toughness results even if saturation is reached in a test specimen. This is attributed to different areas of the interface being exposed to varying degrees of moisture for different periods of time, which could have an effect on fracture toughness results even if test specimens are in a saturated state. Last, wicking of moisture along the interface could also introduce moisture concentration levels that remain unidentified through modeling of the absorption process alone. This would make it difficult to attribute a particular fracture toughness measurement with an associated interfacial moisture concentration level.

In view of these observations, the interfacial fracture test specimen design should be modified with a water-proof perimeter applied to test specimens prior to moisture preconditioning. The application of the water-proof perimeter forces 1D moisture uptake through the top surface of the test specimens and prevents wicking along the interface. Not only does this yield uniform concentrations spatially at the interface, but it also aids in the identification of an interfacial moisture concentration level by utilizing the inherent moisture absorption characteristics of the adhesive to restrict the amount of moisture arriving to the interface. Figures 17.7 and 17.8 depict the moisture concentration distribution in the modified interfacial test specimens.

Although percent weight is dependent on both the specimen volume and density, a comparison between the moisture concentration distributions can be made as a result of both underfills having similar densities (UR-A, $\rho = 1.14 \times 10^{-3}$ g/mm^3 and UR-B, $\rho = 1.16 \times 10^{-3}$ g/mm^3) and volumes. Figures 17.7 and 17.8 illustrate that although UR-A test specimens contain a significantly higher concentration of moisture near the underfill surface, the moisture will actually penetrate the interface first for comparably sized UR-B test specimens. It is clear from the progression of the constant-concentration lines depicted in Figures 17.7 and 17.8 that the moisture traversed much more easily through the UR-B

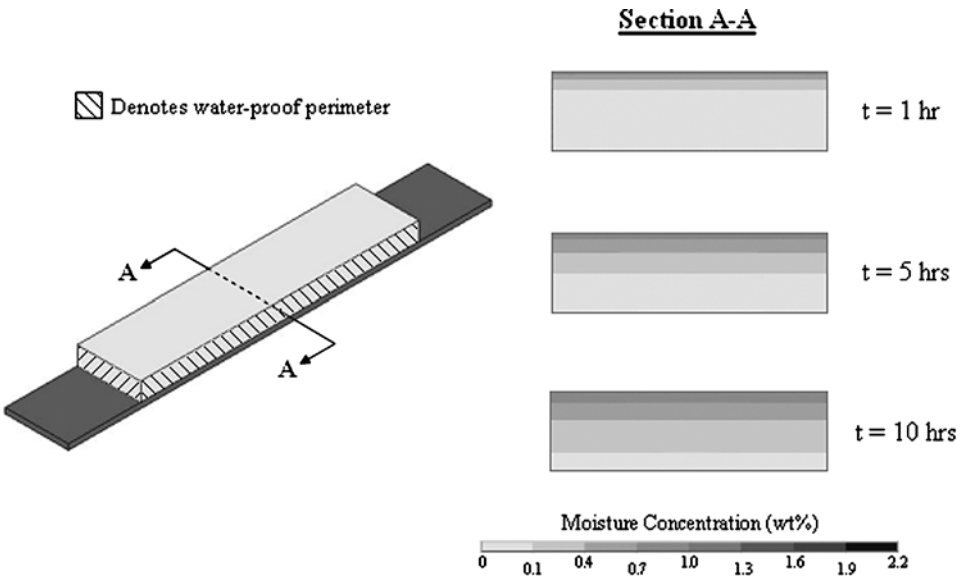


FIGURE 17.8. Moisture concentration distribution for modified UR-B interfacial fracture test specimen at 85°C/85%RH after 1, 5, and 10 hours of exposure.

test specimens. An explanation for this behavior lies in the particular chemistry of each underfill epoxy with respect to the polarity of water molecules. As previously noted, amine functional groups regulate transport through the nanopore channels of the epoxy by either blocking or allowing moisture to traverse the channels depending on the orientation of the resin with respect to nanopore position [48]. UR-A contains amine functional groups, while UR-B is a non-amine containing underfill [20]. Consequently, it would be anticipated that UR-B would have an enhanced diffusion coefficient than UR-A, which was found to be true based on experimental results. As demonstrated in Figures 17.7 and 17.8, the amine functional groups present in UR-A contributed to retard moisture penetration into the amine containing epoxy resin, whereas the moisture diffused more easily through the non-amine epoxy resin, UR-B. From this observation, there are three primary conclusions to consider:

1. Degradation of interfacial adhesion over the entire interface due to the presence of moisture will initially occur in UR-B test specimens prior to comparably sized UR-A test specimens.
2. Further degradation of interfacial adhesion will occur in UR-A test specimens than comparably sized UR-B test specimens for longer exposure times. This is due to UR-A absorbing more aggregate moisture than UR-B at longer durations (Figures 17.1 and 17.2).
3. Amine functional groups present in UR-A retard the rate by which moisture exits the underfill upon redrying. Consequently, UR-A test specimens will take longer to recover the reversible component of adhesion loss upon removal of moisture from the interface than comparably sized UR-B test specimens.

The absorption characteristics, exposure time, and adhesive performance from moisture dictate whether UR-A or UR-B is a more robust product in humid environments. For short exposure times to moist environments and considering only the absorption characteristics, UR-A represents a better encapsulant by retarding the rate of moisture ingress to

the interface through the presence of amine functional groups in its chemistry. Conversely, UR-B represents a better product for longer exposure times to moist environments by absorbing less aggregate moisture than UR-A. Bare in mind neither of the aforementioned statements considers the relative adhesion performance in the presence of moisture nor the extent by which each underfill adhesive bond recovers after multiple exposures to moist environments. These are additional considerations when evaluating the long term reliability of a particular underfill to moist conditions.

17.3. ELASTIC MODULUS VARIATION DUE TO MOISTURE ABSORPTION

The deleterious effect of moisture not only damages interfacial adhesion by being physically present at the interface, but also through the degradation of the elastic modulus of the adhesive and substrate due to moisture uptake. The change in the elastic modulus after moisture uptake can be substantial, which can significantly affect material performance and adhesion results. Consequently, the variation in the elastic modulus of the adhesive and substrate as a function of moisture concentration should be determined to completely characterize the loss in interfacial adhesion due to moisture absorption. Since many of the substrates used in electronic packaging are impermeable to moisture (i.e., copper, aluminum, and silicon), it is often only necessary to characterize the change in the elastic modulus as a function of moisture concentration for the adhesives, which are typically epoxy based and highly susceptible to moisture uptake.

17.3.1. Background

Epoxy adhesives are found in many microelectronic packaging applications and widely used throughout the industry. One of the more substantial developments within the last ten years is underfill, which is an epoxy based encapsulant that mechanically couples the chip to the board. Underfill drastically enhances the fatigue life of microelectronic assemblies when compared to unencapsulated devices [51]; however, since underfills are epoxy based, they are also particularly vulnerable to moisture ingress [20,22,53,56]. Although the absorbed moisture can significantly alter its mechanical performance and the overall microelectronic assembly reliability, very few studies in the electronic packaging literature have addressed the issue of moisture on the mechanical properties of epoxies.

Throughout the literature, the availability of information regarding the effect of moisture on the mechanical properties of epoxy adhesives is in general limited and more work is needed to adequately characterize this response [12,24]. From the work that has been published, it has been found that water absorption can severely modify the mechanical properties of epoxies by decreasing the elastic modulus [36,63], shear modulus [27,64], yield stress [55], and ultimate stress [55] while increasing the failure strain [12,55] as water concentration increases. A representative stress/strain diagram is shown in Figure 17.9 illustrating these effects.

Moisture primarily affects the mechanical properties of epoxy adhesives through three mechanisms: plasticization, crazing, and hydrolysis. The first is considered reversible upon drying, while the latter two are irreversible. Several studies attribute the decrease in modulus due to the plasticizing action of the water on the adhesive [3,12,15,27,49, 55,63,64]. By acting as an external plasticizer to the adhesive, the water spreads the polymer molecules apart and reduces the polymer–polymer chain secondary bonding. This provides more room for the polymer molecules to untangle and move, which results in a softer,

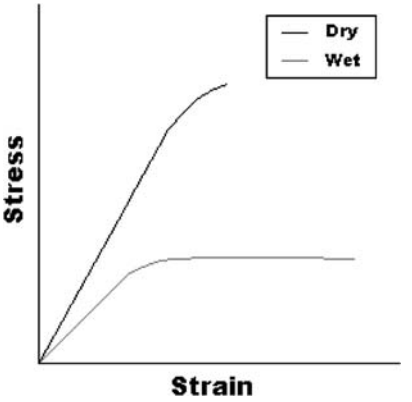


FIGURE 17.9. Representative stress/strain diagram depicting the effect of moisture on the mechanical properties of bulk epoxy adhesives.

more easily deformable mass [42]. Other studies show the decrease in epoxy modulus after moisture absorption resulting from crazing [34,36,37], where the absorbed water can act as a crazing agent continuously decreasing the mechanical strength of epoxies with exposure time in water [34]. This is supported by scanning electron micrographs of epoxies, which show cavities and fractured fibrils that could only be explained by a moisture-induced crazing mechanism [37]. The moisture-induced swelling creates dimensional changes and internal stresses that can ultimately craze and/or crack the material. As a result, lightly cross-linked networks will be more susceptible to crazing than highly cross-linked networks [36]. Last, moisture can also affect the mechanical properties of epoxy adhesives by causing hydrolysis leading to chain scission. Short term exposure to moisture results in chain scission with a chemical addition of water that remains permanently in the epoxy system even after subsequent drying. Long term exposure to moisture can result in an increased probability of chain scission detaching segments from the polymer network, yielding a permanent loss in weight after subsequent drying [59].

Studies by Zanni-Deffarges and Shanahan [63,64] and DeNeve and Shanahan [15] depict the decrease in elastic and shear modulus of an epoxy as a function of time exposure to moisture. Although this information is useful in evaluating the effect of exposure time to moisture on the modulus, it does not depict how the inherent wet modulus values change as a function of concentration. This is due to a gradient of mechanical properties that will exist in the adhesive until saturation is reached, where water concentrations become steady and uniform. Other studies have evaluated the effect of moisture on epoxy adhesives after saturation is established for a given level of moisture preconditioning. These studies have shown a decrease in the elastic modulus of epoxy adhesives of 24% [64], 29% [49], and 86% [49] for saturation concentrations of 4 wt%, 0.9 wt%, and 3.1 wt% respectively; however, they only tested one level of moisture preconditioning to compare to fully dried test results. Consequently, information regarding the mechanical response of epoxy adhesives to different levels of moisture concentrations is incomplete and fundamental insight into the intrinsic response of the adhesives to increasing saturation concentrations of moisture cannot be ascertained.

Even fewer investigations have evaluated the recovery of epoxies upon drying after moisture absorption with little information available regarding the extent of the reversible and irreversible nature of moisture uptake in epoxies. Netravali et al. [38] have shown for

epoxy samples soaked in water at 25°C for 820 hours that much of the loss from moisture results from plasticization and is recoverable upon drying at 30°C for 400 hours; however, samples soaked in water at 70°C for 775 hours were highly irreversible after drying at 70°C for 125 hours. The irreversibility was attributed to water reacting with unreacted epoxide groups. It should be noted that neither groups of samples were completely dry at the time of testing after exposure to water and subsequent drying. Buehler and Seferis [4] found epoxy prepregs soaked in water at 71°C for 1200 hours exhibited varying degrees of reversible and irreversible damage to both the flexural modulus and flexural strength upon drying at 50°C for 450 hours. However, more time was needed to fully dry the specimens in this study as well, with 3% weight concentrations of moisture still existing in the specimens at the time of testing after drying. Wright [57] proposes that the permanent loss of properties that occur due to moisture uptake is most probably due to swelling of the matrix and the production of voids, while Xiao and Shanahan [59] suggest based on absorption behavior that the irreversible damage component of hydrolysis can play a significant role in the degradation process depending on the duration of exposure. Undoubtedly the mechanisms responsible for the observed losses in epoxies from moisture uptake are complex, and the material constitutive damage behavior is not entirely understood.

17.3.2. Effect of Moisture Preconditioning

To help characterize the elastic response of an epoxy adhesive to increasing moisture concentrations, an evaluation of the effect of moisture on the elastic modulus of a no-flow underfill was performed. The particular underfill evaluated was UR-B, which was determined to be ideal for studying the fundamental effect of moisture on interfacial adhesion due to its moisture diffusion kinetics and saturation behavior established from the moisture absorption portion of this research. Flexural specimens were tested in a three-point bend test according to ASTM D790 [2] to determine the effect of moisture on the elastic modulus.

Test specimens were divided into six test groups and subjected to five different levels of moisture preconditioning to ascertain the effect of moisture on the elastic modulus of the underfill. The test groups included fully dry, 85°C only, 85°C/50%RH, 85°C/65%RH, 85°C/85%RH, and 85°C/95%RH, with the latter five test groups being environmentally preconditioned for 168 hours. All test specimens were baked at 115°C for at least 12 hours to remove any moisture that may have been introduced during sample preparation prior to environmental aging, which was performed in a humidity chamber in an atmosphere maintained at a constant temperature (±1°C), humidity (±1%), and pressure (P_{atm}). In addition, all flexural tests were performed with both the surrounding environment and test specimens being at room temperature after environmental preconditioning. No measurable loss in moisture uptake occurred in the test specimens from the time they were removed from the environmental chamber, allowed to cool to room temperature, and experimentally tested.

Figure 17.10 illustrates the effect of moisture preconditioning on the underfill elastic modulus for several different temperature/humidity levels. All moisture preconditioned test specimens were fully saturated at the conclusion of the 168 hour exposure time, hence a gradient of moisture concentration did not occur within the specimens so that the inherent wet modulus was identified. In addition, Differential Scanning Calorimetry (DSC) test results demonstrate that the underfill was fully cured in the flexural specimens for the curing conditions and test specimen size and geometry used [18]. Therefore, incomplete curing

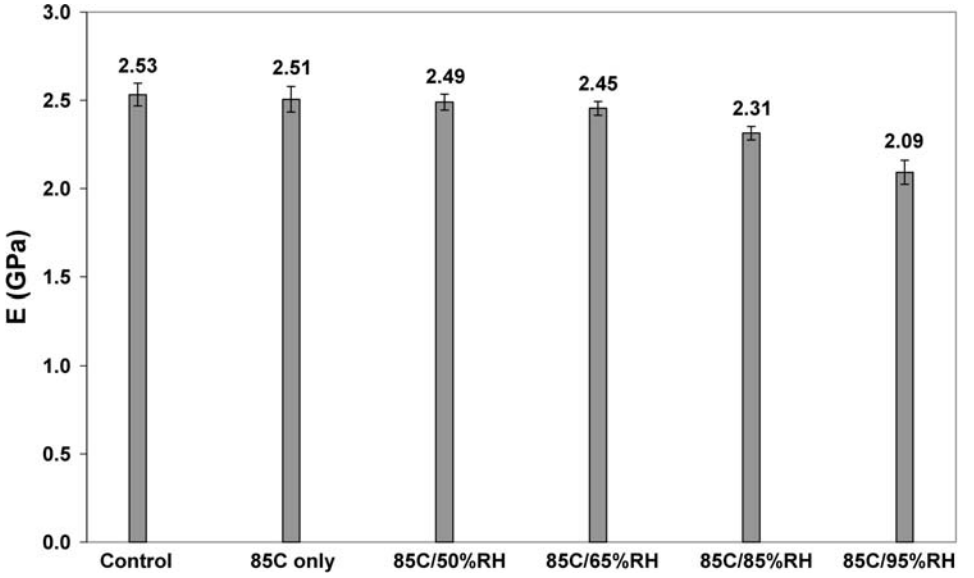


FIGURE 17.10. Effect of moisture preconditioning on the underfill elastic modulus. (Ferguson and Qu [18], reprinted with permission of IEEE.)

of the underfill in the flexural specimens did not influence any observed changes to the elastic modulus of the underfill after moisture preconditioning. Further information on the fundamentals and use of differential scanning calorimetry may be found in the works of Pasztor [40] and Prime [41].

When compared to unaged, control test specimen values, moisture preconditioning at 85°C/50%RH and 85°C/65%RH was found to have little to no effect on the elastic modulus of the underfill. A more noticeable effect occurs at 85°C/85%RH, while conditions of 85°C/95%RH yielded a significant decrease in modulus. To isolate the possible effect of thermal aging at 85°C from moisture preconditioning contributing to the observed changes in the elastic modulus of the underfill, flexural specimens were exposed to conditions of 85°C only for 168 hours and compared to unaged, control test specimen values. As shown in Figure 17.10, thermal aging at 85°C for 168 hours was found to have no effect on the elastic modulus with similar values obtained when compared to the control test group results. Again, it is important to note that all tests were performed at room temperature, hence only the effects of thermal aging were evaluated rather than the effect of testing at elevated temperatures on the elastic modulus. Since all environmental preconditioned test groups were exposed to the same temperature of 85°C and to the same duration of 168 hours, the observed changes in modulus from moisture preconditioning given in Figure 17.10 can be attributed to the effect of moisture and moisture alone.

A summary of the effect of moisture preconditioning on the elastic modulus of the underfill is given in Table 17.1, where C_{sat} represents the saturation concentration of moisture in the test specimens for each respective level of moisture preconditioning and given as both a percent weight change (wt%) and mg H_2O/mm^3.

Since saturation was reached in all moisture preconditioned test specimens prior to removal from the humidity chamber and thermal aging from the 85°C temperature component of moisture preconditioning was found to have no effect on the elastic modulus,

TABLE 17.1.
Change in the underfill elastic modulus from moisture uptake. (Ferguson and Qu [18], reprinted with permission of IEEE.)

T (°C)	RH (%)	C_{sat} (wt%)	C_{sat} (mg H_2O/mm^3)	E (GPa)	Modulus change (%)
Control	—	0	0.0000	2.53 ± 0.06	—
85	50	0.65	0.0075	2.49 ± 0.05	1.6
85	65	0.77	0.0089	2.45 ± 0.04	3.2
85	85	1.02	0.0118	2.31 ± 0.04	8.7
85	95	1.19	0.0138	2.09 ± 0.07	17.4

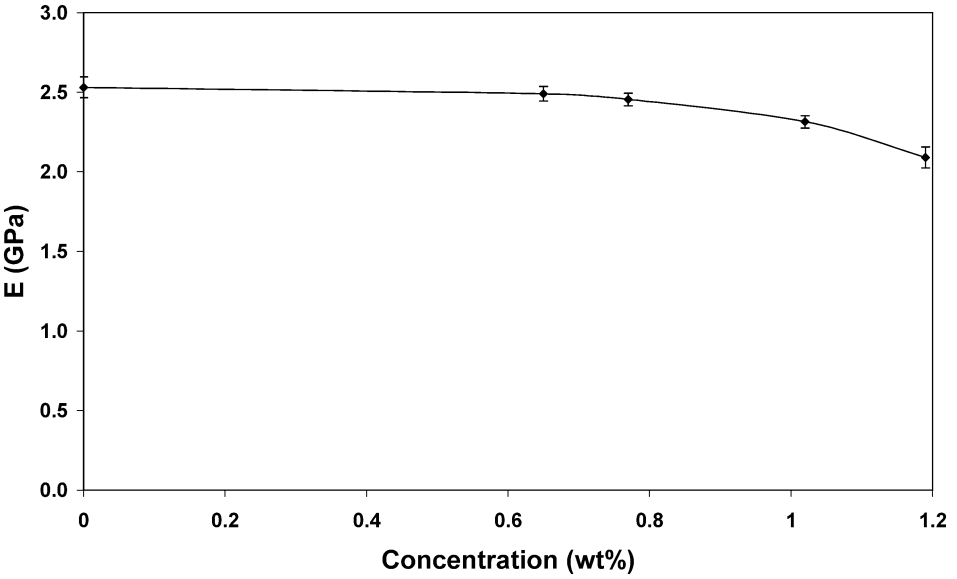


FIGURE 17.11. Underfill elastic modulus variation as a function of moisture concentration (wt%). (Ferguson and Qu [18], reprinted with permission of IEEE.)

the inherent wet modulus was identified and all observed changes in the modulus occurred solely from the influence of moisture. This allows the characterization of the change in modulus of the underfill from moisture uptake as a function of moisture concentration as shown in Figures 17.11 and 17.12.

Figures 17.11 and 17.12 depict the inherent change in the elastic modulus of an epoxy-based adhesive as a function of moisture concentration. Time dependent variation in the elastic modulus after saturation is assumed to be negligible, although it could be a consideration for longer durations of exposure at higher concentrations of moisture as a result of hydrolysis [59]. Previous studies on epoxy adhesives have shown the variation in modulus as a function of the square root of time corrected for specimen thickness [63]; however, this information depicts the change in modulus resulting from a transient, gradient of moisture concentration rather than demonstrating how the inherent wet modulus changes with increasing moisture content. Other studies have identified the inherent wet modulus for a single saturation level and compared to fully dry results [3,49,63]; however, these studies do not show the inherent wet modulus of the same adhesive for several dif-

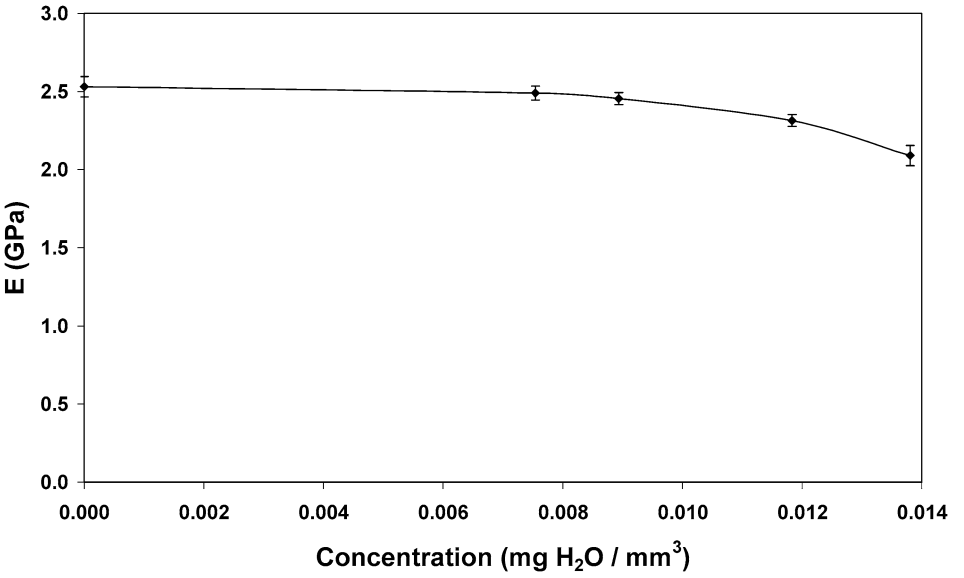


FIGURE 17.12. Underfill elastic modulus variation as a function of moisture concentration (mg H_2O/mm^3). (Ferguson and Qu [18], reprinted with permission of IEEE.)

ferent saturation levels and thus do not show the characteristic response of the adhesive as a function of increasing moisture concentration as given in Figures 17.11 and 17.12. Such information is extremely useful in predictive modeling efforts, where the intrinsic response of the elastic modulus as a function of increasing moisture concentration can be used in a coupled mechanical-diffusion analysis [55] to incorporate the transient effect of the continual variation of elastic modulus as moisture diffuses into the adhesive. These data are not only significant when modeling the effect of moisture on the bulk material behavior, but also on interfacial adhesion, where changes in the mechanical properties of the adhesive due to moisture uptake can play a significant role in the onset of package delamination.

17.3.3. Elastic Modulus Recovery from Moisture Uptake

To further characterize the response of the underfill from moisture uptake and identify the mechanisms responsible for the observed losses in the elastic modulus from moisture absorption, test specimens were moisture preconditioned for 168 hours followed by baking at 95°C until fully dry. A fully dried state was established when there was no measurable change in the weight of a specimen for a period of 24 hours. Since 85°C/85%RH and 85°C/95%RH moisture preconditioning conditions were found to noticeably decrease the elastic modulus of the underfill, only those conditions were evaluated for recovery of the elastic modulus from moisture uptake upon redrying. Figure 17.13 provides a graphical depiction of the recovery results for the underfill elastic modulus.

As shown in Figure 17.13, much of the observed loss in the elastic modulus from moisture uptake was recoverable upon subsequent drying. Since plasticization is the only primary degradation mechanism attributed to moisture that is regarded as a reversible process, the recovery results demonstrate that the majority of the loss in modulus resulted from plasticization of the underfill from moisture uptake. To further evaluate the change

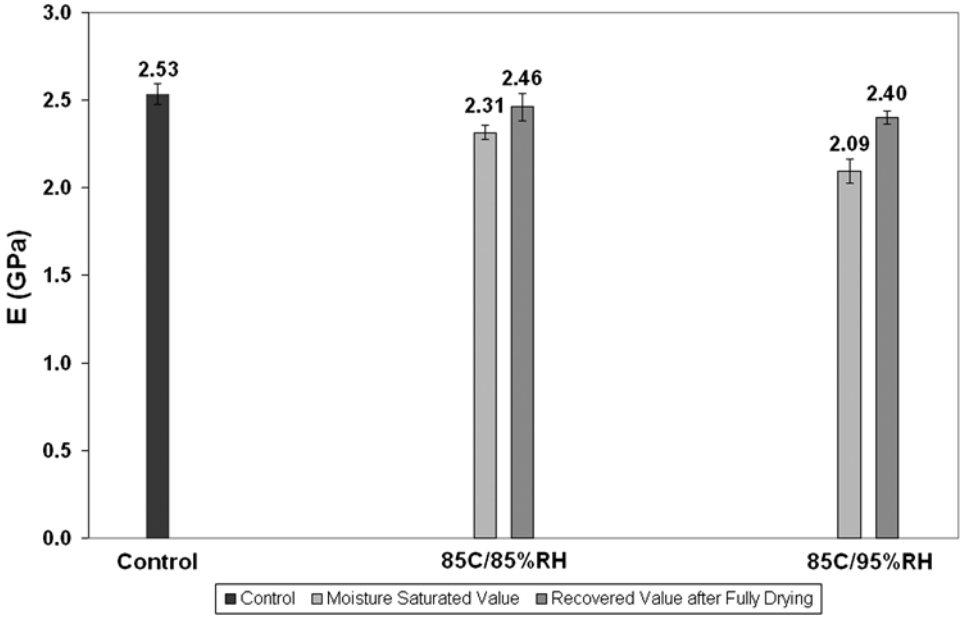


FIGURE 17.13. Recovery of the underfill elastic modulus on removal of moisture. (Ferguson and Qu [18], reprinted with permission of IEEE.)

TABLE 17.2.
Recoverability of the underfill elastic modulus from moisture uptake after subsequent drying. (Ferguson and Qu [18], reprinted with permission of IEEE.)

T (°C)	RH (%)	C_{sat} (wt%)	E_{sat} (GPa)	$E_{recovery}$ (GPa)	*Recoverability* (%)
Control	—	0.00	2.53 ± 0.06	—	—
85	50	0.65	2.49 ± 0.05	—	—
85	65	0.77	2.45 ± 0.04	—	—
85	85	1.02	2.31 ± 0.04	2.46 ± 0.08	68.2
85	95	1.19	2.09 ± 0.07	2.40 ± 0.05	70.5

in elastic modulus from moisture uptake, the recoverability for the elastic modulus will be defined as follows:

$$\textit{Recoverability}\ (\%) = \frac{E_{recovery} - E_{sat}}{E_{dry} - E_{sat}} \cdot 100, \tag{17.9}$$

where $E_{recovery}$ is the value of the elastic modulus upon fully drying from the moisture saturated state, E_{sat} is the saturated value of the elastic modulus after moisture absorption, and E_{dry} is the unaged, control value of the elastic modulus. The recoverability of the elastic modulus is given in Table 17.2.

Although a significant portion of the elastic modulus was recoverable after fully drying, some irreversible, permanent damage did occur. The average recoverable value of the elastic modulus suggests slightly more irreversible damage occurred at higher humidity levels, but it cannot be concluded unequivocally solely based on the modulus results due

to the uncertainty associated within the two measurements. However, it can be concluded when considering the results from moisture uptake data. After fully drying, there was a slight net permanent weight increase in the test specimens, with specimens moisture preconditioned at 85°C/85%RH retaining $1.3 \pm 0.5\%$ of the total absorbed water while specimens moisture preconditioned at 85°C/95%RH retaining $2.3 \pm 0.4\%$ of the total absorbed water. The permanent weight increase in the test specimens after subsequent fully drying suggests that at least part of the irreversible damage resulted from hydrolysis with a greater extent occurring at higher humidity levels. In addition to hydrolysis, it is also possible that moisture-induced crazing also contributed to the irreversible damage to the elastic modulus. Overall, the irreversible damage was small with the majority of the loss in the elastic modulus from moisture uptake being fully recoverable after subsequent drying.

17.4. EFFECT OF MOISTURE ON INTERFACIAL ADHESION

The effect of moisture on interfacial adhesion is governed by two fundamental mechanisms. The first is the rate at which moisture is delivered to the interface, and the second is the change in adhesion performance as a consequence of moisture being present in the adhesive structure. This includes not only the primary effect of moisture being directly present at the interface itself, but also the secondary effect of moisture altering the mechanical performance of the two materials that constitute the bimaterial interface. Having previously quantified both the rate at which moisture is delivered to the interface and the degrading effect of moisture on the elastic modulus of the materials that constitute the bimaterial interface, a model depicting the intrinsic change in interfacial adhesion as a function of moisture concentration is developed. Interfacial fracture mechanics is used to characterize this change to develop relationships that are independent of test specimen geometry.

17.4.1. Background

With interconnect density increasing and package size decreasing, several adaptations to microelectronic assemblies have been developed to accommodate the increasing demand in both cost and performance requirements. In particular, epoxy-based encapsulants have been extensively used in microelectronic devices to enhance package reliability, provide environmental protection, and improve manufacturing yields. To insure these benefits are not compromised, the structural integrity of the adhesive bond must be maintained. Characterizing the primary adhesion mechanisms and identifying the factors that affect the strength and durability of encapsulants are critical to their success.

Traditional encapsulation processes, such as transfer molding, cavity filling, and glob-topping, are used to protect the IC device from environmental pollutants and provide mechanical support. In these devices, copper alloys are typically used as a lead frame material due to their low cost in conjunction with their high electrical and thermal conductivity. However, the adhesion at the epoxy/copper interface is poor [7,9,29,32]. In addition, the copper surface is highly susceptible to oxidation. This is an additional consideration when evaluating the interfacial adhesion of adhesives with copper.

A more recent encapsulant developed within the last ten years is underfill, which is an epoxy-based encapsulant that mechanically couples the chip to the board. Underfill drastically enhances the fatigue life of microelectronic assembles when compared to unencapsulated devices [51], provided the adhesive bond between the underfill and the printed

wiring board, solder mask, copper, silicon, passivation, and solder is maintained. Characterizing the adhesion of underfill to these substrates has been the focus of several studies in adhesion and reliability research [13,14,17,19,60].

Although epoxy encapsulants have many benefits, they are susceptible to moisture uptake. A typical epoxy formulation can absorb between 1 and 7 wt% moisture [47], which can have a detrimental effect on interfacial adhesion and drastically reduce the reliability of encapsulated devices. While it has been shown that moisture can significantly alter adhesive performance in microelectronic packaging [21,34], the interfacial and material constitutive damage behavior from moisture exposure is not well understood. This largely arises due to the difficulty of the problem, which is governed by two fundamental mechanisms. The first is the rate at which moisture is delivered to the interface. The second is the response of the interfacial adhesion to varying levels of moisture concentration, where the deleterious effect of moisture not only affects interfacial adhesion by being physically present at the interface, but also through the degradation of the mechanical properties of the epoxy adhesive due to moisture uptake. Mass transport and in particular the diffusion of moisture in epoxy adhesives has been studied by several sources and is fairly well established [47,48, 52,54,56]; however, the response of interfacial adhesion to moisture is much less understood. Although several studies have addressed the issue of moisture, much more work needs to be completed and there currently exists a lag in fundamental empirical data depicting the loss in interfacial adhesion as a function of interfacial moisture concentration. Since there exists this lag in experimental data, even less effort has been spent on developing predictive models that account for the effect of moisture on interfacial adhesion.

Of particular interest to the long-term reliability of an adhesive bond is ascertaining the permanent damage to the bond from exposure to moisture. Very few studies have examined the reversible and irreversible components of the loss in adhesion from moisture after subsequent drying. This has significant practical aspects, as the recoverability of the interface from moisture will identify the severity of the moisture damage. If the loss in adhesion from moisture is largely unrecoverable and irreversible, then the service life of the adhesive joint will be severely, permanently compromised as a result of exposure to moisture. Such consequences would bring added emphasis to protecting the encapsulated package from moisture ingress and developing more robust, moisture-resistant adhesives.

When evaluating the moisture recovery of an adhesive joint, there are two aspects to consider. The first is the recovery of the materials that constitute the adhesive joint, as absorbed moisture can alter the mechanical performance of those materials and indirectly affect adhesion [18]. The second aspect is the recovery of the interfacial bonding itself, as the direct presence of moisture at the interface can significantly alter adhesion. Butkus [5] examined the permanent change in Mode I fracture toughness of Aluminum/FM73M/Aluminum and Aluminum/FM73M/Boron-Epoxy joints after 5000 hours at 71°C and $>$90%RH followed by 5000 hours of desiccation at 22°C/10%RH prior to testing. Both the Al/FM73M/Al joints and the Al/FM73M/Boron-Epoxy joints recovered very little of their fracture toughness on subsequent drying, demonstrating large, permanent losses in toughness after exposure to moisture. Orman and Kerr [39] have shown that although some of the strength lost in the epoxy-bonded aluminum joints they studied was recovered, there was noticeable permanent damage from moisture suggesting an irreversible disruption at the interface as a result of attack by water. Contrary to this claim, Shaw et al. [43] found that nearly all of the strength lost after immersing steel/epoxy lap shear joints in distilled water for three weeks was recovered after drying. They attributed the loss in strength after moisture preconditioning to plasticization of the epoxy adhesive,

which is generally regarded as a reversible process. Dodiuk et al. [16] found exposure to moisture of their epoxy/aluminum joints caused a reduction in lap shear strength; however, if the moisture concentration was below 0.3 wt%, the strength was fully recoverable after drying indicating a completely reversible process. The authors gave no explanation to this observed behavior other than to state that moisture concentrations exceeding 0.3 wt% would result in an irreversible process and permanent loss of adhesion at the interface. Undoubtedly the mechanisms responsible for the observed losses in both material behavior and interfacial adhesion from moisture uptake are complex, and the material constitutive damage behavior is not entirely understood.

17.4.2. Interfacial Fracture Testing

Interfacial fracture toughness is defined as the critical value of the energy release rate, G_c, at which a bimaterial interface will begin to delaminate. It is a property that characterizes the adhesion of a bimaterial interface, independent of the size and geometry of the cracked body. For a bimaterial interface loaded in four point bending under plane strain conditions, it can be shown that the critical value of the energy release rate, G_c, can be determined using the following equation [26]:

$$G = \frac{1}{2\overline{E_1}}\left(\frac{12M^2}{h^3}\right) - \frac{1}{2\overline{E_2}}\left(\frac{M^2}{Ih^3}\right), \tag{17.10}$$

where

$$\overline{E_i} \equiv \frac{E_i}{1-\nu_i^2}, \tag{17.11}$$

M is the moment, ν is Poisson's ratio, E is the elastic modulus, subscript 1 refers to material 1, subscript 2 refers to material 2, h is the height of material 1, and I is the dimensionless moment of inertia.

Since the interfacial fracture toughness only specifies the magnitude of the crack tip singularity, the mode mixity, ψ, must be determined from the complex stress intensity factor K. For a two-dimensional system, the complex stress intensity factor, K, is given by:

$$K = K_1 + iK_2. \tag{17.12}$$

For four-point loading conditions it can be shown [26]:

$$K = h^{-i\varepsilon}\sqrt{\frac{1-\alpha}{1-\beta^2}}\left(\frac{P}{\sqrt{2hU}} - ie^{i\gamma}\frac{M}{\sqrt{2h^3V}}\right)e^{i\omega} \tag{17.13}$$

with the mode mixity given by:

$$(K_1 + iK_2)L^{i\varepsilon} = |(K_1 + iK_2)|e^{i\psi}, \tag{17.14}$$

$$\psi = \tan^{-1}\left(\frac{\mathrm{Im}(KL^{i\varepsilon})}{\mathrm{Re}(KL^{i\varepsilon})}\right), \tag{17.15}$$

where L is the characteristic length and ε is a dimensionless quantity given by Hutchinson and Suo [26]. As shown in Equation (17.15), the mode mixity for a test specimen requires the specification of some length quantity, L. The choice for L is arbitrary, but it should be selected as a fixed length and reported with the calculated values for the mode mixity.

The flexural beam test for interfacial fracture testing has three primary benefits. First, it yields intermediate values for mode mixity, which is representative of the values experienced by electronic devices during actual application. Second, it provides a means for successful interfacial fracture test specimen construction utilizing substrates and adhesives common to microelectronic packaging. Last, the flexural beam test configuration yields an open-faced test specimen design, which allows saturated, steady state conditions to be reached in the test specimens in a relatively short amount of time. This is due to the large surface area for moisture uptake relative to the short diffusion path to the interface.

17.4.3. Effect of Moisture Preconditioning on Adhesion

Interfacial fracture mechanics was used to characterize the intrinsic effect of moisture on adhesion. The adhesive used was an epoxy-based underfill developed for no-flow assembly, designated as UR-B in this research. This particular underfill was determined to be ideal for studying the fundamental effect of moisture on interfacial adhesion due to its moisture diffusion kinetics and saturation behavior established from the moisture absorption portion of this research. The substrate used was oxygen-free electronic grade copper, alloy 101. The copper substrates were polished to a mirror finish and cleaned using the routine procedure given by Shi and Wong [45] prior to bonding. This was done to isolate the intrinsic effect of moisture on adhesion without mechanical interlocking and/or surface contamination influencing the results. Symmetric interface cracks were introduced into the underfill/copper bilayer test specimens by using a molding compound release agent [19]. Based on the results from the moisture absorption analysis, a water-proof perimeter was applied to the interfacial fracture test specimens during moisture preconditioning and removed before fracture testing. This perimeter served two purposes. First, the application of the perimeter forced 1D diffusion through the top, open surface of the underfill, yielding uniform concentrations of moisture spatially across the entire interface for the full duration of exposure to the humid preconditioning environment. Second, the water-proof perimeter prevented moisture wicking at the interface, which allowed identification of the test specimen moisture concentration by utilizing the inherent moisture absorption characteristics of the adhesive. Completed specimens were tested in a four-point bend test at room temperature to measure the critical load of fracture for the interface. A completed representative interfacial fracture toughness test specimen is shown in Figure 17.14.

Test specimens were divided into five test groups and subjected to four different levels of moisture preconditioning to ascertain the effect of moisture on interfacial fracture toughness. The test groups included fully dry, 85°C only, 85°C/50%RH, 85°C/65%RH, and 85°C/85%RH, with the latter four test groups being environmentally preconditioned for 168 hours. The 85°C temperature component in each moisture preconditioning environment will enhance diffusion rates and drive more moisture into test specimens over a smaller timeframe. In addition, as temperature increases, the moisture capacity of air increases. Consequently, more moisture will be available to diffuse into test specimens at higher relative humidity levels compared to similar, high relative humidity levels at lower temperatures. By gradually increasing the relative humidity while maintaining the temperature constant, the change in interfacial fracture toughness as a function of increasing

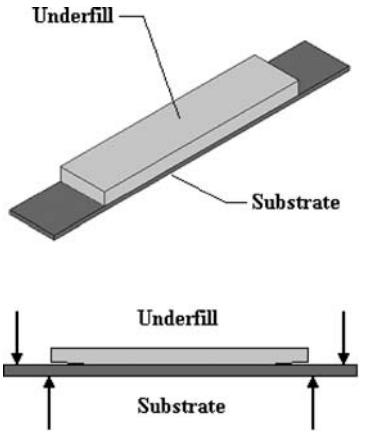


FIGURE 17.14. Interfacial fracture toughness test specimen.

moisture content can be identified. For additional information on psychometrics, refer to Thermodynamics: An Engineering Approach by Cengel and Boles [6].

All test specimens were baked at 115°C for at least 12 hours to remove any moisture that may have been introduced during sample preparation prior to environmental aging, which was performed in a humidity chamber in an atmosphere maintained at a constant temperature (±1°C), humidity (±1%), and pressure (P_{atm}). All interface fracture tests were performed with both the surrounding environment and test specimens being at room temperature after environmental preconditioning. No measurable loss in moisture uptake occurred in the test specimens from the time they were removed from the environmental chamber, allowed to cool to room temperature, and experimentally tested.

Using the experimentally measured value for the critical load of fracture in conjunction with previously identified elastic modulus results, the interfacial fracture toughness of the underfill/copper test specimens was determined using Equation (17.10) for each particular level of moisture preconditioning. Figure 17.15 provides a graphical depiction of the results depicting the effect of environmental preconditioning on the underfill/copper interfacial fracture toughness.

The entire range of mode mixity for all interfacial test specimens fell between −37.41° to −37.64°. The substrate height was used to define the characteristic length for all reported toughness values when evaluating the mode mixity. Since the variation in mode mixity was negligible, the effect of this variation affecting interfacial fracture toughness results between different test groups is insignificant. Consequently, interfacial fracture toughness results for different moisture preconditioned test groups can be compared to one another to ascertain the effect of increasing moisture content on toughness values. In addition, saturation was reached in each moisture preconditioning environment prior to fracture testing. As a result, a gradient of moisture concentration did not exist in the interfacial fracture toughness test specimens during testing. As shown in Figure 17.15, it is clear that the contribution of thermal aging at 85°C did not significantly affect the interfacial fracture toughness of the underfill/copper interface. It is important to remember that all tests were performed at room temperature, hence only the effects of thermal aging were evaluated rather than the effect of testing at elevated temperatures. Since all environmental preconditioned test groups were exposed to the same temperature component of 85°C and duration of 168 hours, any observed changes in the fracture toughness after

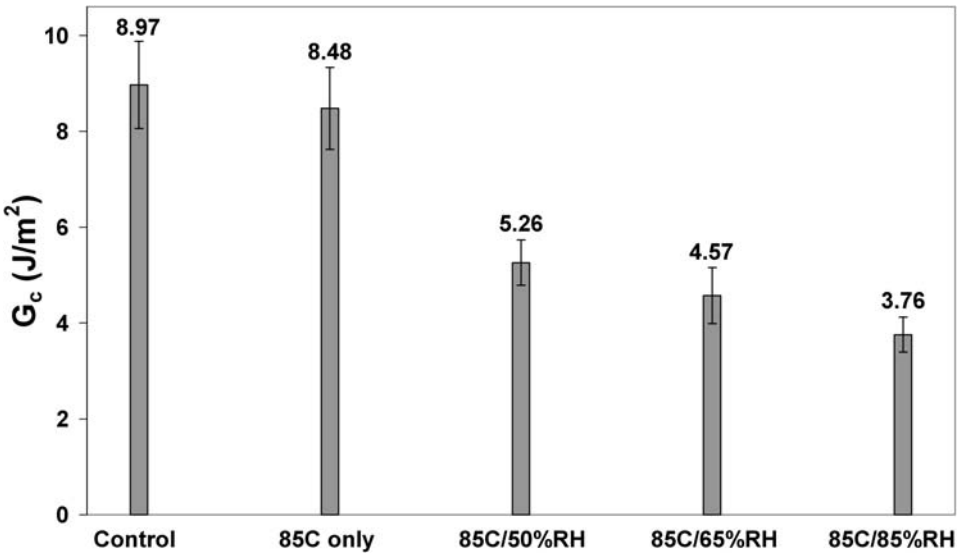


FIGURE 17.15. Effect of environmental preconditioning on the interfacial fracture toughness of the underfill/copper interface.

TABLE 17.3.
Change in the underfill/copper test specimen interfacial fracture toughness from moisture uptake.

T (°C)	RH (%)	C_{sat} (wt%)	C_{sat} (mg H_2O/mm^3)	G_c (J/m^2)	Toughness change (%)
Control	—	0	0.0000	8.97 ± 0.91	—
85	50	0.65	0.0075	5.26 ± 0.47	41.4
85	65	0.77	0.0089	4.57 ± 0.58	49.1
85	85	1.02	0.0118	3.76 ± 0.36	58.1

moisture preconditioning can be attributed to the contribution of moisture. Moisture preconditioning at 85°C/50%RH, 85°C/65%RH, and 85°C/85%RH had a substantial effect on the interfacial fracture toughness and yielded decreases of 41.4%, 49.1%, and 58.1% respectively. A summary of the effect of moisture preconditioning on the interfacial fracture toughness is provided in Table 17.3, where C_{sat} represents the saturation concentration of moisture for each respective level of moisture preconditioning and given as a percent weight change (wt%).

Figures 17.16 and 17.17 depict the inherent change in the underfill/copper interfacial fracture toughness as a function of moisture concentration.

Based on Figures 17.16 and 17.17, it is clear that the change in the interfacial fracture toughness is sensitive to small amounts of moisture. A significant reduction in interfacial adhesion was observed for concentrations as low as 0.65 wt%. Since the moisture did not significantly alter the elastic modulus of the underfill adhesive for the moisture conditions evaluated for the interfacial fracture toughness, plasticization of the underfill from moisture contributed little to the change in the interfacial fracture toughness. As a result, the reduction in toughness is primarily attributed to the weakening of the underfill/copper interface due to the direct presence of moisture at the interface. The moisture at the interface could decrease the adhesion through displacement of the underfill reducing Van der Waals forces

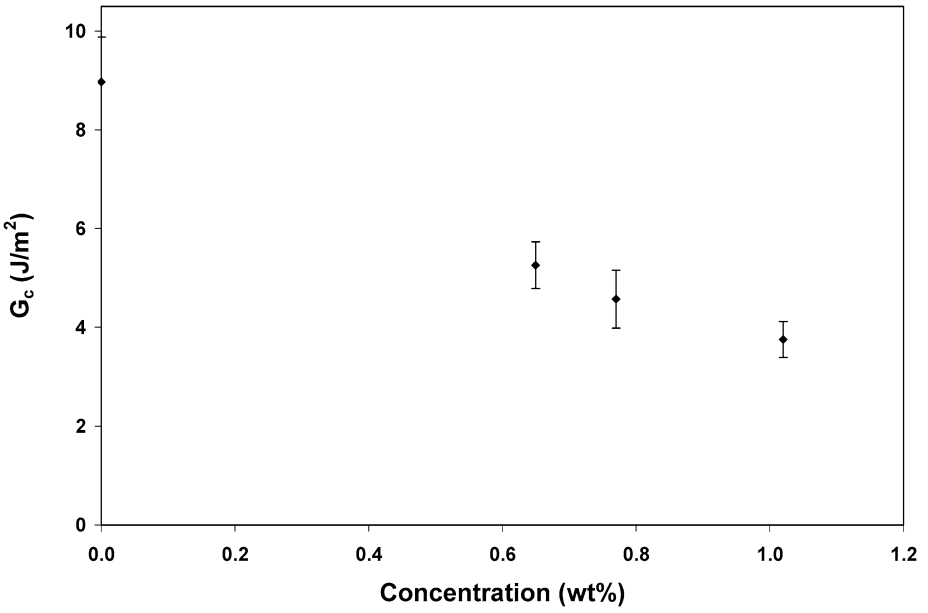


FIGURE 17.16. Underfill/copper interfacial fracture toughness variation as a function of moisture concentration (wt%).

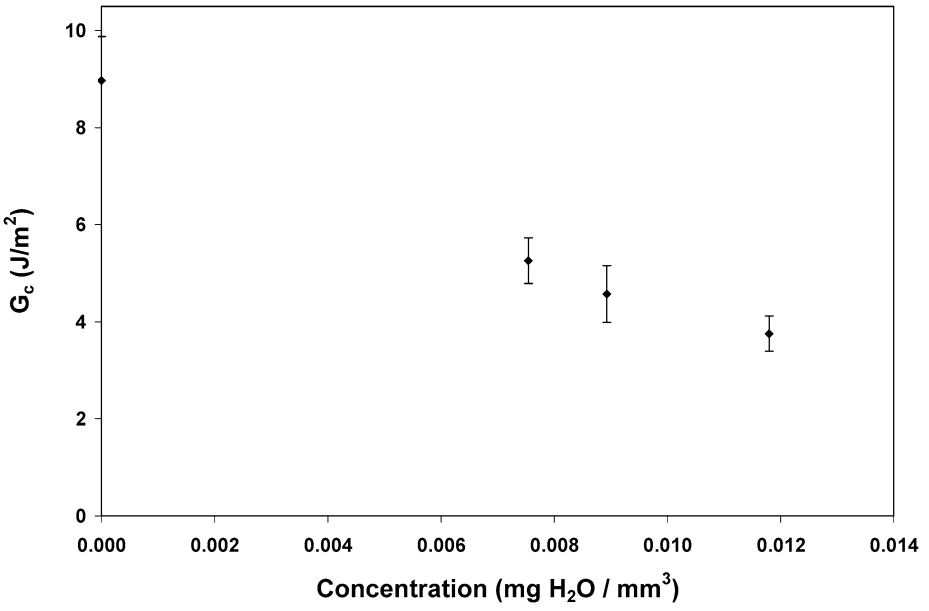


FIGURE 17.17. Underfill/copper interfacial fracture toughness variation as a function of moisture concentration (mg H_2O/mm^3).

as well as possible chemical degradation of adhesive bonds. Further investigations into the exact failure mechanism from moisture at the interface are provided in detail in subsequent sections of this chapter.

17.4.3.1. Moisture Induced Swelling In addition to the mechanical load applied to test specimens during interfacial fracture testing, the interface is also subjected to hygro-swelling and thermal contraction mismatch effects between the adhesive and substrate. These two effects have opposite outcomes on the interface, as the contribution from the hygro-swelling mismatch will cause the underfill to be in compression, while the contribution from the thermal contraction mismatch will cause the underfill to be in tension. This is attributed to the different stress free environments for each case. For the case of the hygro-swelling mismatch, fully dry conditions represent a stress-free state for the interface. As moisture is absorbed in the underfill, it will cause the underfill to expand, while the moisture impermeable substrate will retain its original dimensions. Since the moisture expansion in the underfill will be constrained by the substrate, the expansion in the underfill will yield compressive stresses within the underfill. For the case of the thermal contraction mismatch, the curing temperature of the underfill represents a stress-free state for the interface. Once test specimens are removed from the oven and allowed to cool to room temperature, the thermal mismatch between the copper and the underfill will cause the underfill to be in tension due to it wanting to shrink more than the copper substrate (CTE of experimental materials: underfill = 75 ppm/°C, copper = 17 ppm/°C). Whether the interface is dominated by the hygro-swelling mismatch, thermal contraction mismatch, or possibly neither due to the effects of one another canceling each other out for a particular moisture saturation level will depend on the characteristics of the materials that constitute each bimaterial interface relative to their moisture preconditioning environment.

To investigate the effect of hygro-swelling on interfacial fracture test results, the moisture swelling coefficient, β, of the underfill was experimentally determined for each moisture preconditioning environment. The moisture swelling coefficient is defined as

$$\beta = \frac{\Delta\ell/\ell_o}{C_{sat}}, \tag{17.16}$$

where $\Delta\ell$ is the change in length of the specimen due to moisture absorption, ℓ_o is the initial dry length of the specimen, and C_{sat} is the saturation moisture concentration. Using Equation (17.16) with experimental test data, the moisture swelling coefficient was determined for conditions of 85°C/50%RH ($\beta = 1987$ ppm/wt%), 85°C/65%RH ($\beta = 1907$ ppm/wt%), 85°C/85%RH ($\beta = 1808$ ppm/wt%). Having identified the moisture swelling coefficient for each moisture preconditioning environment, a comparison can be made between the hygro-swelling and thermal mismatch strains for the underfill/copper interface. The hygro-swelling mismatch strain, ε_h, and thermal mismatch strain, ε_t, are defined as follows:

$$\varepsilon_h = \beta_1 C_{sat,1} - \beta_2 C_{sat,2}, \tag{17.17}$$

$$\varepsilon_t = (\alpha_1 - \alpha_2)(T_f - T_i), \tag{17.18}$$

where β is the moisture swelling coefficient, C_{sat} is the equilibrium moisture saturation concentration, α is the coefficient of thermal expansion, T is the temperature, and subscripts 1 and 2 refer to the two materials that constitute the bimaterial interface. The hygro-swelling mismatch strain and thermal expansion mismatch strain were calculated using Equations (17.17) and (17.18) respectively for each moisture preconditioning environment. Since the cooling of the interfacial fracture test specimens from the cure temperature to

TABLE 17.4.
Comparison of hygro-swelling and thermal mismatch strains for the underfill/copper interfacial fracture test specimens.

Environment	β (ppm/wt%)	C_{sat} (wt%)	ε_h	α_{uf} (ppm/°C)	α_{Cu} (ppm/°C)	T_i (°C)	T_f (°C)	ε_t
85°C/50%RH	1987	0.65	0.0013	75	17	190	25	0.0096
85°C/65%RH	1907	0.77	0.0015	75	17	190	25	0.0096
85°C/85%RH	1808	1.02	0.0018	75	17	190	25	0.0096

room temperature will result in a thermal contraction, while the uptake of moisture will result in an expansion from swelling, it should be noted that the hygro-swelling and thermal expansion mismatch strains act in opposite directions. The results are given in Table 17.4.

As shown in Table 17.4, the thermal mismatch strains were significantly greater than the hygro-swelling mismatch strains for all moisture preconditioning environments by roughly an order of magnitude. It is clear that the thermal mismatch strain dominated the interaction at the interface and was only slightly offset by a small contribution from the hygro-swelling mismatch strain for this particular bimaterial interface. As a result, the underfill will be in tension during interfacial fracture testing, effectively preloading the interface and requiring a lower critical load of fracture, P_c, from mechanical testing to advance the interface crack. Consequently, interfacial fracture toughness values will represent a conservative estimate of the interfacial fracture toughness of the interface. In addition, it is clear that increasing the saturation concentration did not significantly increase the hygro-swelling mismatch strain. All interfaces for all environments experienced similar hygro-swelling mismatch strains for the materials and moisture preconditioning environments tested in this study. Consequently, the trends exhibited in the interfacial fracture toughness as moisture concentration increases are essentially independent of the hygro-swelling mismatch relative to one another, and the observed changes between the different moisture preconditioning environments can be predominately attributed to more moisture being present at the interface resulting in a greater loss of adhesion.

17.4.3.2. Interfacial Hydrophobicity The polarity of the water molecule will affect its behavior at the interface, which can influence the extent of environmental degradation of an adhesive joint due to the presence of moisture [33]. The polar behavior of water arises from its structure, which is composed of a single oxygen atom bonded to two hydrogen atoms. The hydrogen atoms are covalently bonded to the oxygen atom through shared electrons. Two pairs of electrons surrounding the oxygen atom are involved in covalent bonds with hydrogen; however, there are also two unshared pairs of electrons (lone-pair) on the other side of the oxygen atom, which shift the electron cloud of the water molecule over to the oxygen atom as shown in Figure 17.18.

This uneven distribution of electron density in the water molecule yields a partial negative charge (δ^-) on the oxygen atom and a partial positive charge (δ^+) on the hydrogen atoms, giving rise to the polarity of the water molecule. Polarity allows water molecules to bond with each other, and hydrogen bonds will form between two oppositely charged ends of a water molecule as shown in Figure 17.19.

The hydrogen bonds have about a tenth of the strength of an average covalent bond, and are being constantly broken and reformed in liquid water. The polarity will also allow water to molecules to bond with other polar molecules, which will affect how the water will wet on different surfaces. Surfaces that contain polar molecules are hydrophilic. They

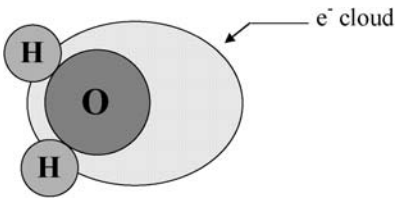


FIGURE 17.18. Electron cloud distribution on a water molecule.

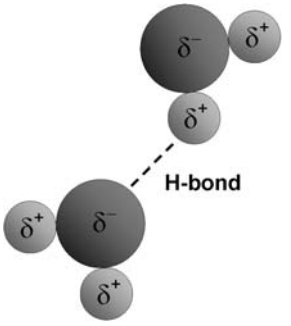


FIGURE 17.19. Hydrogen bonding between water molecules.

interact with the water molecules to enhance wetting, causing the water to smear flat. If a surface contains alcohols, O, or N, it will probably be hydrophilic. Conversely, surfaces that contain nonpolar substances are hydrophobic. They cannot interact with the water molecules, causing it to form a bubble on the surface. In general, if a surface contains C, H, or F, it will probably be hydrophobic.

Most materials will not be purely hydrophobic or hydrophilic, but will have varying degrees to which they are considered one or the other. This is addressed in Hydrophobicity, which is the study of the wetting characteristics of water on surfaces. One method used to test the hydrophobicity of a surface is through measurement of the contact angle, θ, using water as the probe liquid. The contact angle represents a balance between the adhesive forces between the liquid and solid and cohesive forces in the liquid. The adhesive forces cause the liquid drop to spread, while the cohesive forces cause the liquid drop to retain the shape of a sphere. The contact angle is a direct measure of wettability and provides an effective means to evaluate many surface properties such as surface contamination, surface hydrophobicity, surface energetics, and surface heterogeneity. When $\theta > 0$, the liquid is nonspreading and reaches an equilibrium position between the liquid-fluid and solid-liquid interfaces. When $\theta = 0$ the liquid wets without limit and spontaneously spreads freely over the surface. Hydrophobic surfaces repel water and produce high contact angles. Hydrophilic surfaces attract water and produce low contact angles. Figure 17.20 illustrates the contact angle behavior of water on both hydrophobic and hydrophilic surfaces.

By utilizing water as the probe liquid, the interfacial hydrophobicity can be ascertained by measuring the water contact angle of both the adhesive and substrate. To determine the hydrophobicity of interfacial fracture test specimens, contact angle measurements were made for the adhesive and substrate evaluated in this study. Both the clean copper substrate and underfill adhesive exhibited fairly hydrophobic behavior with contact angles of 74° and 83° respectively. Having established the hydrophobicity of the substrate

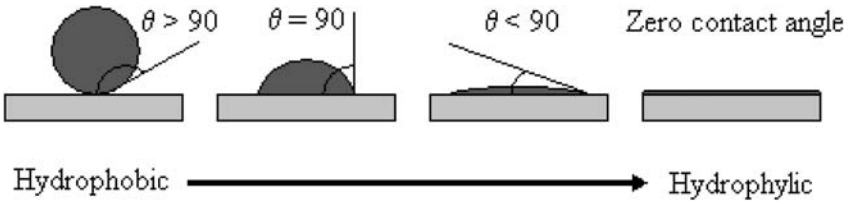


FIGURE 17.20. Hydrophobic and hydrophilic water contact angle behavior.

and adhesive, the interfacial hydrophobicity of the underfill/copper interfacial fracture test specimens can be evaluated. When addressing the relative hydrophobicity of the substrate and adhesive to moisture behavior at the interface, the interaction can become complex. The surface with the most dominant degree of hydrophobicity will govern the shape and response of the water at the interface. For example, if a hydrophobic substrate is bonded with a hydrophilic adhesive, then the water at the interface will want to minimize contact with the substrate and maximize contact with the adhesive. Depending on imperfections in the bonding, surface roughness, and the relative degree of hydrophobicity of the substrate to the adhesive, water at the interface will more or less form a somewhat hemi-spherical shape at the interface, with the spherical end minimizing contact on the substrate and the open end maximizing contact on the adhesive. Naturally, the shape of the water at the interface can have various permutations of the aforementioned shape depending on the degree of hydrophobic behavior of the substrate relative to the hydrophilic behavior of the adhesive, but the general idea remains the same. For other systems with varying degrees of hydrophobicity, the shape of the water at the interface relative to the hydrophobicity of the substrate and adhesive can be extremely difficult to characterize; however, qualitative conclusions can be made. For the case of the underfill/copper interfacial fracture test specimens, the relative hydrophobicity of the adhesive to the substrate was similar; consequently, the wetting behavior of the moisture at the interface would not be significantly dominated by either the adhesive or substrate.

An additional consideration unique to environmental preconditioning is the growth of oxides affecting the interfacial hydrophobicity. Copper has a strong affinity to oxygen, and the development of an oxidation layer between the substrate and adhesive after bonding is inevitable. Initially, cuprous oxide, Cu_2O, will form followed by the formation of a layer of cupric oxide, CuO. The oxidation of copper substrates can be significant, and previous studies have shown that the water contact angle on copper is affected by oxidation [7,25, 28,61]. Due to oxidation growth on the copper substrates, contact angle measurements were made for each preconditioning environment to monitor any change in the hydrophobicity of the copper surface.

Since the copper bonding surface of the interfacial fracture test specimen will be shielded by the underfill adhesive, the oxidation growth rate will be different than for bare copper environmentally aged for a similar duration of time. Therefore, water contact angles for each environmental test group were measured using special test specimens that mimicked the exposure of the copper bonding surface to similar amounts of oxygen and moisture as the interfacial fracture test specimens. These specimens used the same geometry as the interfacial fracture test specimens, but the underfill adhesive was cured separately in an individual mold. After curing the adhesive, the underfill was placed on top of the copper substrate and held in place by c-clamps. Similar to the interfacial fracture test specimens, a water-proof sealant was applied around the perimeter of the test specimen to eliminate

wicking of moisture at the interface and force 1D diffusion through the top surface of the underfill. After environmental preconditioning, the water-proof perimeter, c-clamps, and underfill were removed from the test specimen for contact angle measurement of the copper surface.

Experimentally measured water contact angle results were as follows: 76° for 85°C thermal aging, 76° for 85°C/50%RH moisture preconditioning, 77° for 85°C/65%RH moisture preconditioning, and 77° for 85°C/85%RH moisture preconditioning. All test groups were preconditioning for the same duration of 168 hours, which was the same criteria used in the evaluation of the effect of moisture on interfacial adhesion. Based on these results, it is evident that all levels of environmental preconditioning did not significantly alter the water contact angle and associated hydrophobicity of the interface. As a result, similar interfacial wetting characteristics of moisture at the interface will occur for all preconditioning environments.

Although the contact angle did not significantly change, there did appear to be a slight increase in the water contact angle with moisture preconditioning. Previous studies have shown both an increase [28,61] and decrease [7,25] in the water contact angle of copper with oxidation. The oxidation–reduction chemistry occurring at the interface relative to environmental preconditioning is complex, and the differences in trends could be attributed to the degree of oxidation altering the surface chemistry [7], change in surface roughness of the substrate from oxidation growth [25], and contamination of the surface by hydrocarbons from the environment [33]. In addition, Yi et al. [61] has provided data correlating the oxide layer thickness on copper leadframes to water contact angles. These data shows a slow, gradual increase in oxide thickness from water contact angles ranging from 72°–78°, but depicts a sharp increase in oxide layer thickness for contact angles exceeding 80°. Based on results for the water contact angle on copper in this study, all measurements yielded average contact angles less than 78° with vary little variation with each other. This indicates a similar level of interfacial hydrophobicity and oxide layer thickness for all environmentally preconditioned test groups. Both Mino et al. [35] and Chong et al. [8] have shown that the development of the copper oxide layer thickness is significantly slower and minimal for temperatures below 100°C and 120°C. Since the test specimens in this study had a temperature component of only 85°C, it is anticipated that the oxide layer thickness that developed on test specimens would have a minimal effect on toughness results. This is also supported by X-ray Photoelectron Spectroscopy (XPS) results. XPS showed the presence of cupric oxide not only in the 85°C/50%RH, 85°C/65%RH, and 85°C/85%RH test groups, but also in the 85°C thermal aging test group. As a result, identical oxide chemical formations existed at the interface for all environmentally preconditioned test groups. In addition, similar atomic percentages of cupric oxide were obtained when comparing thermal aging at 85°C to the moisture preconditioning environments of 85°C/50%RH, 85°C/65%RH, and 85°C/85%RH, indicating that the moisture component had a minimal contribution to oxidation growth rates on the copper compared to the available oxygen in the air common to all environmental preconditioned environments. Consequently, a similar level of oxidation thickness existed on all environmentally preconditioned test specimens, which supports the results from the water contact angle measurements.

Since oxides were removed from the copper surface before adhesive bonding and the flux present in the no-flow underfill would have removed any oxides that developed during adhesive curing, it is possible that the oxidation growth from environmental preconditioning would have an effect on the interfacial fracture toughness results. This oxide growth could displace the underfill from the copper substrate after bonding to contribute to the observed loss in adhesion after moisture preconditioning shown in Figure 17.15. Since both

water contact angle measurements and XPS results demonstrate a similar oxidation thickness existed on all environmentally preconditioned test specimens, the 85°C thermal aging results can be compared to the control test results to ascertain the effect of oxidation growth on the loss in adhesion without the contribution from moisture. As shown in Figure 17.15, thermal aging at 85°C produced little to no effect on interfacial fracture toughness results, thus oxidation growth displacing the underfill after adhesive bonding had an insignificant effect on the adhesion loss compared to the effect of moisture from moisture preconditioning.

17.4.4. Interfacial Fracture Toughness Recovery from Moisture Uptake

The underfill/copper interface was found to be very sensitive to moisture, with large decreases in interfacial fracture toughness occurring for moisture preconditioning environments of 85°C/50%RH, 85°C/65%RH, and 85°C/85%RH (Figure 17.15). To further investigate the reversible and irreversible nature of moisture on the interfacial adhesion of the underfill/copper interface, additional test specimens were moisture preconditioned for each condition for 168 hours followed by baking at 95°C until fully dry. A fully dried state was established when there was no measurable change in the weight of a specimen for a period of 24 hours. Upon reaching a dry state, specimens were fracture tested to ascertain the interfacial fracture toughness. The entire range of mode mixity for all interfacial test specimens fell between $-37.43°$ to $-37.48°$. The substrate height was used to define the characteristic length for all reported toughness values when evaluating the mode mixity. Since the variation in mode mixity was negligible, the effect of this variation influencing interfacial fracture toughness results between different test groups is insignificant. Consequently, toughness recovery results for different moisture preconditioned test groups can be compared to one another to ascertain the effect of increasing moisture content on toughness values. Figure 17.21 provides a graphical depiction of the effect of environmental preconditioning and recovery of the underfill/copper interfacial fracture toughness.

As shown in Figure 17.21, most of the loss in interfacial fracture toughness from moisture was not recovered upon fully drying. Since the small change in the underfill elastic modulus from moisture was recoverable upon fully drying, the permanent reduction in the toughness of the underfill/copper interface is attributed to the direct presence of moisture at the interface debonding the underfill adhesive to the copper substrate. Similar in form to the recoverability of the elastic modulus given by Equation (17.9), the recoverability for the interfacial fracture toughness will be defined as follows:

$$\textit{Recoverability}\ (\%) = \frac{G_{c,recovery} - G_{c,sat}}{G_{c,dry} - G_{c,sat}} \cdot 100, \tag{17.19}$$

where $G_{c,recovery}$ is value of the interfacial fracture toughness upon fully drying from the moisture saturated state, $G_{c,sat}$ is the saturated value of the interfacial fracture toughness after moisture absorption, and $G_{c,dry}$ is the unaged, control value of the interfacial fracture toughness. Equation (17.19) only applies when the mode mixity of the interfacial fracture toughness before and after moisture preconditioning remains relatively unchanged, otherwise changes in the toughness due to a contribution from a change in the mode mixity will introduce error in the recoverability results. The recoverability of the underfill/copper interfacial fracture toughness is given in Table 17.5.

As shown by Table 17.5, the irreversible damage on interfacial fracture toughness from exposure to moisture was substantial for the underfill/copper interface. Very little

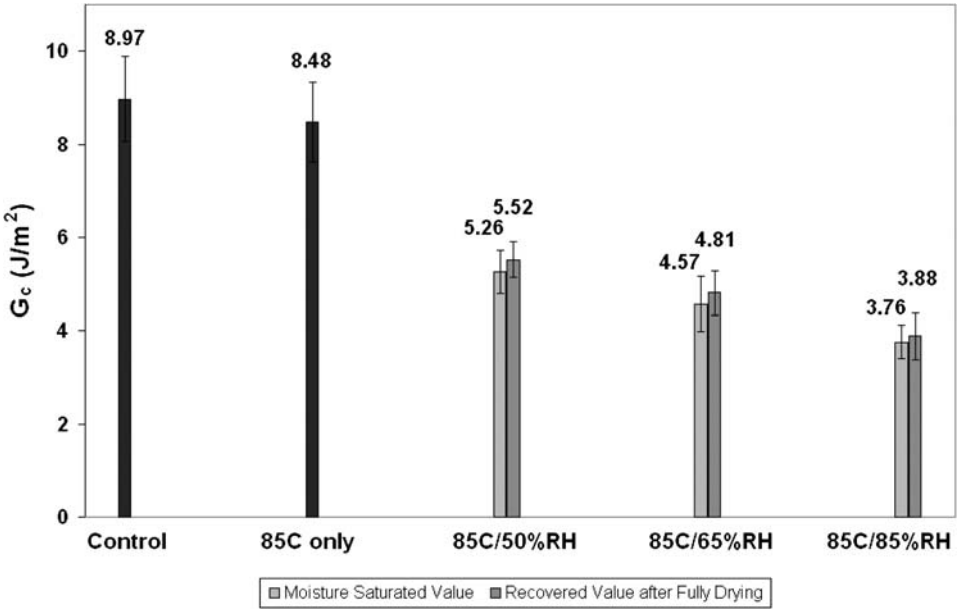


FIGURE 17.21. Recovery of the underfill/copper interfacial fracture toughness on removal of moisture.

TABLE 17.5.
Recoverability of the underfill/copper interfacial fracture toughness from moisture uptake after subsequent drying.

T (°C)	RH (%)	C_{sat} (wt%)	$G_{c,sat}$ (J/m^2)	$G_{c,recovery}$ (J/m^2)	*Recoverability* (%)
Control	—	0.00	8.97 ± 0.91	—	—
85	50	0.65	5.26 ± 0.47	5.52 ± 0.38	7.0
85	65	0.77	4.57 ± 0.58	4.81 ± 0.47	5.5
85	85	1.02	3.76 ± 0.36	3.88 ± 0.50	2.3

of the underfill/copper interfacial fracture toughness was recoverable after fully drying, with recoverability values for all moisture preconditioning environments less than 7%. It is also evident that a relatively small amount of moisture reaching the interface causes the structural integrity of the adhesive bond to be noticeably, permanently compromised.

17.4.5. Interfacial Fracture Toughness Moisture Degradation Model

Having implemented an extensive experimental program to ascertain the role of moisture in adhesion degradation and the physical mechanisms responsible for the change in interfacial adhesion, the focus now shifts to developing a model depicting the intrinsic loss in interfacial fracture toughness as a function of the critical parameters relevant to moisture. At the root of this model is characterizing the dominant mechanism for adhesion between the adhesive and substrate. There are four primary mechanisms for adhesion which have been proposed. They include mechanical interlocking, diffusion theory, electronic theory, and adsorption theory [30]. For the underfill/copper interface, the contributions of interfacial diffusion and electrostatic forces between the adhesive and substrate causing

adhesion is far lower than the effects of mechanical interlocking and adsorption. Since the copper substrates in this study were polished to a mirror finish, the effects from mechanical interlocking of the adhesive into irregularities present on the substrate surface will be small compared to the effects from intermolecular secondary forces (i.e., Van der Waals) between the atoms and molecules in the surfaces of the adhesive and substrate. Consequently, adsorption theory will dominate the adhesive bonding at the underfill/copper interface of our test specimens.

Provided adsorption theory governs adhesion and only secondary forces are acting across an interface, the stability of an adhesive/substrate interface in the presence of moisture can be ascertained from thermodynamic arguments. The thermodynamic work of adhesion, W_A, in an inert medium is given by [30]:

$$W_A = \gamma_a + \gamma_s - \gamma_{as}, \tag{17.20}$$

where γ_a is the surface free energy of the adhesive, γ_s is the surface free energy of the substrate, and γ_{as} is the interfacial free energy. In the presence of a liquid, the thermodynamic work of adhesion, W_{Al}, is given by:

$$W_{Al} = \gamma_{al} + \gamma_{sl} - \gamma_{as}, \tag{17.21}$$

where γ_{al} and γ_{sl} are the interfacial free energies between the adhesive/liquid and substrate/liquid interfaces, respectively. Typically the thermodynamic work of adhesion of an adhesive/substrate interface in an inert medium, W_A, is positive, which indicates the amount of energy required to separate a unit area of the interface. However, the thermodynamic work of adhesion in the presence of a liquid, W_{Al}, can be negative, which indicates the interface is unstable and will separate when it comes in contact with the liquid. Thus, the calculation of W_A and W_{Al} can indicate the environmental stability of the adhesive/substrate interface. Kinloch [30] has shown that W_A and W_{Al} may be calculated from the following expressions:

$$W_A = 2\sqrt{\gamma_a^D \gamma_s^D} + 2\sqrt{\gamma_a^P \gamma_s^P}, \tag{17.22}$$

$$W_{Al} = 2(\gamma_{lv} - \sqrt{\gamma_a^D \gamma_{lv}^D} - \sqrt{\gamma_a^P \gamma_{lv}^P} - \sqrt{\gamma_s^D \gamma_{lv}^D} - \sqrt{\gamma_s^P \gamma_{lv}^P} + \sqrt{\gamma_a^D \gamma_s^D} + \sqrt{\gamma_a^P \gamma_s^P}), \tag{17.23}$$

where γ^D is the dispersion component of surface free energy, γ^P is the polar component of surface free energy, and γ_{lv} is the surface free energy of the liquid. Table 17.6 gives the polar and dispersion surface free energies of epoxy, copper, and water.

TABLE 17.6.
Polar and dispersion surface free energies of epoxy, copper, and water [30].

Substance	γ (mJ/m^2)	γ^D (mJ/m^2)	γ^P (mJ/m^2)
Epoxy	46.2	41.2	5.0
Copper	1360	60	1300
Water	72.2	22.0	50.2

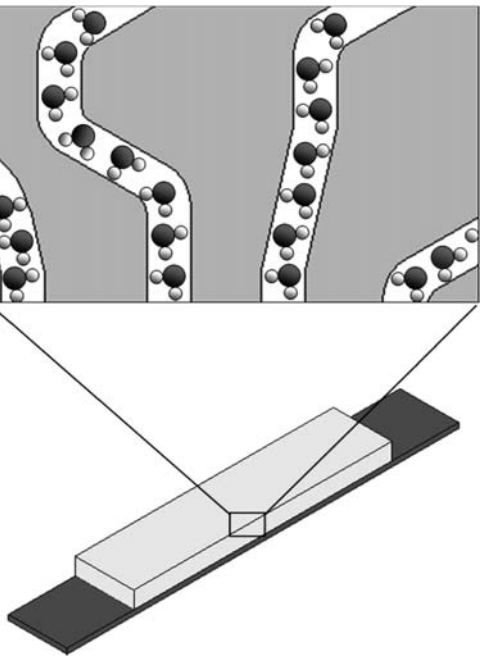


FIGURE 17.22. Moisture transport through the bulk epoxy of an interfacial fracture test specimen.

Using the values given in Table 17.6 and substituting into Equation (17.22), the thermodynamic work of adhesion of the epoxy/copper interface is 260.7 mJ/m^2. If water is present at the epoxy/copper interface, the thermodynamic work of adhesion given by Equation (17.23) is -270.4 mJ/m^2. Therefore, since the work of adhesion is positive before exposure to moisture and negative after exposure, all adhesion of the epoxy/copper interface is lost if water comes in contact with the interface. This is supported by the recovery interfacial fracture toughness results presented in Section 17.4.4, where virtually none of the observed loss in adhesion from moisture exposure was recovered upon fully drying.

Using adsorption theory as the physical basis for the loss in adhesion from moisture, expressions are now developed depicting the amount of moisture delivered to the underfill/copper interface. Since the interfacial fracture test specimens were designed to prevent wicking of moisture at the interface and the copper substrate provides a barrier for moisture transport, the moisture transport to the interface is governed by the epoxy network of the underfill. Soles and Yee [48] have shown that water traverses within the epoxy through the network of nanopores inherent in the epoxy structure. A typical nanopore ranges from 5.0 to 6.1 Å in diameter. Figure 17.22 illustrates the transport of moisture through the bulk epoxy of an interfacial fracture test specimen.

Assuming that the nanopore channels are the only mechanism by which moisture can be delivered to the interface, the saturation concentration in the epoxy expressed in mg H_2O/mm^3 is given by:

$$C_{sat} = \frac{\rho(NV)}{V_{tot}}, \tag{17.24}$$

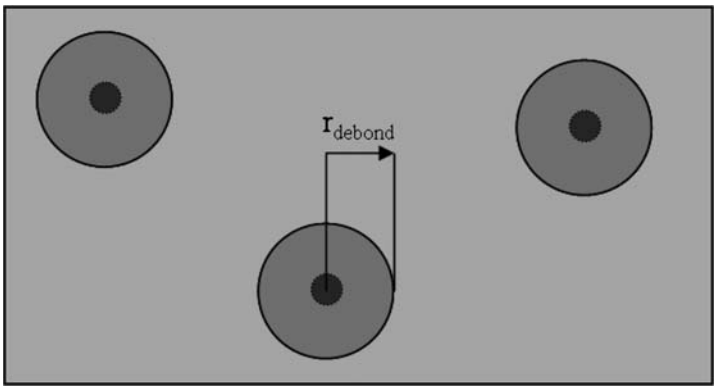


FIGURE 17.23. Graphical illustration of the parameter, r_{debond}, at the interface.

where ρ is the density of water measured in milligram per cubic millimeter (mg/mm^3), N is the number of nanopores actively participating within the epoxy network, V is the volume occupied by a single nanopore in the epoxy network, and V_{tot} is the total volume of the epoxy. After rearrangement of Equation (17.24), the number of nanopores actively participating within an epoxy system for a given saturation concentration is as follows:

$$N = \frac{4AC_{sat}}{\pi \rho D^2}, \tag{17.25}$$

where A is the total area of the interface and D is the nanopore diameter. Assuming adsorption theory holds, the adhesive bond area, A_{bond}, that remains intact after exposure to moisture will depend on the area occupied by the moisture at the interface, A_{H_2O}:

$$A_{bond} = A - A_{H_2O}. \tag{17.26}$$

Relating this adhesive bond area to the number of nanopores actively participating in transport yields:

$$A_{bond} = A - \pi N r_{debond}^2, \tag{17.27}$$

where r_{debond} represents the debond radius of moisture at the interface that occurs at each nanopore. The debond radius must be greater or equal to the nanopore radius and is governed by the interfacial hydrophobicity of the adhesive/substrate interface. Figure 17.23 provides a graphical depiction of the parameter, r_{debond}, at the interface.

Substituting Equation (17.25) into (17.27) provides an expression for the adhesive bond area that remains intact after exposure to a particular moisture saturation concentration:

$$A_{bond} = A - \frac{4AC_{sat} r_{debond}^2}{\rho D^2}. \tag{17.28}$$

We now want to employ a fracture mechanics development to relate the change in bond area due to the presence of moisture at the interface. Recall from fracture mechanics the general form of the stress intensity factor:

$$K = S\sigma\sqrt{\pi a}, \tag{17.29}$$

where S is a dimensionless constant that depends on the geometry and mode of loading, σ is the remotely applied stress, an a is the crack length. The stress intensity factor is related to the fracture toughness, G_c, by the following expression:

$$G_c = Z\sigma^2, \tag{17.30}$$

where

$$Z = \frac{\pi a S^2(1-\upsilon^2)}{E}.$$

Based on the thermodynamic work of adhesion for the epoxy/copper interface, the interface will become unstable and debond in the presence of moisture; however, since interfacial fracture toughness is a material property that characterizes the adhesion of the interface, the toughness must be the same in all areas that remain bonded after exposure to moisture. Using mode I loading and making the following three assumptions: (1) Adsorption theory dominates the interfacial bonding; (2) The change in the mechanical properties of both the adhesive and substrate from moisture is small relative to the change in bond area from moisture, and (3) The relative change in fracture toughness from moisture remains constant irrespective to the means of measuring the toughness for a given moisture saturation concentration, an expression is obtained relating the change in bond area due to the presence of moisture to the change in the critical load of fracture:

$$\frac{P_{wet}}{A - \pi N r_{debond}^2} = \frac{P_{dry}}{A}. \tag{17.31}$$

Rearranging Equation (17.31) to obtain an expression for P_{wet} and substituting that value into Equation (17.30) for the wet, saturated case yields the following expression:

$$G_{c,wet} = \left(1 - \frac{\pi N r_{debond}^2}{A}\right)^2 G_{c,dry}. \tag{17.32}$$

As the saturation moisture concentration increases, so will the number of active nanopores participating. The incremental change in fracture toughness due to the participation of a single additional nanopore, $N+1$, is given by:

$$G_{c,wet} = \left(1 - \frac{\pi (N+1) r_{debond}^2}{A}\right)^2 G_{c,dry}. \tag{17.33}$$

For convenience, define f such that for N nanopores participating:

$$f_N = \frac{\pi N r_{debond}^2}{A}. \tag{17.34}$$

For $N+1$ nanopores participating:

$$f_{N+1} = \frac{\pi r_{debond}^2}{A}(N+1). \tag{17.35}$$

Restating Equations (17.32) and (17.33) in terms of f:

$$G_{c,wet}(f_N) = \left(1 - \frac{\pi N r_{debond}^2}{A}\right)^2 G_{c,dry}, \tag{17.36}$$

$$G_{c,wet}(f_{N+1}) = \left(1 - \frac{\pi r_{debond}^2}{A}\right)^2 G_{c,wet}(f_{N_N}). \tag{17.37}$$

Subtracting (17.36) from (17.37) and dividing by $f_{N+1} - f_N$ gives:

$$\frac{G_{c,wet}(f_{N+1}) - G_{c,wet}(f_N)}{f_{N+1} - f_N} = \frac{\left[1 - \left(\pi r_{debond}^2/A\right)\right]^2 G_{c,wet}(f_N) - G_{c,wet}(f_N)}{f_{N+1} - f_N}. \tag{17.38}$$

Utilizing a Taylor series expansion of f_N with first order accuracy and substituting Equations (17.34) and (17.35) into (17.38) yields:

$$\frac{dG_{c,wet}(f_N)}{df_N} = \frac{\left[1 - \left(\pi r_{debond}^2/A\right)\right]^2 G_{c,wet}(f_N) - G_{c,wet}(f_N)}{\left(\pi r_{debond}^2/A\right)}. \tag{17.39}$$

Simplification and elimination of higher order terms gives the following differential equation characterizing the loss in interfacial fracture toughness due to moisture:

$$\frac{dG_{c,wet}(f_N)}{df_N} = -2G_{c,wet}(f_N), \tag{17.40}$$

subject to the boundary condition:

$$G_{c,wet}(f_N = 0) = G_{c,dry}. \tag{17.41}$$

Solution of Equation (17.40) gives:

$$G_{c,wet} = G_{c,dry} \exp\left[\frac{-8C_{sat} r_{debond}^2}{\rho D^2}\right]. \tag{17.42}$$

Equation (17.42) characterizes the loss in interfacial fracture toughness from moisture in terms of key parameters relevant to moisture. Using the value for the density of water at room temperature (0.998 mg/mm^3), an average nanopore diameter of 5.5 Å, and the saturation concentration determined from the experimental portion of this study in conjunction with Equations (17.25) and (17.42), the number of active nanopores participating, N, and value of r_{debond} can be determined by the intrinsic response of the material system to each level of moisture preconditioning. The results are shown in Table 17.7.

TABLE 17.7.
Key parameters relevant to moisture for the underfill/copper interface.

Environment	Substrate	Adhesive	C_{sat} (mg H_2O/mm^3)	N	r_{debond} (mm)
85°C/50%RH	Copper	Underfill	0.0075	1.006×10^{13}	1.640×10^{-6}
85°C/65%RH	Copper	Underfill	0.0089	1.194×10^{13}	1.692×10^{-6}
85°C/85%RH	Copper	Underfill	0.0118	1.583×10^{13}	1.669×10^{-6}

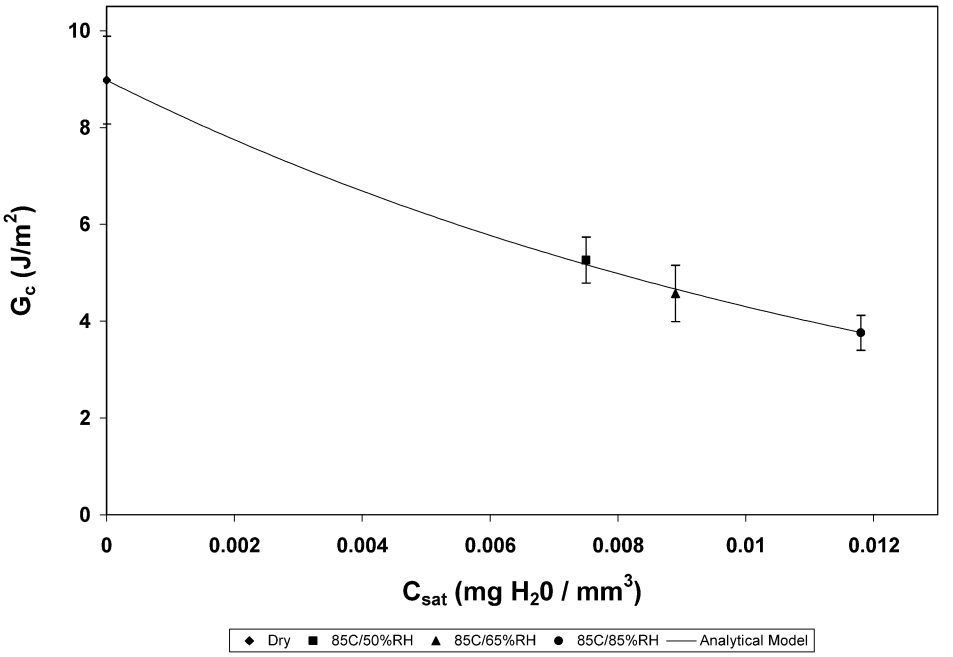


FIGURE 17.24. Analytical prediction of the loss in interfacial fracture toughness from moisture for the underfill/copper interface.

As shown in Table 17.7, the number of nanopores participating increases with saturation concentration. This is expected since an increase in saturation concentration would increase the available moisture for transport through the nanopores. In addition, the values for r_{debond} were similar for each moisture preconditioning environment for both respective interfaces, which is also expected since X-ray Photoelectron Spectroscopy and water contact angle results did not indicate a change in the interfacial hydrophobicity of the copper surface from moisture preconditioning. The slight variation in the values for r_{debond} could in part be attributed to experimental scatter. Since the results were similar, they were averaged to obtain a representative value for r_{debond} in the presence of moisture for each interface.

Using the moisture parameters identified for each interfacial material system, Equation (17.42) was used to predict the interfacial fracture toughness for the underfill/copper interface as a function of increasing saturation concentration.

As shown in Figure 17.24, Equation (17.42) accurately predicted the loss in interfacial fracture toughness as a function of increasing moisture concentration. Since Equa-

tion (17.42) was based on the physics of adsorption theory, it will yield a loss in interfacial fracture toughness provided there is moisture at the interface, no matter how small the concentration. This contradicts the results of previous studies, who have reported a critical concentration of water may exist below which there is no measurable loss in adhesion [10,23,31]. Based on the results of adsorption theory, it does not appear possible that a critical concentration of water could exist in theory. It is possible in those studies that other mechanisms for adhesion in addition to adsorption theory governed the adhesion at the interface, which could explain why a critical concentration of water was observed. An additional consideration is the method of testing used to obtain adhesion results. The aforementioned studies used lap shear test specimens to determine the interfacial strength after moisture preconditioning. Due to lacking a precrack at the interface and the applied load being distributed over the entire bonding area, these test specimens are not as sensitive to interfacial failure; consequently, possibly also explaining why in part a critical concentration of water appeared to exist for low concentrations of moisture. Conversely, interfacial fracture toughness test specimens are designed for interfacial failure through the use of a precrack at the interface, making them more sensitive to subtle changes in adhesion at the interface. The work of Wylde and Spelt [58] supports this observation. Using interfacial fracture toughness test specimens with a similar material system previously reported to exhibit a critical concentration of water from lap shear results, they found a decrease in the interfacial toughness from moisture for all concentrations of moisture, including those lower than the previously reported critical concentration of water. Consequently, provided adsorption theory dominates the adhesive bonding at the adhesive/substrate interface and the assumptions in the development of the model are satisfied, Equation (17.42) should accurately predict the loss in interfacial fracture toughness for a given moisture concentration.

REFERENCES

1. H. Ardebili, E.H. Wong, and M. Pecht, Hydroscopic swelling and sorption characteristics of epoxy molding compounds used in electronic packaging, IEEE Transactions on Components and Packaging Technologies, 26(1), pp. 206–214 (2003).
2. ASTM D790, Standard Test Methods for Flexural Properties of Unreinforced and Reinforced Plastics and Electrical Insulating Materials, Annual Book of ASTM Standards, Vol. 08.01, 1999.
3. D. Brewis, J. Comyn, A. Raval, and A. Kinloch, The effect of humidity on the durability of aluminum-epoxide joints, International Journal of Adhesion and Adhesives, 10, pp. 247–253 (1990).
4. F. Buehler and J. Seferis, Effect of reinforcement and solvent content on moisture absorption in epoxy composite materials, Composites: Part A: Applied Science and Manufacturing, 31, pp. 741–748 (2000).
5. L. Butkus, Environmental durability of adhesively bonded joints, Doctoral Thesis, Georgia Institute of Technology, Woodruff School of Mechanical Engineering, Atlanta, GA, 1997.
6. Y. Cengel and M. Boles, Thermodynamics: An Engineering Approach, McGraw-Hill, Inc., New York, 1994.
7. K. Cho and E. Cho, Effect of the microstructure of copper oxide on the adhesion behavior or epoxy/copper leadframe joints, Journal of Adhesion Science and Technology, 14(11), pp. 1333–1353 (2000).
8. C. Chong, A. Leslie, L. Beng, and C. Lee, Investigation on the effect of copper leadframe oxidation on package delamination, Proceedings of the 45th Electronic Components and Technology Conference, 1995, pp. 463–469.
9. P. Chung, M. Yuen, P. Chan, N. Ho, and D. Lam, Effect of copper oxide on the adhesion behavior of epoxy molding compound-copper interface, Proceedings of the 52nd Electronic Components and Technology Conference, 2002, pp. 1665–1670.
10. J. Comyn, C. Groves, and R. Saville, Durability in high humidity of glass-to-lead alloy joints bonded with and epoxide adhesive, International Journal of Adhesion and Adhesives, 14, pp. 15–20 (1994).
11. J. Crank, The Mathematics of Diffusion, Clarendon Press, Oxford, 1956.

12. A. Crocombe, Durability modeling concepts and tools for the cohesive environmental degradation of bonded structures, International Journal of Adhesion and Adhesives, 17, pp. 229–238 (1997).
13. X. Dai, M. Brillhart, and P. Ho, Adhesion measurement for electronic packaging applications using double cantilever beam method, IEEE Transactions on Components and Packaging Technology, 23, pp. 101–116 (2000).
14. X. Dai, M. Brillhart, M. Roesch, and P. Ho, Adhesion and toughening mechanisms at underfill interfaces for flip-chip-on-organic-substrate packaging, IEEE Transactions on Components and Packaging Technology, 23(1), pp. 117–127 (2000).
15. B. DeNeve and M. Shanahan, Effects of humidity on an epoxy adhesive, International Journal of Adhesion and Adhesives, 12, pp. 191–196 (1992).
16. H. Dodiuk, L. Drori, and J. Miller, The effect of moisture in epoxy film adhesives on their performance: I. Lap shear strength, Journal of Adhesion, 17, pp. 33–44 (1984).
17. L. Fan, C. Tison, and C. Wong, Study on underfill/solder adhesion in flip-chip encapsulation, IEEE Trans. on Advanced Packaging, 25(4), pp. 473–480 (2002).
18. T. Ferguson and J. Qu, Elastic modulus variation due to moisture absorption and permanent changes upon redrying in an epoxy based underfill, IEEE Transactions on Components and Packaging Technologies, 29(1), pp. 105–111 (2006).
19. T. Ferguson and J. Qu, Moisture and temperature effects on the reliability of interfacial adhesion of a polymer/metal interface, Proceedings of the 54th Electronic Components and Technology Conference, 2004.
20. T. Ferguson and J. Qu, Moisture absorption analysis of interfacial fracture test specimens composed of no-flow underfill materials, ASME Journal of Electronic Packaging, 125, pp. 24–30 (2003).
21. T. Ferguson and J. Qu, Effect of moisture on the interfacial adhesion of the underfill/soldermask interface, ASME Journal of Electronic Packaging, 124, pp. 106–110 (2002).
22. T. Ferguson and J. Qu, Moisture absorption in no-flow underfills and its effect on interfacial adhesion to solder mask coated FR-4 printed wiring board, International Symposium and Exhibition on Advanced Packaging Materials, Processes, Properties, and Interfaces, 2001, pp. 327–332.
23. R. Gledhill, A. Kinloch, and J. Shaw, A model for predicting joint durability, Journal of Ahdesion, 11, pp. 3–15 (1980).
24. B. Harper and V. Kenner, Effects of temperature and moisture upon the mechanical behavior of an epoxy molding compound, ASME Advances in Electronic Packaging, 1, pp. 1207–1212 (1997).
25. K. Hong, H. Imadojemu, and R. Webb, Effects of oxidation and surface roughness on contact angle, Experimental Thermal and Fluid Science, 8, pp. 279–285 (1994).
26. J. Hutchinson and Z. Suo, Mixed mode cracking in layered materials, Advances in Applied Mechanics, Vol. 29, Academic Press, New York, 1992.
27. R. Jurf and J. Vinson, Effect of moisture on the static and viscoelastic shear properties of epoxy adhesives, Journal of Materials Science, 20, pp. 2979–2989 (1985).
28. S. Kim, The role of plastic package adhesion in IC performance, Proceedings of the 41st Electronic Components and Technology Conference, 1991, pp. 750–758.
29. J.K. Kim, M. Lebbai, J. Liu, J.H. Kim, and M. Yuen, Interface adhesion between copper lead frame and epoxy molding compound: effects of surface finish, oxidation, and dimples, Proceedings of the 50th Electronic Components and Technology Conference, 2000, pp. 601–608.
30. A.J. Kinloch, Adhesion and Adhesives Science and Technology, Chapman and Hall, London, 1987.
31. A. Kinloch, Interfacial fracture mechanical aspects of adhesive bonded joints—a review, Journal of Adhesion, 10, pp. 193–219 (1979).
32. H. Lee and J. Qu, Microstructure, adhesion strength and failure path at a polymer/roughened metal interface, Journal of Adhesion and Science Technology, 17(2), pp. 195–215 (2003).
33. S. Luo, Study on adhesion of underfill materials for flip chip packaging, Doctoral Thesis, Georgia Institute of Technology, School of Textile and Fibers Engineering, Atlanta, GA, 2003.
34. S. Luo and C.P. Wong, Influence of temperature and humidity on adhesion of underfills for flip chip packaging, Proceedings of the 51st Electronic Components and Technology Conference, 2001, pp. 155–162.
35. T. Mino, K. Sawada, A. Kurosu, M. Otsuka, N. Kawamura, and H. Yoo, Development of moisture-proof thin and large QFP with copper lead frame, Proceedings of the 48th Electeonic Components and Technology Conference, 1998, pp. 1125–1131.
36. R. Morgan, J. O'Neal, and D. Fanter, The effect of moisture on the physical and mechanical integrity of epoxies, Journal of Materials Science, 15, pp. 751–764 (1980).
37. R. Morgan, J. O'Neal, and D. Miller, The structure, modes of deformation and failure, and mechanical properties of diaminodiphenyl sulphone-cured tetragylcidyl 4,4′ diaminodiphenyl methane epoxy, Journal of Materials Science, 14, pp. 109–124 (1979).

38. A. Netravali, R. Fornes, R. Gilbert, and J. Memory, Effects of water sorption at different temperatures on permanent changes in an epoxy, Journal of Applied Polymer Science, 30, pp. 1573–1578 (1985).
39. S. Orman and C. Kerr, in D.J. Alner, Ed., Aspects of Adhesion, University of London Press, London, 1971.
40. A. Pasztor, in F. Settle, Ed., Handbook of Instrumental Techniques for Analytical Chemistry, Chapter 50, Prentice Hall, New Jersey, 1997.
41. B. Prime, in E. Turi, Ed., Thermal Characterization of Polymeric Materials, Vol. 2, Chapter 10, San Diego, Academic Press, San Diego, 1997.
42. S. Rosen, Fundamental Principles of Polymeric Materials, John Wiley and Sons, New York, 1993.
43. G. Shaw, C. Rogers, and J. Payer, The effect of immersion on the breaking force and failure locus in an epoxy/mild steel system, Journal of Adhesion, 38, pp. 255–268 (1992).
44. C.H. Shen and G.S. Springer, Moisture absorption and desorption of composite materials, Journal of Composite Materials, 10, pp. 2–20 (1976).
45. S. Shi and C.P. Wong, Study of the fluxing agent effects on the properties of no-flow underfill materials for flip-chip applications, Proceedings of the 48th Electronic Components and Technology Conference, 1998, pp. 117–124.
46. R. Shook and J. Goodelle, Handling of highly-moisture sensitive components—an analysis of low-humidity containment and baking schedules, Proceedings of the 49th Electronic Components and Technology Conference, 1999, pp. 809–815.
47. C. Soles, F. Chang, D. Gidley, and A. Yee, Contributions of the nanovoid structure to the kinetics of moisture transport in epoxy resins, Journal of Polymer Science: Part B: Polymer Physics, 38, pp. 776–791 (2000).
48. C. Soles and A. Yee, A discussion of the molecular mechanisms of moisture transport in epoxy resins, Journal of Polymer Science: Part B: Polymer Physics, 38, pp. 792–802 (2000).
49. N. Su, R. Mackie, and W. Harvey, The effects of aging and environment on the fatigue life of adhesive joints, International Journal of Adhesion and Adhesives, 9, pp. 85–91 (1992).
50. E. Suhir, Failure criterion for moisture-sensitive plastic packages of integrated circuit (IC) devices: application of von-Karman's equations with consideration of thermoelastic strains, International Journal of Solids and Structures, 34(23), pp. 2991–3019 (1997).
51. D. Suryanarayana, R. Hsiao, T. Gall, and J. McCreary, Enhancement of flip-chip fatigue life by encapsulation, IEEE Transactions on Components, Hybrids, and Manufacturing Technology, 14(1), pp. 218–223 (1991).
52. M. Uschitsky and E. Suhir, Moisture diffusion in epoxy molding compounds filled with particles, Journal of Electronic Packaging, 123, pp. 47–51 (2001).
53. M. Uschitsky and E. Suhir, Moisture diffusion in epoxy molding compounds filled with silica particles, ASME Structural Analysis in Microelectronics and Fiber Optics, 21, pp. 141–170 (1997).
54. M. Vanlandingham, R. Eduljee, and J. Gillespie, Moisture diffusion in epoxy systems, Journal of Applied Polymer Science, 71, pp. 787–798 (1999).
55. M. Wahab, A. Crocombe, A. Beevers, and K. Ebtehaj, Coupled stress-diffusion analysis for durability study in adhesively bonded joints, International Journal of Adhesion and Adhesives, 22, pp. 61–73 (2002).
56. E. Wong, K. Chan, T. Lim, and T. Lam, Non-fickian moisture properties characterization and diffusion modeling for electronic packages, Proceedings of the 49th IEEE Electronic Components and Technology Conference, 1999, pp. 302–306.
57. W. Wright, The effect of diffusion of water into epoxy resins and their carbon-fibre reinforced composites, Journal of Composites, 12, pp. 201–205 (1981).
58. J. Wylde and J. Spelt, Measurement of adhesive joint fracture properties as a function of environmental degradation, International Journal of Adhesion and Adhesives, 18, pp. 237–246 (1998).
59. G.Z. Xiao and M. Shanahan, Water absorption and desorption in an epoxy resin with degradation, Journal of Polymer Science: Part B: Polymer Physics, 35, pp. 2659–2670 (1997).
60. D. Yeung, M. Yuen, D. Lam, and P. Chan, Measurement of interfacial fracture toughness for microelectronic packages, Journal of Electronics Manufacturing, 10, pp. 139–145 (2000).
61. S. Yi, C. Yue, J. Hsieh, L. Fong, and S. Lahiri, Effects of oxidation and plasma cleaning on the adhesion strength of molding compounds to copper leadframes, Journal of Adhesion Science and Technology, 13, pp. 789–804 (1999).
62. O. Yoshioka, N. Okabe, S. Nagayama, R. Yamagishi, and G. Murakami, Improvement of moisture resistance in plastic encapsulants MOS-IC by surface finishing copper leadframe, Proceedings of the 39th Electronic Components and Technology Conference, 1989, pp. 464–471.
63. M. Zanni-Deffarges and M. Shanahan, Diffusion of water into an epoxy adhesive: comparison between bulk behaviour and adhesive joints, International Journal of Adhesion and Adhesives, 15, pp. 137–142 (1995).
64. M. Zanni-Deffarges and M. Shanahan, Bulk and interphase effects in aged structural joints, Journal of Adhesion, 45, pp. 245–257 (1994).

18

Highly Compliant Bonding Material for Micro- and Opto-Electronic Applications

E. Suhir[a] and D. Ingman[b]

[a]*University of California, Santa Cruz, CA, University of Maryland, College Park, MD, USA*
[b]*Technion, Israel*

18.1. INTRODUCTION

Bonded assemblies (joints) that experience thermal and/or mechanical loading are widely used in micro- and opto-electronics. These assemblies are typically subjected to thermal stresses due to the thermal expansion (contraction) mismatch of the dissimilar materials of the adherends and/or because of temperature gradients. In other cases, bonded assemblies experience mechanical loading. It has been established that the most reliable adhesively bonded or soldered assemblies are characterized by stiff adherends and a compliant adhesive. It has been established also that the employment of low modulus bonding materials and thick bonding layers can lead to an appreciable stress relief [1–7]. This is true for both the interfacial stresses (which are responsible for the adhesive and the cohesive strength of the bonding material) and the stresses acting in the cross-sections of the adherends (these stresses are responsible for the strength of the bonded components) [8–10]. Since the interfacial stresses concentrate at the assembly ends, a substantial stress relief could be expected, if a low modulus material is used at the peripheral portions of the joint [11,12]. Low modulus and thick (up to 4 mils or even thicker) bonding layers are currently employed in micro- and opto-electronic packaging in order to provide a desirable strain buffer between the adherend materials. Certainly, there is always a need in the art for better bonding materials that combine good adhesive properties with high interfacial compliance.

In this connection it should be point out that the reliability of a bonded assembly is due to both high bearing capacity of the joint (i.e., its ability to withstand high stresses of any nature, as well as repetitive loading) and to the low enough level of loading. High compliance of the bonding layer is aimed at reducing the level of the loading, thereby making sufficiently reliable even those joints whose bearing capacity might not be very high. It is also noteworthy that in "conventional" assemblies (which are characterized by moderately compliant bonding layers), the interfacial compliance is due to both the bonding

layer and the bonded components. It is not advisable, however, and, in many cases, is not even possible, to increase the interfacial compliance of the adherends. The most attractive way to increase the interfacial compliance and, hence, the reliability of a bonded joint is to employ a highly compliant, yet thin enough (say, for lower thermal resistance) bonding layer. In assemblies with highly compliant bonds, the role of the adherends, as far as the interfacial compliance is concerned, is insignificant. It is the bonding material only that is responsible for the high and favorable interfacial compliance.

In this chapter we suggest an analytical stress model that enables one to quantitatively assess the expected relief in the interfacial shearing stress due to the application of a highly compliant bonding layer. We use this model to demonstrate that the application of a wire array (WA) and/or a suitable modification of a newly developed nano-particle material (NPM) and structure can improve dramatically the compliance of a bonding layer. This NPM has extraordinary mechanical and environmental properties [13–15]. It has been recently tested in application to a new generation of cladding and coatings for fiber optic systems [16–18]. We expect that this material will find a wide application in micro- and opto-electronics, and beyond.

18.2. EFFECT OF THE INTERFACIAL COMPLIANCE ON THE INTERFACIAL SHEARING STRESS

The objective of the developed analytical stress model (see Appendix 18.A) is to evaluate the effect of the interfacial compliance on the interfacial shearing stress. We use this model to demonstrate the importance of highly compliant bonds. This could certainly be done in addition to, or sometime even instead of, employing adherends with a good thermal extension (contraction) match. The following major conclusions can be drawn from the analysis based on the developed model.

The interfacial shearing stress is proportional to the parameter of the interfacial compliance. This parameter is defined as a square root of the ratio of the axial compliance of the assembly (which, in typical assemblies with thin and/or low modulus bonds, is due to the adherends only and should be made as low as possible) to the assembly's interfacial compliance (which, in typical assemblies is due to both the adherends and the bonding layer). In "conventional" assemblies, i.e., in assemblies with a moderately compliant bonding layer, the interfacial compliance is due to both the adherends and the adhesive: thick and high-modulus adherends and a thin and low-modulus adhesives typically contribute more or less equally to the total interfacial compliance. If, however, a highly compliant bonding layer is employed, it is only this layer, and not the adherends, that is responsible for the interfacial compliance of the bond, and, hence, for the structural reliability of the joint.

Although the predictive model was developed for a "flat" (planar) joint, the above conclusions are valid for any assembly, whether circular (e.g., in fiber optics applications), flat (as in microelectronics applications), elliptical, etc.

The level of stress relief that could be expected by employing a WA as a compliant attachment can be assessed based on the following simple reasoning. Examine a bonded structure, in which the bonding layer is constructed of an array of wires, particularly, nanowires. Let the wires have circular cross-sections and can be considered rigidly clamped at

the ends. Treating each wire as a beam clamped at its ends and experiencing ends offset of the magnitude δ, we find that the lateral forces at the wire ends are

$$N = 12\frac{EI\delta}{h_0^3} = \frac{3\pi}{16}\frac{Ed^4\delta}{h_0^3}, \tag{18.1}$$

where E is the Young's modulus of the wire material, d is the wire diameter, and h_0 is the wires' height, i.e., the thickness of the WA (bonding layer). In the Equation (18.1),

$$I = \frac{\pi \cdot d^2}{64}$$

is the moment of inertia of the wire cross-section. The interfacial compliance of the bonding layer can be found by multiplying the δ/N ratio by the wire cross-sectional area

$$A = \frac{\pi \cdot d^2}{4}. \tag{18.2}$$

This results in the following formula for the interfacial compliance:

$$\kappa_w = \frac{16h_0^3}{3\pi Ed^4}\frac{\pi d^2}{4} = \frac{4}{3}\frac{h_0^3}{Ed^2}. \tag{18.3}$$

Using the Equation (18.A7) for the interfacial compliance of a "conventional" (adhesively bonded or soldered) assembly, we conclude that the ratio

$$\eta = \frac{4}{3}\frac{G_0}{E}\left(\frac{h_0}{d}\right)^2 \tag{18.4}$$

could be used to assess the advantage of a WA-based bonding layer in comparison with a "conventional" layer. Let, for instance, a conventional bonding material has a shear modulus of 5000 psi, the Young's modulus of the WA material be $E = 15 \times 106$ psi, the thickness of the bonding layer be $h_0 = 50$ μm, and the diameter of a single wire be 100 nm. Then the Equation (18.4) yields: $\eta = 111$. Hence, a significant increase in the interfacial compliance could be expected by using a WA-based bonding structure, and, because the interfacial shearing stress is inversely proportional to the square root of the interfacial compliance, one could expect an order of the magnitude decrease in the interfacial shearing stress.

Although the above analysis was carried out for a "clamped-clamped" wire, the results are equally applicable to wires with other boundary conditions at the ends. Additional increase in the interfacial compliance could be achieved, if the wires experience axial compression. In order to assess the expected effect of such a compression, we have examined, as an illustration, in Appendix 18.B, a cantilever wire (beam) subjected at its free end to a lateral (bending) force, P, and an axial compressive force, T. The force P can produce a significantly larger lateral deflection, if the wire is subjected to an external axial compression. Such an effect should be expected, of course, since a compressed wire (beam) experiences lateral deflections even in the absence of any lateral force, provided that the axial force reaches and exceeds its critical value.

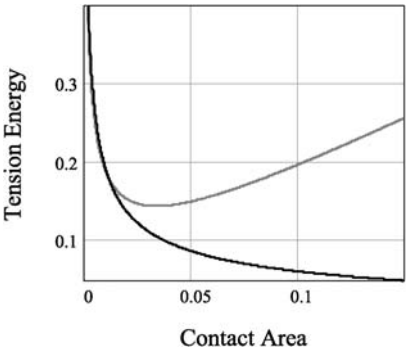


FIGURE 18.1. Surface potential energy vs. "bonded" surface-compound (adhesion) area. Red line — $\sigma_{SC} > \sigma_{SA}$ Blue line — $\sigma_{SC} < \sigma_{SA}$ (see Appendix 18.C).

18.3. INTERNAL COMPRESSIVE FORCES

In addition to the external compression that can be imposed deliberately to enhance the lateral compliance of the assembly, there exist also "internal" compressive forces caused by surface tension.

Consider two bonding surfaces (e.g., parallel surfaces, although these surfaces do not have to be necessarily parallel) and a certain amount of the bonding material (compound) that touches the bonded surfaces. Let us suppose, for the sake of simplicity, that this bonding compound has a shape of a circular cylinder, whose bases touch the bonded surfaces. The lateral surface of the cylinder is "free," i.e., is in contact with air. Let us address a situation, when the area of each of the two bonded surfaces is greater than the bonding area, i.e., the area of the contact with the bonding material. If the surface tension characteristics of the bonding and the bonded materials are properly chosen, then the attraction forces will arise between the bonding surfaces. This phenomenon is similar to the phenomenon observed during the formation of solder balls. So, the bonding compound material will tend to decrease its "free" area and increase its contact area with the bonded surfaces. The total volume of the cylinder remains unchanged. Such a "self-adjusting" process stops when the potential energy of the system (which is as a function of the distance between the bonded surfaces) becomes minimal. A typical plot for this potential energy vs. the contact area is shown in Figure 18.1. The development of the expression for the potential energy is given in the Appendix 18.C.

If the distance between the bonded surfaces becomes larger than the minimum distance that corresponds to the minimum strain energy of the system, then the bonding material behaves as a, sort of, a stretched spring, trying to bring the bonded surfaces closer, i.e., providing axial compression. This compression is proportional to the derivative of the elastic energy with respect to the distance between the bonded surfaces.

18.4. ADVANCED NANO-PARTICLE MATERIAL (NPM)

NPM is an unconventional inhomogeneous "smart" composite material that is equivalent to a "hypothetical" homogeneous material with the following major properties:

- Low Young's modulus.
- Immunity to corrosion.

- Good adhesion to the adjacent material.
- Non-volatile.
- Stable properties at temperature extremes (from $+350°C$ to as low as $-220°C$).
- Very long (practically infinite) lifetime.
- Strong hydrophobicity (against both water and water vapor).
- Ability for "self-healing:" ability to restore its dimensions and initial structure when damaged.
- Ability for "healing" the surfaces of the adjacent materials, i.e., to fill in the existing and/or the developed defects (surface cracks, flaws, and other imperfections) and to slow down their propagation and/or even to "arrest" them completely.
- Very low effective refractive index (if needed).
- High dielectric constant (if needed).

The NPM can be designed, depending on the particular application, in such a way that its important particular properties are enhanced. The conducted tests have confirmed these properties. In general, it is desirable to provide application-specific modifications of the NPM to master (optimize) its properties and performance. Because it is a nano-material, its surface chemistry and its performance depend a lot upon the contact materials and surfaces. The NPM applications include, but are not limited, to the following ones:

- Hermetic sealing of packages, components and devices, such as laser packages, MEMS, displays, plastic LEDs, etc.
- Effective protection (coating) for various metal and non-metal surfaces, well beyond the area of micro- and opto-electronic packaging: (cars, aerospace structures, off-shore and ocean structures, marine vehicles, bridges, towers, tubes, pipes and pipelines, etc.). These applications benefit because the NPM is actively hydrophobic, does not induce additional stresses (owing to its low modulus), is inexpensive, is easy-to-apply, has practically infinite lifetime, and is self-healing. Application of the NPM can result in a significant resistance of a metal surface to corrosion, and, in addition, in substantial increase in the fracture toughness of the material, both initially and during the system's operation (use).
- The NPM can be added in the formulation of various coatings such as paints, thereby providing protective benefits without changing the application techniques.
- Because of a low refractive index, the NPM can be used, if necessary, as an effective cladding of optical silica fibers. The use of the NPM cladding eliminates the need to dope silica for obtaining light-guide cores. The new preform will consist of a single (undoped and, hence, less expensive) silica material.
- A derivative application is flexible light-guides. Multicore flexible fiber cables employing NPM are able to provide high spatial image resolution. As such, they might find important applications, when there is a need to provide direct high-resolution image transmission from secluded areas. Possible applications can be found in biomedicine, nondestructive evaluations, oil and other geological explorations, in ocean engineering, or in other situations when an image needs to be obtained and transmitted from relatively inaccessible locations.
- Another derivative application is a multicore fiber cable. Ultra-small diameter glass fibers with an NPM-based cladding/coating can be placed in large quantities within a NPM medium ("multiple cores in a single cladding"). In addition, owing to a much better inner-outer refractive index ratio in the NPM-based fibers, such cables will be characterized by very low signal attenuation.

- Yet another derivative application is sensor systems. The NPM-based fibers can be used in optical sensor systems that employ optical fibers embedded in a laminar or a cast material. Such systems are used, for instance, in composite airframes. With the NPM used as a cladding or, at least, as a coating of the silica optical fiber, the optical performance, the mechanical reliability and the environmental durability of the light-guide will be improved dramatically compared to the conventional systems.
- Ultra-thin planar light-guides are yet another derivative application of the NPM. In the new generation of the planar light-guides, NPM can be used as the top cladding material. It will replace silicon or polymer claddings, which are considered in today's planar light-guides. All the advantages of the NPM cladding material discussed above for optical fibers are equally applicable to planar light-guides. These are thought to have a "bright" future in the next generation of computers and other photonic devices.

18.5. HIGHLY-COMPLIANT NANO-SYSTEMS

In the current application, intended primarily for micro- and opto-electronic assemblies, we have developed a modification of the NPM for use as a highly compliant bonding structure. Here are some major characteristics of this material (additional to those described in the previous section).

The NPM for the application in question is highly inhomogeneous, anisotropic and thixotropic. It allows for an elevated, but still limited displacement in the in-plane direction, i.e., in the direction of the interfacial shearing stress. The material includes a filler and nanoparticles. The bonding elements are of a "strings-like" structural type. These elements are slightly compressed in the through-thickness direction. This enhances their compliance (flexibility) in the in-plane direction. The thixotropic matrix is quasi-solid in the absence of the applied stress field (disturbance) and becomes much less viscous and even quasi-liquid, when thermal or mechanical stress is applied. In other words, the bonding structure becomes much more compliant as a result of experiencing mechanical or thermal loading. The structure provides a highly effective strain buffer between the adherends.

The role of such a plurality of links (strings) are played by numerous nanotubes and/or nanowires embedded into the bonding matrix. Nanotubes and/or nanowires could be subjected to externally and/or internally deliberately induced compression in the through-thickness direction for even higher lateral compliance. In bonded structures of a circular (such as coated optical fibers or cables) or elliptic shape, the radial compression (hoop stresses) arises because of the thermal contraction mismatch in the radial direction of the high CTE coating and low CTE cladding. In flat (planar) bonded joints the compression in the through-thickness direction could be introduced deliberately (externally or internally).

The filler of the bonding compound has to have a high wetting ability with respect to the bonded surfaces. This means that the surface tension at the interface between the bonding compound (structure) and the bonded surface has to be much lower, than at the "free" surface of the bond. This allows for creating the desirable distance-depending attraction forces

In order to keep the system at the near-optimal distance between the bonded surfaces, the NPM-based compound has to maintain the adequate gap between the surfaces. This can be achieved by introducing microspheres into the compound. Such micro-spheres could serve as, sort of, bearing rolls that increase the system's stability and provide the desirable anisotropy of bonding matrix.

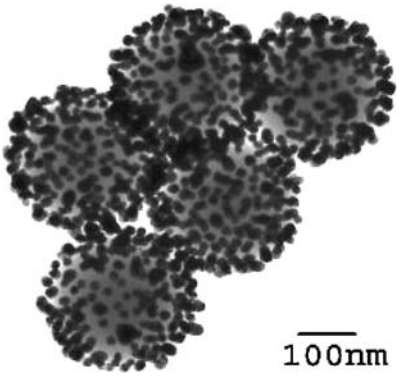


FIGURE 18.2. Microspheres with "protrusions."

The highly-compliant nano-systems in question are extremely heterogeneous. They consist of a bonding matrix with numerous embedded nanoparticles, microspheres and nanotubes/nanowires (whiskers). All these nano- and micro-particles embedded into a thixotropic matrix enable one to create a multi-purpose thixotropic structure. Figure 18.2 illustrates a possible combination of microspheres with "protrusions." These "protrusions" are formed by surface nanoparticles' clusters attracted to the microspheres.

18.6. CONCLUSIONS

The following conclusions can be drawn from the carried out analysis:

- Based on the developed simplified analytical stress model, we demonstrated that the employment of highly compliant materials and structures as bonding layers in bimaterial assemblies (joints) can lead to a significant relief in both the thermally induced and mechanical stresses. The model indicates that the interfacial shearing stress in an adhesively bonded or a soldered assembly is inversely proportional to the square root of the interfacial compliance, and that in "conventional" bimaterial assemblies (which are characterized by not-highly-compliant bonding layers), the interfacial compliance is due to both the bonding layer and the bonded components themselves. However, in assemblies with highly compliant bonds, it is the bonding material only that provides the high and favorable interfacial compliance.
- An appropriate wire array (WA) fabricated on one or both bonded components can be used as a suitable compliant bond. Based on the developed stress model, we demonstrate that its application can lead to a significant, about two orders of magnitude, increase in the interfacial compliance, thereby leading to a reduction in the maximum interfacial shearing stress of about an order of magnitude (compared to the bonded joints using "conventional" adhesives or solders).
- A modification of the newly developed nano-particle material (NPM) be used to increase dramatically the compliance of the bonding layer.
- The combination of both approaches could be effectively used to provide a highly compliant and a highly reliable bonding material and structure. In this case the NPM is employed as a suitable embedding material for the WA.
- Since the NPM has extraordinary mechanical and environmental properties, and, in combination with the appropriate WA, can make an extremely highly compliant

bond, we expect that the NPM and WA, used independently or in combination, might find a wide application in assemblies employed in micro- and opto-electronics, and beyond.

REFERENCES

1. E. Suhir, Thermal stress failures in microelectronics and photonics: prediction and prevention, Future Circuits International, (5) (1999).
2. E. Suhir, Buffering effect of fiber coating and its influence on the proof-test load in optical fibers, Applied Optics, 29(18) (1990).
3. Z. Kovac, et al., Compliant interface for semiconductor chip and method therefor, U.S. Patent #6,133,639, 2000.
4. T.H. DiStefano, et al., Compliant microelectronic mounting device, U.S. Patent #6,370,032, 2002.
5. Z. Kovac, et al., Methods for making electronic assemblies including compliant interfaces, U.S. Patent #6,525,429, 2003.
6. E. Paterson, et al., Mechanical highly compliant thermal interface pad, U.S. Patent #6,910,271, 2005.
7. Z. Kovac, et al., Methods of making microelectronic assemblies including compliant interfaces, U.S. Patent #6,870,272, 2005.
8. E. Suhir, Stresses in bi-metal thermostats, ASME Journal of Applied Mechanics, 53(3) (1986).
9. E. Suhir, Calculated thermally induced stresses in adhesively bonded and soldered assemblies, Proc. of the Int. Symp. on Microelectronics, ISHM, 1986, Atlanta, Georgia, Oct. 1986.
10. E. Suhir, Interfacial stresses in bi-metal thermostats, ASME Journal of Applied Mechanics, 56(3) (1989).
11. E. Suhir, Thermal stress in a polymer coated optical glass fiber with a low modulus coating at the ends, Journal of Materials Research, 16(10) (2001).
12. E. Suhir, Thermal stress in an adhesively bonded joint with a low modulus adhesive layer at the ends, Applied Physics Journal, (April) (2003).
13. E. Suhir, New nano-particle material (NPM) for micro- and opto-electronic packaging applications, IEEE Workshop on Advanced Packaging Materials, Irvine, March 2005.
14. D. Ingman and E. Suhir, Optical fiber with nanoparticle cladding, Patent Application, 2001 (allowed).
15. D. Ingman, V. Ogenko, and E. Suhir, Moisture-resistant nano-particle material and its applications, Patent Application, 2003 (pending).
16. E. Suhir, New hermetic coating for optical fiber dramatically improves strength, DARPA and Navair Workshop, St. Louis, MO, August 2004.
17. E. Suhir, Polymer coated optical glass fiber reliability: could nano-technology make a difference? Polytronic'04, Portland, OR, September 13–15, 2004.
18. E. Suhir, Mechanics of coated optical fibers: review and extension, ECTC'2005, Orlando, Florida, 2005.

APPENDIX 18.A. BIMATERIAL ASSEMBLY SUBJECTED TO AN EXTERNAL SHEARING LOAD AND CHANGE IN TEMPERATURE: EXPECTED STRESS RELIEF DUE TO THE ELEVATED INTERFACIAL COMPLIANCE

The following major assumptions are used in the analysis:

- Approximate methods of structural analysis (strength-of-materials) and materials science, rather than methods of elasticity, can be used to evaluate stresses and displacements.
- The bonded components can be treated, from the standpoint of structural analysis, as elongated rectangular strips that experience linear in-plane elastic deformations.
- At least one of the components ("substrate") is thick (stiff) enough so that bending deformations of the assembly as a whole do not occur and need not be considered.
- All the materials can be treated as linearly elastic.

- The interfacial shearing stresses can be evaluated based on the concept of the interfacial compliance, without considering the effect of the "peeling" stresses (normal interfacial stresses acting in the through-thickness direction).

Let a bimaterial assembly be subjected to both an external shearing force, $\widehat{T}$, and change, Δt, in temperature. Assuming, for instance, that the assembly was manufactured at an elevated temperature and subsequently cooled down to a low (say, room) temperature, one can seek, in an approximate analysis, the longitudinal interfacial displacements, $u_1(x)$ and $u_2(x)$, of the assembly components #1 and #2, respectively, in the form:

$$u_1(x) = -\alpha_1 \Delta t x + \lambda_1 \int_0^x T(\xi) d\xi - \kappa_1 \tau(x),$$
$$u_2(x) = -\alpha_2 \Delta t x - \lambda_2 \int_0^x T(\xi) d\xi + \kappa_{21} \tau(x). \tag{18.A1}$$

In this equations, α_1 and α_2 are the coefficients of thermal expansion (CTEs) of the dissimilar materials,

$$\lambda_1 = \frac{1-\nu_1}{E_1 h_1}, \quad \lambda_2 = \frac{1-\nu_2}{E_2 h_2} \tag{18.A2}$$

are the axial compliances of the assembly components, E_1 and E_2 are the Young's moduli of the materials, ν_1 and ν_2 are their Poisson's ratios, h_1 and h_2 are the thicknesses of the assembly components,

$$T(x) = x \int_{-l}^x \tau(\xi) d\xi \tag{18.A3}$$

are the thermally induced forces acting in the cross-sections of the assembly components, $\tau(x)$ is the thus far unknown interfacial shearing stress, l is half the assembly length,

$$\kappa_1 = \frac{h_1}{3G_1}, \quad \kappa_2 = \frac{h_2}{3G_2} \tag{18.A4}$$

are the interfacial compliances of the assembly components, and

$$G_1 = \frac{E_1}{2(1+\nu_1)}, \quad G_2 = \frac{E_2}{2(1+\nu_2)} \tag{18.A5}$$

are the shear moduli of the adherend materials. The origin, O, of the coordinate, x, is at the mid-cross-section of the assembly. The condition of the compatibility of the longitudinal interfacial displacements (18.A1) can be written as follows:

$$u_1(x) = u_2(x) - \kappa_0 \tau(x), \tag{18.A6}$$

where

$$\kappa_0 = \frac{h_0}{G_0} = 2(1+\nu_0)\frac{h_0}{E_0} \tag{18.A7}$$

is the interfacial compliance of the bonding layer, h_0 is the thickness of this layer, E_0 and ν_0 are the elastic constants of the bonding material, and G_0 is its shear modulus. Introducing the Equation (18.A1) into the compatibility condition (18.A6), we obtain the following equation for the sought interfacial shearing stress function, $\tau(x)$:

$$\kappa\tau(x) - \lambda\int_0^x T(\xi)d\xi = \Delta\alpha\Delta t x, \tag{18.A8}$$

where

$$\kappa = \kappa_0 + \kappa_1 + \kappa_2 \tag{18.A9}$$

is the total interfacial compliance of the assembly,

$$\lambda = \lambda_1 + \lambda_2 \tag{18.A10}$$

is the total axial compliance, and $\Delta\alpha = \alpha_1 - \alpha_2$ is the thermal contraction mismatch of the adherend materials.

Differentiating the Equation (18.A8) with respect to the coordinate, x, we obtain:

$$\kappa\tau'(x) - \lambda T(x) = \Delta\alpha\Delta t. \tag{18.A11}$$

The next differentiation, taking into account the Equation (18.A3), yields:

$$\kappa\tau''(x) - \lambda T(x) = 0. \tag{18.A12}$$

The boundary conditions

$$T(-l) = 0, \qquad T(l) = \widehat{T} \tag{18.A13}$$

for the induced force can be translated, using the Equation (18.A11), into the boundary conditions for the interfacial shearing stress, $\tau(x)$, as follows:

$$\tau'(-l) = \frac{\Delta\alpha\Delta t}{\kappa}, \quad \tau'(l) = \frac{\Delta\alpha\Delta t + \lambda\,\widehat{T}}{\kappa}. \tag{18.A14}$$

Equation (18.A12) has the following solution:

$$\tau(x) = C_1\sinh kx + C_2\cosh kx, \tag{18.A15}$$

where

$$k = \sqrt{\frac{\lambda}{\kappa}} \tag{18.A16}$$

is the parameter of the interfacial sharing stress. Introducing the solution (18.A15) into the boundary conditions (18.A14), solving the obtained equations for the constants C_1

and C_2 of integration, and substituting the formulas for the constants of integration into the solution (18.A15), we obtain the following expression for the interfacial shearing stress:

$$\tau(x) = k\left\{\frac{\Delta\alpha\,\Delta t}{\lambda}\frac{\sinh kx}{\cosh kl} + \hat{T}\frac{\cosh[k(x+l)]}{\sinh 2kl}\right\}. \tag{18.A17}$$

The maximum value of this stress occurs, in this simplified stress model, at the end $x = l$:

$$\tau_{\max} = \tau(l) = k\left(\frac{\Delta\alpha\,\Delta t}{\lambda}\tanh kl + \hat{T}\coth 2kl\right). \tag{18.A18}$$

In sufficiently long and/or stiff assemblies, this formula yields:

$$\tau_{\max} = \tau(l) = k\left(\frac{\Delta\alpha\,\Delta t}{\lambda} + \hat{T}\right). \tag{18.A19}$$

Thus, it is the parameter, k, of the interfacial shearing stress that is responsible for both the level of this stress and the length of the end portions of the assembly that experience elevated interfacial stresses. As evident from the Equation (18.A16), this parameter decreases with a decrease in the axial compliance, λ, of the assembly and with an increase in the assembly's interfacial compliance κ. The axial compliance, λ, as evident from the Equations (18.A2), is due to the adherends only, and it is desirable, as evident from the Equation (18.A6), that these adherends are stiff, i.e., are characterized by a low axial compliance. As to the interfacial compliance, κ, this compliance, for conventional assemblies with not a very compliant bonding layer, is due to both the interfacial compliance of the bonding layer and the adherends themselves: the bonding layer is typically thin and low modulus, while the adherends are thick and high modulus, so, as evident from the Equations (18.A4) and (18.A7), the contribution of the compliances of the adherends and the bonding layer to the total compliance expressed by the Equation (18.A9), might be quite comparable. From the obtained formulas it is clear also that there is no incentive to increase the interfacial compliance of the adherends: by doing so, one would inevitably increase the axial compliance of the assembly as well, which is highly undesirable. Hence, for lower interfacial shearing stress, one should design a joint with a highly compliant bonding layer. This will make the factor k lower, will bring down the maximum interfacial shearing stress (which is inversely proportional to the square root of the interfacial compliance) and will spread the interfacial shearing loading over larger areas at the assembly ends.

APPENDIX 18.B. CANTILEVER WIRE ("BEAM") SUBJECTED AT ITS FREE END TO A LATERAL (BENDING) AND AN AXIAL (COMPRESSIVE) FORCE

Let a cantilever wire ("beam") be subjected at its free end to a lateral (bending) force, P, and an axial compressive force, T. The equation of bending of such a beam is as follows:

$$EIw''(x) + Tw(x) = 0. \tag{18.B1}$$

Here $w(x)$ is the deflection function of the beam, and EI is its flexural rigidity. The origin, O, of the coordinate x is at the clamped end of the wire. The following boundary conditions should be satisfied:

$$w(0) = 0,\ w'(0) = 0,\ w''(l) - 0,$$
$$EIw'''(l) + Tw'(l) + P - 0. \tag{18.B2}$$

In order to satisfy the four conditions (18.B2), the Equation (18.B1) is differentiated twice, and the solution to the obtained differential equation of the fourth order is as follows:

$$w(x) = C_0 + C_1 kx + C_2 \cos kx + C_3 \sin kx, \tag{18.B3}$$

where

$$k = \sqrt{\frac{T}{EI}} \tag{18.B4}$$

is the parameter of the compressive force. Introducing the solution (18.B3) into the boundary conditions (18.B2), we find:

$$C_0 = -C_2 = \frac{P}{kT}\tan kl,$$
$$C_1 = -C_3 = -\frac{P}{kT}, \tag{18.B5}$$

and the solution (18.B3) results in the following expression for the deflection function:

$$w(x) = \frac{P}{kT}\left\{\tan kl - kx - \frac{\sinh[k(l-x)]}{\cosh kl}\right\}. \tag{18.B6}$$

The maximum deflection at the wire end is

$$w(l) = \frac{P}{kT}(\tan kl - kl). \tag{18.B7}$$

The maximum deflection at the free end of a cantilever beam subjected to a lateral force P applied at this end is

$$\delta = \frac{Pl^3}{3EI}. \tag{18.B8}$$

Comparing the Equations (18.B5) and (18.B6), we conclude that the parameter

$$\varsigma = 3\frac{\tan kl - kl}{(kl)^3} \tag{18.B9}$$

considers the effect of the axial compression on the lateral compliance of a cantilever wire ("beam"). Within the framework of the linear approximation, the parameter ς is infinitely

large, when the compressive force, T, reaches its critical (Euler) value. Indeed, this parameter is infinitely large, when $kl \to \pi/2 = 1.571$. In this case the Equation (18.B3) yields

$$T = \frac{\pi^2 \cdot EI}{4l^2},$$

which is a well known expression for a critical force for a cantilever beam. But even when the parameter kl is only $kl = 0.866$, the parameter ς is as high as about 6.4, and in such a case a 2.5 fold decrease in the interfacial shearing stress could be expected. Thus, there is a definite incentive, as far as the interfacial shearing stress is concerned, for using nanowire-based bonding structures subjected to compression.

APPENDIX 18.C. COMPRESSIVE FORCES IN THE NPM-BASED COMPOUND STRUCTURE

The desirable compression in the compound structure could be achieved by application of external forces, but it certainly can stem from the bonding compound features, as quantitatively shown below.

Let us define the system parameters as following:

S_S overall are of surfaces to be bonded,
S_B interfacial area of contact between the bonded surface and bonding compound,
S_C side area of the compound (in contact with air),
σ_{SA} interfacial tension coefficient between the bonded surface and air,
σ_{SC} interfacial tension coefficient between the bonded surface and the compound,
σ_{CA} interfacial tension coefficient between the compound and air,
d gap between the bonded surfaces,
R_C radius of the contact area,
and
V_C volume of the compound located between the bonded surfaces.

Constraint of the compound volume conservation during any spread/wetting process should be considered in the expression for the system surface tension energy:

$$E = 2 \cdot \sigma_{SA} \cdot (S_S - S_B) + \sigma_{CA} \cdot S_C + 2 \cdot \sigma_{SC} \cdot S_B$$

provided

$$S_B \cdot d = V_C. \tag{18.C1}$$

It is clear that

$$S_B = \pi \cdot R_C^2,$$

and

$$S_C = 2 \cdot \pi \cdot R_C \cdot d. \tag{18.C2}$$

Then, the system energy can be easily expressed vs. d or S_B. Let us bring the dependence of this energy on the contact area S_B:

$$E(S_B) = 2 \cdot \left[\sigma_{SA} \cdot S_S + \sigma_{CA} \cdot \frac{V_C \cdot \sqrt{\pi}}{\sqrt{S_B}} + (\sigma_{SC} - \sigma_{SA}) \cdot S_B \right]. \tag{18.C3}$$

In a case of $\sigma_{SC} < \sigma_{SA}$ there the energy function does not have any minimum, and the system tends to the complete bonded surface coverage with the compound, and the equilibrium is achieved only with the help of repulsive forces provided by the nanowires, microspheres and other embedded "elastic" elements. On the other hand, in an opposite case, when $\sigma_{SC} > \sigma_{SA}$, the energy has a minimum at

$$S_B = \sqrt[3]{\frac{\pi \cdot V_C^2}{4} \cdot \frac{\sigma_{SC}^2}{(\sigma_{SC} - \sigma_{SA})^2}},$$

or

$$d = \sqrt[3]{\frac{4 \cdot V_C}{\pi} \cdot \frac{(\sigma_{SC} - \sigma_{SA})^2}{\sigma_{SC}^2}}. \tag{18.C4}$$

As a quasi-elastic system (actually, visco-elastic system), this system tries to achieve a state of equilibrium that corresponds to the minimum of the elastic energy associated with surface tension and the nanowires buckling and microspheres compression. Then the system will operate in the condition of a dynamic steady-state (stable) equilibrium with the "distractive" reactive forces caused by the compressed nano-wires and the compression forces of the surfaces tension.

19

Adhesive Bonding of Passive Optical Components

Anne-Claire Pliska and Christian Bosshard

CSEM SA, Untere Grundlistrasse 1, 6550 Alpnach Dorf, Switzerland

19.1. INTRODUCTION

Fiber-optic communication involves generation, transmission, amplification, detection and processing of optical signals. For each of these functions, a multiplicity of optical components is needed: transmitters, receivers, modulators, splitters, fibers, filters, couplers, optical isolators, pump lasers, amplifiers. The proliferation of wavelength channels and passive components has triggered a dramatic increase in the amount of parts found in optical networks. As 60 to 80% of the cost of an optical component is in the packaging, the prospect of housing several components in the same package becomes attractive, especially when this strategy can eliminate several fiber pigtailing operations and lossy transitions between discrete components. The integration of electronic and optoelectronic chips on the same platform is expected to be the next breakthrough in optoelectronic packaging and to reduce the overall device cost. The benefits of squeezing more components into a single optical module are: reduced inventory costs, space savings, simplified mechanical design, reduction in fiber splicing. On the electronic side, as data rate continues to increase, there is a need to bring electronic circuitry (e.g., drivers, pre-amplifiers) and active optical components (e.g., lasers, amplifiers, photodiodes ...) closer together to avoid parasitic capacitances.

Optics has not found an equivalent of the transistor, i.e., a single building block from which all other functions can be constructed in a single process. As a result, almost every optical function needs its own discrete element requiring its own fabrication process. In order to increase functionality into a single housing, two main strategies are being currently pursued:

- Monolithic integration uses a single-material to fabricate several components on a common substrate. So far, it is not clear which technology platform can handle fully integrated products and whether this platform can do it cost effectively.
- Hybridization relies on the assembly on the same board of specialty subcomponents, each optimized in its own material system.

The hybridization scheme and the need for reduced assembly costs are driving the development of new assembly technologies and new optical designs providing relaxed positioning tolerances:

- Optical building blocks with a collimated beam output [61].
- New passive waveguide technology: silicon waveguides, polymer waveguides [56], hollow waveguides [29].
- Disruptive fixing technologies including adhesive bonding or clipping mechanisms [37,48,55].
- MEMS-based alignment structures.
- Silicon-based heat pipes for improved thermal management [35].
- Local heating options or local curing options for multi-chip assembly [3].
- Development of dedicated automation tools [14].

To assemble the bare dies and optical fibers onto a common substrate or into housings, the following packaging processes are available today: soldering, laser welding, thermo-compression, mechanical interlocking, resistance welding and adhesive bonding.

Although adhesive bonding has been used for many years in the microelectronics industry, there is still a mental barrier for using adhesives in optical assemblies and packages. Historically, the reliability of adhesives in opto-electronic assemblies has often been questioned (outgassing, mechanical stability, photostability). With the release of new adhesives specifically developed for the assembly of fiber-optic components and a better understanding of the assembly processes, adhesives are nowadays increasingly used in optical packages, e.g., to fix optical fibers or lenses. Adhesives are applicable to a wide variety of materials and optical assembly tasks. Other advantages of adhesive bonding over alternative fixing techniques include low processing temperature and low-cost equipment. Additionally, a suitable adhesive can be selected for each optical assembly application from the large variety of products available on the market, making adhesive bonding a versatile fixing technology.

In order to better understand the requirements for a bonding process in opto-electronics packaging, we will first discuss the typical positioning tolerances when an optical connection is involved. Then, the influence of the thermal stress at a chip and package level, as well as the impact of creep on the performances of optical assemblies, will be evaluated. A comparison between adhesive and solder materials will be carried out.

In a next section, we will review the existing adhesive bonding theories and the requirements that must be met to create a reliable and low-stress joint. A special emphasis will be put on the surface energy concept and the surface preparation of adherends. Additionally, the influence of the glass transition temperature (T_g) and the coefficient of thermal expansion (CTE) on the relative displacement between optical devices will be discussed.

Ultimately, we will describe two examples of optical assemblies where an adhesive material is used to fix optical fibers. The first application consists in a laser pigtailing operation: we describe the alignment and fixing process and we present thermal cycling results. The second application concerns a channel waveguide where a fiber has to be aligned and fixed on both sides of the device.

19.2. OPTICAL DEVICES AND ASSEMBLIES

19.2.1. Optical Components

Through the exploitation of the unique properties of integrated, free space and fiber optics, a wide variety of active and passive optical devices are currently offered to supply the telecom and datacom market. Active devices require driving electronics using typical wiring technologies or external optical pump signals. These devices are used for:

- Opto-electronic conversion (optical-to-electrical, or vice-versa): diode laser (edge emitters, VCSEL,[1] LED,[2]) photodiode (surface or edge illuminated).
- Manipulation of the signal: modulators, amplifiers (SOA,[3] EDFA[4]), attenuators, switches.

Passive devices do not have driving electronics. They simply filter or route the signal based on wavelength, intensity or polarization. They include lenses, mirrors, prisms, isolators, couplers, multiplexers, and demultiplexers.

Table 19.1 and 19.2 list a variety of active and passive optical components along with the corresponding key material systems addressing the 1.55 μm telecommunication window. The applications of optoelectronic devices, ranging from long-haul to metro and access devices, include a large mix of material classes and properties. Thus, assembling several devices out of different materials in the same module requires special attention to packaging design (substrate, bonding material, etc.). So far, with an exception of low-cost TO[5]-based packages, the assembly processes remain proprietary and the standardization of packaging techniques are yet to come.

19.2.2. Opto-electronics Assemblies: Specific Requirements

In photonics, packaging must provide not only electrical connections and mechanical support but also thermal management and, more critically, optical connections. Figure 19.1 and 19.2 presents a "butterfly" housing used, e.g., in laser diodes or in high-speed photodetector packaging where the light from a semiconductor laser diode is coupled into an optical fiber through microlenses in free space propagation. Optical interconnects are highly directional. In order to achieve optimal signal transmission, they require precise control of the relative location between components and an attachment process that maintains the alignment over time. In single-mode fiber applications, the positional tolerance is typically in the sub-micron range. Therefore, the assembly of optoelectronic components provides unique requirements and challenges.

19.2.2.1. Optical Connection Optimizing optical connection is the most critical step in the assembly of an optoelectronic package. In a laser to fiber coupling configuration, the optical coupling efficiency η of the laser beam ψ_i in the optical fiber characterized by its

[1] VCSEL: vertical cavity surface emitting laser.
[2] LED: light emitting diode.
[3] SOA: semiconductor optical amplifier.
[4] EDFA: erbium-doped fiber amplifier.
[5] TO: transistor outline.

TABLE 19.1.
Examples of active functions and the corresponding material classes.

Active components	Material	References
Transmitters	III-V semiconductor	[4,15]
Receivers–photodetectors	InGaAsP	[1,22]
Modulators	III-V semiconductor, $LiNbO_3$, polymer, silicon	[17,60] [18,36]
Amplifiers	III-V semiconductor, doped polymer, doped glass fiber	[65] [8] [2]
Pump lasers	III-V semiconductor	[43]
Switch	Polymer, silicone III-V semiconductor	[45] [58]

TABLE 19.2.
Examples of passive functions and the corresponding material classes.

Passive components	Material	References
Arrayed waveguide grating	Silica, SiON, InP, polymer	[26,30]
Passive optical waveguides	Silica, SOI, SiON polymer, sol-gel, hollow waveguide	[29,56]
Coupler, splitter	Silica, SOI, SiON polymer, sol-gel, hollow waveguide	[23]
Lenses	Silica, silicon	[61]

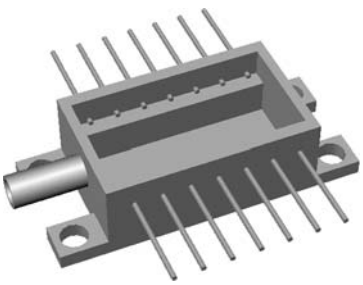


FIGURE 19.1. Butterfly package used for the packaging of transmitters, pump lasers, edge detectors. Internal package dimensions are: 18.9 mm × 10.2 mm × 6.4 mm.

fundamental optical mode ψ_f is related to the overlap integral between the two electric fields and is defined as:

$$\eta = \frac{\left| \iint \psi_i \psi_f^* dA_f \right|^2}{\left(\iint |\psi_i|^2 dA_i \right) \left(\iint |\psi_f|^2 dA_f \right)}. \tag{19.1}$$

In Equation (19.1), A_i and A_f refer to the integration surface in the plane of the incident field and the optical fiber fundamental field, respectively. The small refractive

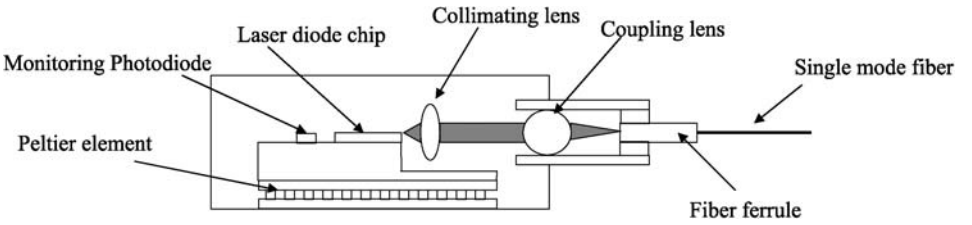


FIGURE 19.2. Typical structure of a laser module: the laser chip, a monitoring photodiode and a thermistor (not represented) are soldered on a submount. The laser beam is collimated by a first lens (collimating lens) and is then focused by the second lens (coupling lens) on the facet of the single mode optical fiber. A Peltier element provides active cooling of the laser chip during operation.

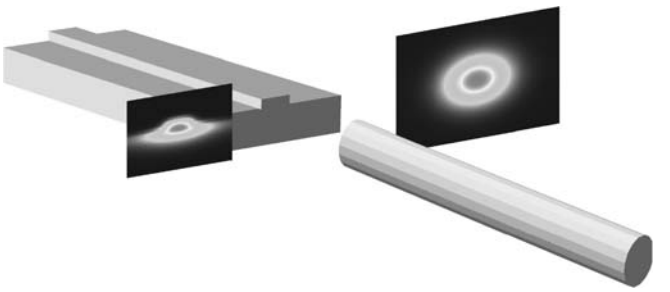


FIGURE 19.3. End-fire coupling between a laser diode chip and a single mode optical fiber. The mode shape of the ridge laser diode is highly asymmetric and smaller than the fundamental mode in the optical fiber.

index difference in a glass fiber of $\Delta n \approx 5 \times 10^{-3}$ results in a weakly guided optical mode with a typical mode size of 8–10 μm. In planar waveguides, including semiconductor integrated optical devices, Δn is often larger than 10^{-2}, leading to a mode size smaller than 3 μm. Moreover, unlike the circular mode in a fiber, the mode shape in a planar device is elliptical, resulting in an additional mode mismatch.

In order to estimate the coupling efficiencies and alignment tolerances, the Gaussian field approximation can be used in most cases [53]. In addition, the integration of intermediate microlenses or the use of lensed fibers can be simulated with the ABCD propagation matrix theory.

In the following example, the computation of the coupling efficiency of a 1550 nm laser diode into a single mode fiber is considered. We assume Gaussian profiles for the laser diode field distributions in the directions parallel and orthogonal to the junction plane with beam spot sizes $2\omega_{//} = 3$ μm and $2\omega_{\perp} = 1$ μm (divergence: $\theta_{//} = 9.4°$, $\theta_{\perp} = 28.3°$). The fundamental field HE_{11} in the single mode optical fiber is approximated by a Gaussian beam having a beam spot size $2\omega_f = 9.8$ μm. In an end-fire coupling configuration (direct fiber coupling without using intermediate lenses, see in Figure 19.3), the calculated optical coupling efficiency is roughly 16%. This percentage decreases if the alignment is not optimal. The solid line in Figure 19.4 shows coupling efficiency variations when a lateral (orthogonal) misalignment is introduced in the end-fire coupling situation. The coupling tolerance, defined as the lateral misalignment yielding 1 dB of additional coupling losses, is here 1.7 μm in the direction orthogonal to the junction plane.

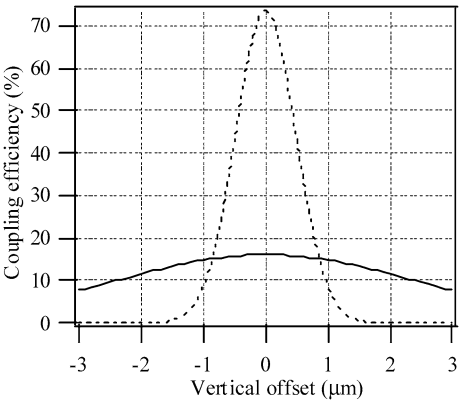


FIGURE 19.4. Evaluation of the coupling of a 1.55 μm laser diode with beam dimensions of 2×3 μm into a single mode fiber using a cleaved fiber (solid line) and a ball lens fiber (dash line). With a cleaved fiber, the optical coupling efficiency is roughly 16% and the "orthogonal" positioning tolerance for 1 dB additional coupling losses is 1.7 μm. With a ball lens fiber, the optical coupling efficiency rises up to 73% but the "orthogonal" positioning tolerance for 1 dB additional coupling losses becomes much tighter (0.4 μm).

Lensed fibers provide a highly effective way to improve the coupling efficiency between fibers and optical devices [42]. Various tip designs eliminate the need for a separate lens thereby reducing the return loss and the assembly costs. In the above example of the laser diode pigtailing, the coupling efficiency can be dramatically increased if a ball lens fiber is used. Indeed, an optimal coupling efficiency of 73% is found when the fiber lens radius is 4 μm and the distance between the laser facet and the fiber tip is 8 μm. Unfortunately, this optimized coupling configuration is achieved at the expense of reduced alignment tolerances. The dash curve in Figure 19.4 shows the coupling efficiency variations when a lateral (orthogonal) misalignment is introduced in the lensed fiber pigtailing situation. The coupling tolerance is here as low as 0.4 μm in the direction orthogonal to the junction plane.

In practice, in order to achieve optimal alignment, the fiber is held in a gripper mounted on an actuator-controlled stage in front of the powered laser diode endface. Stray light ("first light") is coupled into the optical fiber and detected with an optical power meter connected at the output of the fiber. Using dedicated alignment algorithms (hill-climb, triangulation, raster scan, spiral scan), the optical fiber stage is moved until maximum optical power is detected.

Fiber loading, alignment and fixing are still the bottleneck of an optoelectronic assembly. Indeed, it usually takes more than 4 minutes to align an optical fiber to a semiconductor laser [62]. Using machine vision and pattern recognition, stray light detection is sped up and the alignment time can be shortened down to less than 1.5 minutes [47].

Passive alignment is an alternative approach to increase the integration of optoelectronic components and to drive down the packaging costs. This technique does not rely on emitted light for accurate coupling to single mode fibers and allows the alignment of optical fibers with passive devices, such as switches, array waveguide grating. In this assembly technique, optoelectronic chips, lenses and optical fibers have to be passively aligned. The flip-chip process on a structured substrate is the common tool for micropositioning the components [64] provided that:

- The alignment process of the chip is either based on vision if the machine specifications allow for a pre-bond accuracy below 1 μm [16] or/and on the C4[6] process first developed by IBM in the 60's [41].
- The fiber is, e.g., passively aligned in a micro-machined groove in front of the opto-electronic chip.

Chip to fiber alignment performance achievable with this technology is still somewhat reduced compared with standard active alignment techniques and cannot be used where sub-micron positioning accuracy is required.

19.2.2.2. Substrate Materials The choice of the submount material is mainly driven by the need of matching the CTE of both the optical components and the submount [44,57] as well as providing an efficient thermal pathway in uncooled devices or in high power applications [43].

19.2.2.2.1. Thermal Stresses. In this section, the thermal stresses in a tri-material assembly are considered. We evaluate the influence of the substrate material for a given GaAs laser chip and we compare the ability of the intermediate solder or adhesive material to accommodate the CTE-mismatch and to limit the thermal stresses in the assembly.

A GaAs laser chip soldered onto a heat spreading copper substrate with a AuSn (80 wt% Au) solder represents an extreme case of the CTE mismatch. The CTE of GaAs and Copper are, respectively, 6.5×10^{-6} K^{-1} and 17.8×10^{-6} K^{-1}. During the heating step of the soldering process, the semiconductor laser die and the copper substrate are free to expand. The copper substrate, due to its larger CTE, will expand more than the GaAs die. During the cooling phase of the process, the stress-free displacements between the die and the submount are prevented by the solder layer, and thermal stresses are induced in the assembly [57]. These include:

- Normal stresses acting over the cross sections of the components. These stresses are responsible for the strength of the components (die or substrate) themselves.
- Shearing and transverse normal ("peeling") stresses at the interfaces. These stresses are responsible for the adhesion or cohesion strength of the die attach material.

Using 2D modeling and assuming that there is no bending, the CTE mismatch induced shear stress $\tau_{\alpha,s}$ at the substrate-solder and at the chip–solder interface is maximum at the edge of the chip. Its value is given by [12]:

$$\tau_{\alpha,s} = \Delta\alpha \Delta T G_d \frac{\tanh \beta L}{\beta t_d}, \tag{19.2}$$

where β is defined as:

$$\beta = \sqrt{\frac{G_d}{t_d}\left(\frac{1}{E_c t_c} + \frac{1}{E_s t_s}\right)}. \tag{19.3}$$

The letters c, s and d stand for chip, substrate and die attach materials, respectively, $\Delta\alpha$ is the CTE mismatch between the chip and the substrate, ΔT is the temperature change. E and G are the elastic moduli in tension and shear, respectively, t represents the materials thickness, and $2L$ is the chip length.

[6] C4: controlled chip collapse connection.

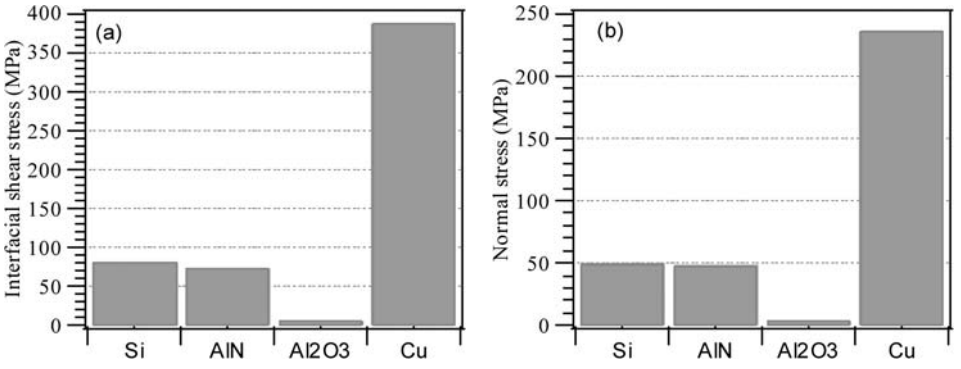


FIGURE 19.5. (a) Comparison of the CTE-mismatch induced interfacial shear stresses when a 2400 μm GaAs chip is soldered to different substrate materials. The intermediate solder material is AuSn (80%wt Au) and the temperature difference between the peak soldering temperature and the room temperature is 300°C. The shear stresses at the chip–solder and at the substrate–solder interfaces are minimum when an Alumina submount is used (7.4 MPa). The influence of the bow in the substrate and in the chip is not considered in these calculations. (b) Corresponding normal stresses in the GaAs die. The normal stress in the GaAs die soldered on a Copper substrate exceeds the GaAs strength fracture. Hence, the die is expected to crack upon a rapid cooling phase of the soldering process.

Similarly, the maximum normal stresses $\sigma_{\alpha,n}$ (at the center of the chip) are given by:

$$\sigma_{\alpha,n} = \frac{\Delta\alpha \Delta T G_d}{\beta^2 t_d}\left(1 - \frac{1}{\cosh \beta L}\right). \tag{19.4}$$

In Figure 19.5(a), we numerically evaluate the influence of the substrate on the maximum interfacial shear stress built-up during soldering of a 2400 μm long GaAs chip using a AuSn (80 wt% Au) solder. We assumed a solder material thickness of 20 μm in the calculations and a maximum temperature variation between room and soldering temperature of 300°C. The following submount materials are compared: copper, silicon, AlN and Al_2O_3. In many applications involving active optoelectronic semiconductor chips based on InP or GaAs, the preferred submount material is AlN due to high thermal dissipation and good CTE matching properties. On the other hand, Al_2O_3 and silicon are the materials of choice when cost savings are critical [7,64]. The substrates and solder materials properties involved in the calculations are summarized in Table 19.3.

The thermally induced shear stresses at the edge of the AuSn solder layer are approximately 10 times smaller when using an Al_2O_3 (7.4 MPa) instead of an AlN (73.5 MPa) or a silicon (80 MPa) substrate. For all the three substrates, the normal stresses displayed in Figure 19.5(b) remain below the GaAs fracture strength (85 MPa). On the other hand, the estimated interfacial shear stresses are 387 MPa when the GaAs chip is bonded onto a copper substrate. Assuming that the solder exhibits a linear elastic behavior at such stress values, the normal (tensile) stresses in the GaAs chip are estimated to be 236 MPa, i.e., above the fracture strength of GaAs. However, as the estimated interfacial shear stresses exceed the yield stress of the gold-tin solder (275 MPa), a plastic deformation is induced in the solder during the cooling phase of the soldering process. This plastic deformation redistributes the stresses in the solder layer and finally reduces the normal stresses in the GaAs chip [33]. Moreover, considering the creep-induced stress relaxation and slowing down the cooling process, once could further reduce the interfacial stress. This effect has proven to

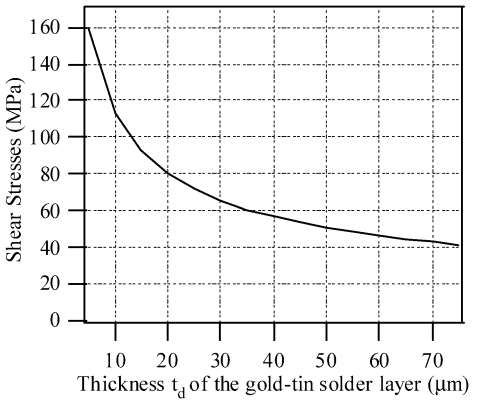


FIGURE 19.6. Dependence of the CTE mismatch induced shear stresses in a gold-tin solder as a function of the solder thickness assuming a difference between room and soldering temperature of 300°C. The interfacial shear stresses are better accommodated with a thick solder layer.

TABLE 19.3.
Thermal and mechanical properties of common packaging substrates materials.

Material	CTE (ppm/K)	Thermal conductivity (W/m K)	Young modulus (GPa)	Shear modulus (GPa)
Si	4.2	150	130	52
GaAs	6.5	54	86	33
Cu	17.8	400	110	46
CuW	7	180	260	
Al	23.6	240	70	25
Kovar (Fe:Ni:Co)	5.8	15	138	52
AlN	4.5	170	350	140
Al_2O_3	6.7	21	390	125
AlSiC	8	200	188	76
Diamond	2	2000	800	
AuSn (80% Au wt)	16	58	68	25
AgSn (96.5% Ag wt)	22	36	50	19
PbSn (37% Pb wt)	21	21	40	14
Epoxy	50	0.3	3	1.2
Acrylate	220		1.1	0.4

reduce also normal stresses from 130 MPa to 80 MPa in the case of GaAs chips soldered on diamond substrates [9] and could eventually be applied to the GaAs chip on copper to avoid chip cracking.

One can decrease the interfacial shear stress by increasing the die attach material thickness t_d. Figure 19.6 represents the dependence of the maximum induced shear stress in the gold-tin solder layer in a GaAs chip on a silicon submount, when the maximum temperature change between the room and the soldering temperature is 300°C. The shear stresses at the substrate-solder and at the chip–solder interfaces are reduced by approximately 40%, when the solder layer thickness is increased from 20 μm to 60 μm. By this

means, the normal stresses in the GaAs die are also reduced and the assembly integrity is ensured.

For a defined CTE-mismatch between the substrate and the chip, the interfacial shear stresses is limited, if a compliant and/or thick die attach material is used. Indeed, the use of a compliant epoxy material requiring a reduced temperature excursion during the bonding process decreases the shear stresses from approximately 80 MPa (with a AuSn solder) down to 8.7 MPa.

Based on these calculations, we conclude that a compliant epoxy material limits the thermo-mechanical stresses and reduce the likelihood of cracks generation in the die or the substrate compared with a solder alloy material. Additionally, the reduced stress in adhesive bonds limits the adverse effect of creep in adhesively bonded assemblies.

19.2.2.2.2. Thermal Conductivity and CTE Trade-Off. A good trade-off between the CTE matching and thermal conductivity of the submount material must be found. Al_2O_3 offers ideal CTE properties, however, its low thermal conductivity can be detrimental for a proper operation of an active GaAs device. AlN or silicon materials are good alternatives to Al_2O_3. Indeed, a stack with a 150 μm thick GaAs chip mounted on an AlN substrate and a CuW package base, both 500 μm thick, has a thermal resistance of approximately 22 $K\,W^{-1}$. This assembly compares favorably to the Al_2O_3 substrate and package base configuration where a thermal resistance of 44 $K\,W^{-1}$ is estimated. A stack with a silicon substrate and a light AlSi package base give identical heat flow performances as the AlN/CuW combination.

In addition to good CTE matching with other semiconductor materials and high thermal conductivity, silicon offers an extensive hybridization potential. Standard IC photolithography and structuring processes not only allow the fabrication of electrical interconnects but also opens the way to the fabrication of high-precision alignment features needed to mount optical devices [20]:

- V-grooves or U-grooves for fiber alignment [50,55].
- Microlens mounting [61].
- Standoff and indentations for flip-chip mounting of semiconductor chips [64].
- Reservoirs for adhesive [32].

Increased functionality of the silicon bench can be achieved with the monolithic integration of resistors, capacitors or inductors. For example, miniature polysilicone heaters can be integrated on a silicon board to enable local heating for reflow soldering of multichip modules [50]. Excellent RF properties can be achieved using Polyimide or BCB intermediate layers [59]. Additionally, deposition of TaN thin film resistor on silicon microbench can provide electrical damping for 10 Gbits/s driving signals in transceiver modules [3]. Optical functions can also be monolithically integrated on the silicon platform using either silica or silicon waveguides and devices [36].

19.2.2.3. Die Attach Material Requirements The most commonly used die attach materials in opto-electronics are solder alloys, adhesives and glass solders. Advantages and disadvantages of these materials are briefly summarized in Table 19.4. Minimizing the thermal stresses in the assembly process or in device operation is a concern in both microelectronics and opto-electronics. We have seen in Section 19.2.2.2.1 that an epoxy material compares favorably with a gold-tin solder as far as CTE mismatch induced shear stresses are concerned. In addition to the minimization of the mechanical stresses in the assembly, bonding materials in optical assemblies must fulfill the following requirements: low

TABLE 19.4.
Advantages and limitations of the three main die attach materials used in the opto-electronics industry. Cycling times are similar for all three materials.

Bonding material	Advantages	Limitations
Adhesive	Low curing temperature Low Young modulus (reduced stresses) Ease of automation	Rework not possible Large CTE Outgassing Low thermal and electrical conductivity Environmental sensitivity (moisture)
Solder	Good thermal and electrical conductivity Low CTE Rework possible	Require metalized surfaces Processing temperatures $>200°C$ Need an inert gas atmosphere
Glass	Good thermal and electrical conductivity Low CTE Limited stress relaxation	Processing temperatures $>300°C$ Rework is difficult

creep to minimize components shift over time and low stress-induced birefringence in the attached component (fiber, microlens, prisms).

19.2.2.3.1. Effect of Creep. In optical assemblies, there are very tight alignment requirements that do not allow extended movements of the semiconductor chip with respect to the optical fiber when heated or exposed to mechanical shock. Therefore bonding materials are supposed to take up the CTE mismatch, vibration and other stresses that may cause a die or a fiber to move. Moreover, the die attach material should exhibit low creep properties to maintain the optical alignment over time and under stress.

Driving forces of the creep are the internal and the external stress built-up during the solder cooling or as a result of adhesive polymerization and during device operation. The creep strain–stress relationship is determined by the material properties, i.e., by the microstructure and the diffusion properties of the solder alloy [40], or the free-volume ratio in adhesives [51]. Following Andrade's work on metals, the creep strain is a function of the applied stress (internal or external), σ, the time t and the temperature T:

$$\varepsilon = f(\sigma)g(t)h(T). \tag{19.5}$$

Several expressions can be found in the literature for each of these functions. A detailed description of the generally accepted expressions can be found in [34]. For example, the steady-state creep strain can be defined as:

$$\varepsilon = C\sigma^n t \exp\left(-\frac{\Delta H}{kT}\right), \tag{19.6}$$

where n, ΔH, k and C are the stress factor, the creep activation energy, the Boltzmann's constant and a material constant.

The stress factor n is usually a function of the stress level and temperature. For low stress conditions (typ. $\sigma \ll 1$ MPa) and high temperature, $n \sim 1$, i.e., the strain rate is

TABLE 19.5.
Creep parameters assuming a Norton-type stress dependence $d\varepsilon/dt = C \cdot \sigma^n \cdot \exp(-\Delta H/kT)$ for various solder alloys. ΔH is the creep activation energy and n is the stress factor at room temperature. The data for Sn-3.5Ag, Sn-9Zn, Sn-37Pb, Sn-2Cu-0.8Sb-0.2Ag are taken from the NIST database[a] and the parameters for the Sn-80Au are derived from [9].

Solder	C ($\mathrm{MPa}^{-n}\,\mathrm{s}^{-1}$)	Activation energy ΔH (eV)	Stress factor n	Melting temperature (°C)
Sn-3.5Ag	1.5×10^{-3}	0.825	11.3	221
Sn-9Zn	2.17×10^{-2}	0.677	5.7	199
Sn-37Pb	0.205	0.49	5.25	183
Sn-2Cu-0.8Sb-0.2Ag	3.031	0.85	8.9	285
Sn-80Au	5.29×10^{6}	1.24	7	280

[a] http://www.boulder.nist.gov.

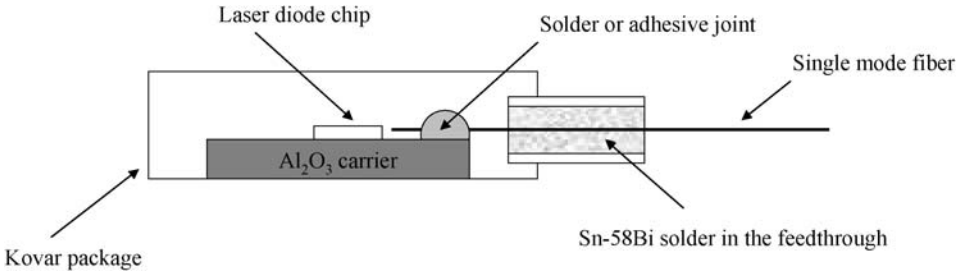


FIGURE 19.7. Assembly configuration used in the calculation of the creep in a solder and adhesive joint. The optical fiber is soldered or adhesively bonded on the laser carrier and is subsequently soldered in the fiber feedthrough with a bismuth-based solder.

proportional to the stress level. At higher stresses or lower temperatures, the stress exponent lies in the range of 3–20 depending on the material. This behavior is governed by a change in the dominant origin of the creep mechanism: diffusion-induced at low stresses or high temperature and dislocation induced at higher stresses or low temperature.

Table 19.5 lists the creep parameters for five common solders used in microelectronics and opto-electronics. These solders exhibit activation energies ranging from 0.6 eV to 1.3 eV, i.e., above the default activation energy specified in the Telcordia Generic Requirements GR 468[7] for module wear-out failure estimation (0.4 eV). It is noteworthy that the creep activation energy for all the listed binary solders increases linearly with the solder melting temperature.

In the following section, we compare the creep-induced displacement in a solder joint for different solder materials listed in Table 19.5 for the configuration sketched in Figure 19.7:

- An optical fiber is soldered on a planar substrate in front of a laser in a Kovar butterfly module. The mechanical properties of this solder are particularly critical as they

[7] GR468: generic reliability insurance requirements for optoelectronic devices used in telecommunications equipment, Bellcore, Issue 1, R4-62, 1998.

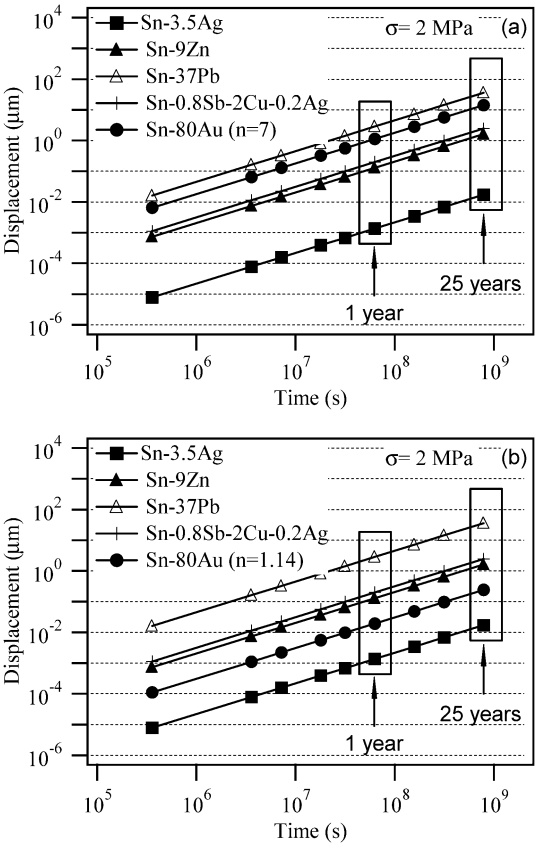


FIGURE 19.8. Solder joint displacement as a function of the time at room temperature using (a) $n = 7$ and (b) $n = 1.14$ for the gold-tin solder. From all solders listed in Table 19.5, Sn-3.5Ag exhibits the lowest creep compliance with an estimated displacement of approximately 20 nm after 25 years of device operation. Assuming a diffusion process as the dominant creep mechanism at room temperature for the Sn-80 Au, the displacement in the joint in the gold-tin solder is reduced from 14 μm down to 250 nm after 25 years of device operation.

define the evolution over time of the coupling efficiency between the laser beam and the optical fiber.

- Once properly aligned and soldered in front of the laser, the optical fiber is soldered at a second location in the fiber feedthrough of the butterfly module with a BiSn (58 wt% Bi) solder.

When the fiber is soldered in the module feedthrough, axial stresses are built up in the fiber when the assembly is cooled down from the manufacturing temperature ($T_m = 138°C$) and the room temperature T_a due to CTE mismatch between the housing, and the fiber substrate (the thermal expansion of the solder and the fiber itself is neglected). Assuming an alumina substrate and a fiber Young's modulus of 72 GPa, the initial axial stress in the fiber is estimated to be around 20 MPa. This corresponds to a force of 0.2 N acting on the solder joint. Based on a 2D modeling proposed by Suhir [57], the average shear stress in the solder is 2 MPa assuming a solder joint length of 500 μm.

The creep strain in the solder joint can be evaluated using Equation (19.6) for all the solders listed in Table 19.5. The results are displayed in Figure 19.8 as a function of

time. The solder Sn-Ag (3.5%wt Ag) exhibits the lowest creep-induced displacement with an estimated fiber offset of approximately 20 nm after 25 years of device operation.[8] For all the others investigated solders, the creep compliances are expected to adversely impact device reliability after a few months of the device operation. Indeed, the expected fiber offset is already as large as 1.5 μm with the PbSn (37%wt Pb) after 1 year of service, which is already larger than the positioning tolerances in most single-mode fiber applications. The estimated behavior of the gold-tin solder is not as good as expected. The stress factor for the gold-tin solder at room temperature, extrapolated from the data at temperatures above 100°C found in [9] could be slightly overestimated. If we assume now that the predominant creep mechanism in the gold-tin solder remains atomic diffusion and that the stress factor keeps the value measured at 100°C (value of 1.14), the expected ultimate fiber displacement is 10 nm and 250 nm after 1 and 25 years of device operation, respectively. If the stress level is increased from 2 MPa to 5 MPa, this displacement remains below 1 μm after 25 years.

The strain developed over time in the solder fixing the fiber inside the module depends on the operating temperature, through the temperature dependence of the stress acting on the solder and the temperature dependence of the creep in the solder itself. Using the Norton model to describe the stress dependence of the creep, we can define a temperature acceleration factor as:

$$a(T) = \frac{\varepsilon(T)}{\varepsilon(T_a)} = \left(\frac{T_m - T}{T_m - T_a}\right)^n \exp\left[-\frac{\Delta H}{k}\left(\frac{1}{T} - \frac{1}{T_a}\right)\right]. \tag{19.7}$$

In this particular configuration where the stress experienced by the solder decreases with temperature and vanishes at $T = T_m$, the acceleration factor is not a monotonically increasing function of the temperature. Indeed, this function goes through a maximum for $T = T_c$ and then decreases for $T > T_c$. T_c can be easily evaluated for each of the solders under investigation. The calculations give 30°C for Sn-3.5Ag, 58°C for Sn-9Zn, 53°C for Sn-37Pb, 46°C for Sn-2Cu-0.8Sb-0.2Ag solder and 76°C for Sn-80Au. The acceleration factor is plotted in Figure 19.9 as a function of the package temperature T for the five solder materials. In fact, in the case of the Sn-3.5Ag solder, the acceleration factor assumes a maximum value of 1.1 at 28°C and for all temperature above 40°C, the acceleration factor is below 1: increase in the module temperature above 40°C decreases the creep compliance.

Note that, in these calculations, we assumed a single external stress. Other stresses in the solder will affect the true creep compliance and its dependence on temperature. The effect of internal stresses is not considered here. The presence of voids in the solder is not considered either.

As adhesive bonding is used in optical assemblies, it is necessary to evaluate the creep strain in adhesive joints and to compare it with the data obtained with solder joints. Polymeric materials do not necessarily follow the creep strain–stress relationship defined in Equation (19.6). While the creep mechanism in solders is governed by grain size and atomic diffusion or dislocation phenomenon, the creep in polymeric materials is associated with a viscoelastic behavior where the free volume ratio has a determinant effect. The intermolecular space between long polymer chains (defining the free volume) allows for chain mobility over time in response to an imposed mechanical deformation. A decrease

[8]The extrapolation of the fiber displacement after 25 years is only given for design rule purposes as the behavior of the solder material will become nonlinear over time.

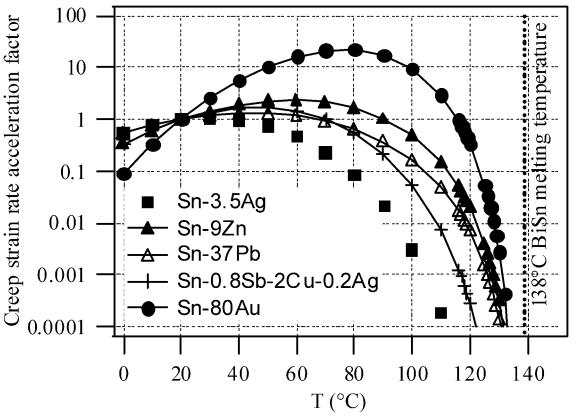


FIGURE 19.9. Steady-strain state acceleration factor as the function of the package temperature for Sn-3.5Ag, Sn-9Zn, Sn-37Pb, Sn-2Cu-0.8Sb-0.2Ag and Sn-80Au solders. The temperature dependence of the shear stress induced in the solder counteracts the Arrhenius term and strongly limits the usual adverse effect of the temperature on the creep compliances.

in the free volume yields a reduced chain mobility and a slower deformation under the imposed mechanical stress. In other words, the free volume directly impacts the time scale of a polymeric material strain–stress response. Moreover, the free volume and the T_g of a polymer are correlated: the larger the free volume, the smaller the T_g. Thus, the T_g and the viscoelastic properties of a polymer material are interrelated: the larger the T_g, the slower the viscoelastic deformation under a given mechanical load.

According to the Telcordia Generic Requirement GR1221,[9] adhesives used in optical assemblies should have a T_g above 95°C in order to limit joint deformations, i.e., components movements, under internal or external stress. However, published data on the creep strain of structural adhesive joints in optical assemblies are rather limited. Plitz et al. [46] have estimated the influence of post-curing aging on various mechanical properties, including CTE and creep compliances. Reith et al. [52] have evaluated the dimensional stability of adhesives in optical connectors ferrules and its influence on optical performance. They have shown that, although all the tested adhesives cannot restrain a permanent fiber pushback after a mechanical loading, fibers adhesively bonded into ferrules using high T_g adhesives ($T_g \sim 120$°C) exhibited a reduced fiber displacement of 10 nm after 400 hours testing. This displacement slightly increased to 15 nm when the test was performed at 65°C, however this pushback was not detrimental to the connector optical performances.

A rough estimation of the influence of the T_g on the creep compliance of optical adhesives can be performed using the data from [46]. They have evaluated the creep compliance at 40°C over time for two commercial UV adhesives having different T_g. After an initial time period where the creep-induced deformation steeply increases, both adhesives exhibited a steady-state creep compliance. The mechanical load applied on the bonded parts under investigation is small enough (0.3 MPa), so that a linear stress dependency model can be assumed. Assuming a similar assembly geometry as in the calculation of the creep strain in solders (Figure 19.7) and a similar strain–stress model than for solder materials, we can estimate the influence of the T_g on the creep displacement in an optical fiber bonded with an adhesive material. In order to account for the dependence of the creep properties on T_g,

[9] GR1221: generic requirements for passive optical components, R4-24, Issue 2 (1999).

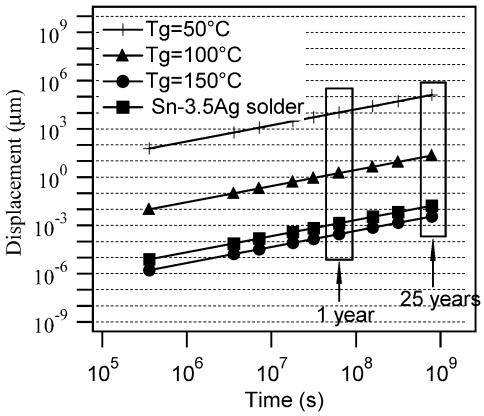


FIGURE 19.10. Adhesive joint displacement as a function of the adhesive glass transition temperature T_g. The results are given for an operating temperature of 40°C. In order to reach the performances of the Sn-3.5Ag material, the T_g of the adhesive should be above 150°C.

we assumed that the creep activation energy ΔH increased linearly with T_g and compared the displacement in the joint over time for adhesives having a $T_g = 50$°C, $T_g = 100$°C and $T_g = 150$°C. The obtained results are shown in Figure 19.10. For comparison, the creep displacement in a Sn-3.5Ag solder at 40°C is also presented. We see that, in order to limit the displacement to an acceptable level (a fraction of a micron after 2 years), the T_g of the adhesive should be around 120–130°C. Again, the extrapolation of the fiber displacement after 25 years is only given as an illustration. It is expected that the behavior of the adhesive material becomes nonlinear over time.

19.2.2.3.2. Stress-Induced Birefringence. An adhesive or a solder joint can affect the optical properties of a bonded optical element through the build-up of a mechanical stress-induced birefringence in the optical component material. Whenever the polarization of the light propagating through the bonded element is critical, this stress should be controlled in order to obtain the desired state of propagation at the output of the element. For example, to achieve polarization-independence gain in semiconductor laser amplifier after flip-chip soldering, it has been shown that an additional bulk tensile strain needs to be preset into the semiconductor material when the component is soldered p-down on the substrate [19].

Other examples of mechanical stress-affected birefringence include:

- Optical fiber-based temperature sensor [25]. In this apparatus, the birefringence of the optical fiber, hence the state of polarization of the outcoming light, depends on the temperature-sensitive mechanical stress induced by the CTE mismatch between the glass fiber and an adhesive surrounding the fiber in a capillary tube. The sensitivity of the sensor depends on the CTE of the adhesive, the original birefringence of the fiber, as well as the length of the capillary tube.
- Reduction of birefringence in laser gyroscopes prisms using a soft indium solder as a bonding material [27]. The use of a soft indium solder reduces the optical birefringence at the basis of the prisms by a factor of 2–3 compared to an optical contact configuration.

19.3. ADHESIVE BONDING IN OPTICAL ASSEMBLIES

Common attach materials include solder alloy materials, adhesives and glass solders. Adhesive bonding provides advantages over other bonding techniques used in optoelectronic assemblies:

- ability to bond dissimilar materials,
- low processing temperature,
- refractive index-matching properties,
- no metalization required on the parts (as e.g., in the case of soldering or welding).

On the other hand, challenges in adhesive bonding include:

- limited operational temperature range,
- sensitivity to moisture,
- rework not possible,
- outgassing.

We will see in this paragraph that the main parameters that determine the performance of an adhesive bond are similar to the others bonding materials, i.e.,

- Surface preparation of the adherend.
- Physical and thermal properties of the adhesive (T_g, CTE, modulus).
- Joint design and shrinkage control.

We will not consider here electrically conductive adhesives. Their use and their properties have been described in [38].

19.3.1. Origin of Adhesion

Several theories of adhesion describe different types of adhesive bonding [28,31]: adsorption, chemical bonding, diffusion, mechanical interlocking, electrostatic bonding. However, no single model is able to explain a specific adhesive bond. The properties of an adhesive joint usually reflect the interplay of the effects described by various models. Nevertheless, it is clear that adherend–adhesive interface properties, in particular intermolecular forces at the interface, have a major contribution to the strength, as well as to the environment-induced degradation, of adhesive bonds.

19.3.1.1. Adsorption Adsorption is responsible for adhesion when intermolecular attractive forces, Van der Waals forces (electric dipole interactions) and/or hydrogen bonds [21], build up between the adhesive and the adherend. It is often believed that it is the most important adhesion mechanism and it has been experimentally demonstrated that the mechanism of adhesion in many adhesive joints only involves these interfacial secondary forces [5]. In most cases, adsorption is the relevant model when using adhesives in optical assemblies.

19.3.1.1.1. Wetting. According to this mechanism, the wetting of the adherend by the adhesive is a key factor in determining the bond strength. Wetting is defined as the tendency of a liquid to spread over a surface. The wetting process is controlled by three parameters: surface energy of the adherend γ_{SV} (solid–vapor interface), surface tension of the adhesive γ_{LV} (interface liquid–vapor) and the interfacial surface energy between the adherend and the adhesive γ_{SL}. The surface energy of a medium is defined as the free energy change when the surface area of this medium is increased by unit area. In other words, the surface

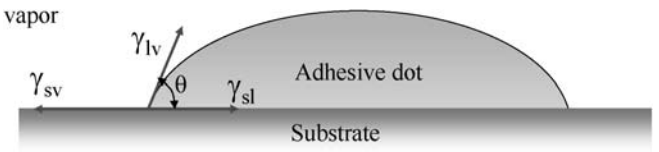


FIGURE 19.11. Adhesive drop resting at equilibrium on a solid surface. γ_{SV}, γ_{LV} and γ_{SL} are the surface energy of the adherend (solid–vapor interface), the surface tension of the adhesive (interface liquid–vapor) and the interfacial surface energy between the adherend and the adhesive. θ is the contact angle.

tension of the liquid γ_{LV} defines the energy required to create a new liquid surface area by moving molecules from the bulk liquid to the surface. The greater the surface tension, the greater is the energy needed to convert bulk molecules into surface molecules. Similarly, the surface energy of a solid γ_{SV} is equivalent to half the work of cohesion of this solid.

Young's equation [31] relates these three parameters at the three phase contacts to the equilibrium contact angle θ (Figure 19.11):

$$\gamma_{SV} = \gamma_{SL} + \gamma_{LV} \cos\theta. \tag{19.8}$$

When dispensing a liquid onto a substrate, the following equilibrium configurations can be reached:

- $\theta = 0°$: spontaneous spreading.
- $0° < \theta < 90°$: partial wetting.
- $90° < \theta < 180°$: partial non-wetting.
- $\theta = 180°$: negligible wetting.

For spontaneous spreading to occur, we need:

$$\gamma_{SV} > \gamma_{SL} + \gamma_{LV}. \tag{19.9}$$

By ignoring the interfacial free energy, Sharpe and Schornhorn [54] have proposed the following criteria:

For good wetting:

$$\gamma_{SV} > \gamma_{LV}. \tag{19.10}$$

For poor wetting:

$$\gamma_{SV} < \gamma_{LV}. \tag{19.11}$$

In other words, in order to wet the adherend, the surface tension of the adhesive must be lower than the surface energy of the adherend.

The surface tension of the adhesive is a given parameter in a dispensing process and typically ranges from 25 to 50 mN m^{-1}. It is not possible to change it without affecting the properties of the adhesive. Thus, in order to improve the adhesive wetting properties, it is advisable to look for an appropriate surface treatment to increase the surface energy of the substrate.

Table 19.6 presents the variation of the diameter of adhesive (acrylate) dots dispensed on a silicon substrate with the surface treatment. The raw silicon substrate exhibits a surface

TABLE 19.6.
Correlation between surface energy and dispensed dot size for various silicon surface preparations. The adhesive used in this experiment was an acrylate. The higher the surface energy of the silicon, the larger the adhesive dot due to improved wetting properties of the silicon surface.

Cleaning process	Surface energy ($mJ\,m^{-2}$)	Dot size (μm)
Raw silicon substrate	42 ± 4	130 ± 10
2-propanol	52 ± 4	170 ± 10
O_2 plasma	>105	200

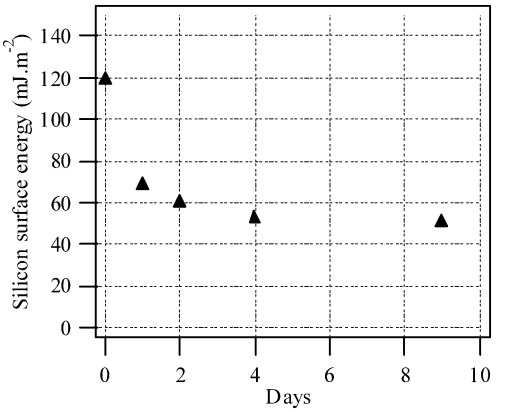


FIGURE 19.12. Evolution of the surface energy of an oxygen-plasma treated silicon substrate over time in ambient atmosphere.

energy as low as 42 ± 4 mJ m^{-2}. Two surface treatments were evaluated in these experiments: a wet cleaning with a 2-propanol solvent and a O_2 plasma etching. Using 2-propanol as a cleaning solution, the surface energy of the substrate and the diameter of the adhesive are increased by 25% and 40%, respectively. After an oxygen plasma treatment, the surface energy of the silicon substrate is higher than 105 mJ m^{-2}, and the diameters of the adhesive dot are increased by more than 50% compared to the uncleaned raw material. These results clearly demonstrate that the higher the surface energy of the silicon, the larger the adhesive dot due to improved wetting properties of the silicon surface.

Once properly treated, the substrates should be kept in a non-contaminating environment. Figure 19.12 shows the evolution of the surface energy of silicon substrates stored in ambient atmosphere. After two days, the substrate has lost half of its initial surface free energy through most probably the adsorption of the water molecules.

19.3.1.1.2. Kinetics of Wetting. The viscosity of a liquid arises from the intermolecular forces and steric-induced anchoring effects. The stronger the forces hindering the motion of the molecules, the higher the viscosity. Typical intermolecular forces include hydrogen bonding and dipole–dipole interactions. Hydrogen bonding accounts for the high viscosity of water compared to the aromatic benzene molecule, where there is no hydrogen bonding. Glycerol ($C_3H_7O_3$) as well is very viscous owing to the number of hydrogen bonds its molecule can form. Heavy hydrocarbon oils, which are not hydrogen bonded, are also

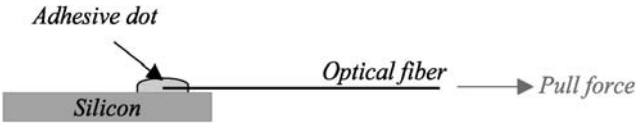


FIGURE 19.13. Description of the mechanical test assembly.

viscous. Their viscosity arises partially from the dipole interaction between molecules as well as steric effects (the long chainlike molecules get tangled to each other). On the other hand, the surface tension of a liquid is only determined by the intermolecular forces. These forces counteract the wetting of the liquid on a substrate. Steric-related effects have no influence on liquid surface tension.

There is no relationship between viscosity and surface tension of liquids. In other words, decreasing the viscosity will not decrease the contact angle nor affect adhesive dot size, but will speed up the kinetics of wetting [10].

The surface morphology of the substrate affects the kinetics of wetting as well. A liquid having a contact angle above 90° will spread along fine pores, scratches and other inhomogeneities by capillary action, even if it does not wet spontaneously a planar surface. It has been reported that random surface scratches can increase the spreading rate of liquids by as much as 50% [6]. This spreading is also observed when dispensing adhesive in V-grooves or U-grooves etched in silicon for passive alignment of optical fibers. The integration of larger sections along the main groove will prevent the adhesive flow down to the fiber tip owing to a decrease of the capillary pressure driving the adhesive flow and a reservoir-like functionality.

19.3.1.1.3. Effect of Surface Energies on Bond Strength in Optical Assemblies. The adhesive joint strength depends on the ability of the adhesive to spread spontaneously on the substrate [54]. Thus, the adhesive strength is reduced by the presence of contaminants, including hydrocarbons or moisture. The mechanical and environmental resistance of an adhesive joint will be improved through the prior application of dedicated surface treatments: solvent cleaning, wet chemical etching, plasma cleaning, UV radiation, silane adhesion promoter, ion-beam, laser surface treatment [31].

The adhesive bond strength can be assessed by a mechanical shear test. In order to evaluate the influence of adherend handling and surface preparation in passive optical components assemblies, stripped Corning SMF28 fibers have been bonded to silicon submounts and the joint strength has been measured through the application of a tensile stress on the free hanging part of the fiber until fracture. The tensile stress applied on the fiber translates into the shear stress in the adhesive. The test set-up is shown in Figure 19.13. To fulfill the Bellcore GR468,[10] such assemblies must be able to withstand 0.8 GPa tensile stress (120 kpsi or 1 kg load). The graph in Figure 19.14 shows three different failure cases of pull-tested adhesive-bonded fibers that are observed depending on adherend preparation:

- Adhesive failure of the adhesive joint on the silicon substrate [(Figure 19.15(a)]. Such failures occur for relatively low stress values (average value: 200 MPa tensile stress) when untreated silicon substrates are used.
- Fiber cohesive failure [(Figure 19.15(b)]. The applied tensile stress at failure is 500 MPa on average. This is typically observed when the silicon substrate has been

[10]GR468: generic reliability insurance requirements for optoelectronic devices used in telecommunications equipment, Bellcore, Issue 1, R4-40, 1998.

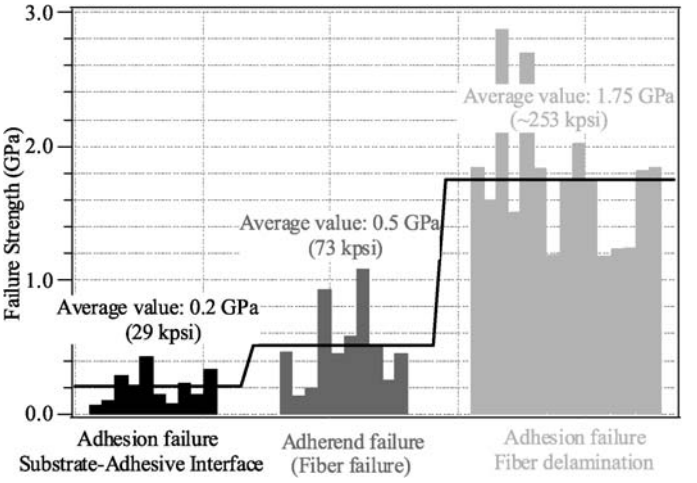


FIGURE 19.14. Tensile stress failure results on adhesive bonded fibers on silicon submounts. Results scattering is related to slight variations in adhesive dot size.

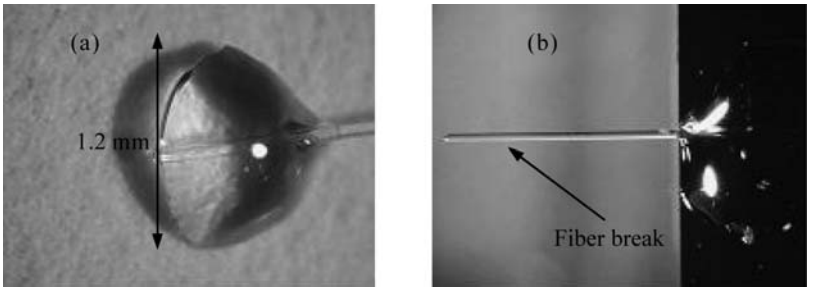


FIGURE 19.15. (a) Adhesive failure of the adhesive bond: the adhesive dot lifts off the silicon submount. (b) Fiber failure: the glass fiber broke at mid-span (here 1.5 mm away from the adhesive dot).

oxygen plasma treated and the stripping of the fiber has introduced micro-cracks at the bare fiber surface.

- Fiber delamination (adhesion failure at the adhesive–fiber interface). The applied tensile stress in this case reaches 1.75 GPa on average. This is typically observed with plasma treated silicon substrates and a proper fiber handling.

The Bellcore requirements are met only in the third case.

The low-stress adhesive failure case in Figure 19.15(a) is explained by a low surface energy value of the silicon submount (28 ± 5 mJ m^{-2}). This value is related to the presence of hydrocarbons or water on the surface of the substrate. After Oxygen plasma treatment of the substrates, surface energy of the silicon submount is above 105 mJ m^{-2} and subsequent pull test experiments with treated samples lead to failure cases 2 and 3 only (fiber cohesive failure or fiber delamination). Additional experiments showed that the transition between failure cases 1 and 2 or between 1 and 3 with 1 mm diameter dot occurs when the surface energy of the silicon submount is 48 ± 5 mJ m^{-2}. Fiber cohesive failure in Figure 19.15(b) is avoided with a proper handling of the fiber during and after stripping. In particular, we have seen that thermo-mechanical stripping compares favorably to standard mechanical

stripping as micro-cracks at the surface of the glass appear using purely mechanical stripping techniques. The stripped length and the cleaning process must be carefully optimized to avoid subsequent mechanical contact on the bare fiber.

19.3.1.2. Mechanical Interlocking The mechanical interlocking model states that, when the surface of the adherend exhibits pores, holes and other irregularities, adhesive bonding can be enhanced through the mechanical interlocking of the adhesive and the adherend material. The adhesive should not only wet the substrate, but also have the right rheological properties to penetrate pores and openings. Since good adhesion can occur between smooth adherend surfaces as well, it is clear that while interlocking helps promote adhesion, it is not really a generally applicable adhesion mechanism. Pre-treatment techniques resulting in microroughness on the adherend surface can improve bond strength and durability [13]. Indeed, a larger contact area resulting from the roughening of the adherend surface contributes to the enhancement of the adhesive joint strength. Additionally, according to [11], the influence the bond strength depends not only on the contact angle and surface energies but also on the kinetics of wetting. The roughening of the substrate, speeding up the adhesive spreading, would then contribute to an increase of the bonding strength. However, this theory has not received a strong echo from adhesion scientists.

In most cases, the adhesive bonds found in the optical assemblies can be explained with the adsorption theory or the mechanical interlocking phenomenon or both. However, in order to explain all possible adhesive bond configurations, e.g., a plastic optical component on a plastic substrate, the models described in the following section may be necessary.

19.3.1.3. Other Models: Chemical Bonding, Diffusion Bonding and Electrostatic Bonding *Chemical bonding* is responsible for adhesion when, in addition to an adsorption mechanism, there is a surface reaction, i.e., establishment of primary chemical bonds (covalent or ionic). Primary chemical forces have energies ranging between 60–1100 kJ/mol, which are considerably higher than the secondary bond energies have (0.08–5 kJ/mol) [31]. Chemical bonding will be the primary adhesion mechanism when silane-based adhesion promoters are used before application of the adhesive material. The chemical reactions occurring at the interface through the use of silane coupling agents have been reviewed by E. Plueddemann [49].

The diffusion bonding theory predicts a diffusion of molecules across the interface when the adherend and the adhesive have mutual solubility [63]. This theory may apply when both the adhesive and the adherend are polymers (e.g., when the optical element is bonded on a plastic substrate). The strength of the adhesive bond is related to the extent of the interdiffusion across the interface. The diffusion theory, however, is not justified where the adherend and adhesive are not soluble or when chain movement of the polymer materials is constrained by its highly crosslinked structure, or when it is below its glass transition temperature.

Electrostatic bonding is related to the formation of an electrical double layer of charges of opposite sign across the interface when the adherend and the adhesive have permanent electrical dipole moments or polar molecules. There are still some controversies around this theory because the electrical double layer cannot identified without separating the adhesive bond.

19.3.2. Adhesive Selection and Dispensing

19.3.2.1. Adhesives Detailed formulations of adhesives are usually proprietary information of the manufacturers and not available to the end-user. However, the active agent re-

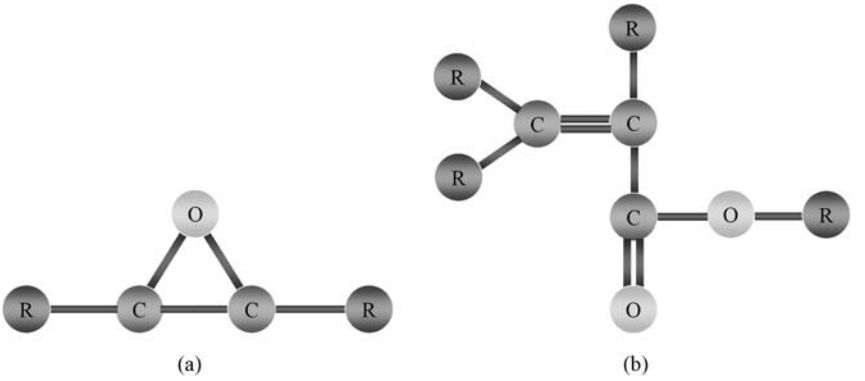


FIGURE 19.16. Chemical structures of (a) an epoxy functional group, (b) acrylate functional group.

sponsible for the chemical activity of the material is usually specified. In the fiber-optics industry, the adhesives are mostly based on either an *epoxy* or an *acrylate* group. The polymerization process is either light-based (UV or visible), heat-based or both.

The term "epoxy" refers to a chemical group consisting of an Oxygen atom bonded to two carbon atoms forming a ring structure. The simplest epoxy is a three-member ring structure known by the term "alpha-epoxy" or "1,2-epoxy" [Figure 19.6(a)]. *Thermal epoxies* are usually two component adhesives: the epoxy resin and the hardener (curing agent). The hardener, often an amine, is used to cure the epoxy by an addition reaction where two epoxy groups react with each amine site. This forms a complex three-dimensional molecular structure. Since the amine molecules co-react with the epoxy molecules in a fixed ratio, it is essential that the correct mix ratio is obtained between resin and hardener to ensure that a complete reaction takes place. If amine and epoxy are not mixed with the right stoichiometry, unreacted resin or hardener will remain within the matrix which will affect the final properties after cure. One component thermal epoxies are also available today. However, as they typically require 150°C curing temperature during one hour soaking time, they are cumbersome to use in low alignment tolerance optical assemblies.

The term acrylate refers to a chemical group consisting of a carbon–carbon double bond bonded to an ester functional group COOR [Figure 19.16(b)]. The resin base consists of a light molecular weight polymer (oligomer) having one or several acrylate functional groups. As opposed to epoxies, acrylates are polymerized with a catalyst rather than a hardener so that the curing proceeds as a chain reaction rather than an addition reaction. Typically, peroxides ROOR′ are added to provide the resin with a source of free radicals. Upon curing, the peroxide in the resin base decomposes to yield free radicals RO. These radicals then initiate polymerization through the condensation of acrylate groups on the resin oligomers.

In UV curable adhesives, the polymerization proceeds in a chain mechanism involving cationic (epoxies) or free radical (acrylate) intermediates generated through the photolysis of a photoinitiator. The curing process of UV curable adhesives is fast, making them well suited for optical assemblies when the alignment between parts must be guarantied during the polymerization process.

19.3.2.2. Critical Parameters When selecting an appropriate adhesive for an optical assembly application, the following parameters should be considered: T_g, CTE, propensity to creep, shrinkage, index-matching properties and photostability as well as moisture resis-

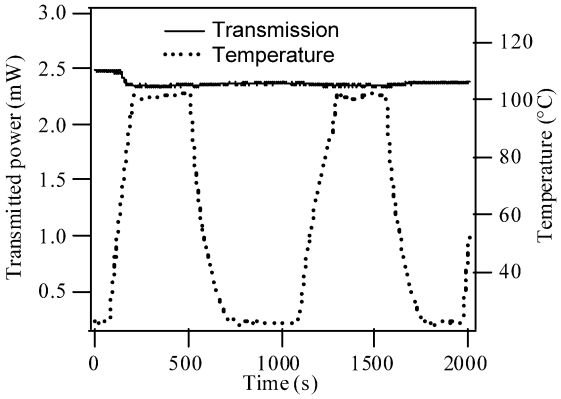


FIGURE 19.17. Thermal cycling test for a fiber-to-fiber assembly on a PEEK substrate. No change in coupling efficiency is observed up to 100°C.

tance and outgassing properties. In this section, we review the importance of each of these parameters.

19.3.2.2.1. CTE, Creep and Glass Transition Temperature. The glass transition temperature T_g is the temperature at which the material changes from a hard, glassy substance to a soft, rubbery one. In practice, this material transition occurs over a wide temperature range (up to 50°C) and only the central temperature is given on adhesive datasheets. CTE increases for $T > T_g$ whereas the Young's modulus and the hardness decrease for $T > T_g$.

To fulfill the Telcordia Generic Requirement GR1221, adhesives used in structural assemblies should have a T_g above 95°C measured by Differential Scanning Calorimetry (DSC). Indeed, the thermal expansion and the creep of adhesives, both T_g dependent,[11] can induce dimensional instabilities and ultimately misalignment between adhesive bonded optical devices.

The T_g of a UV adhesive is mainly defined by the adhesive temperature during curing: if the maximum temperature during curing is 60°C, the T_g of the UV-cured adhesive will be approximately 60°C. However, the T_g of this UV-cured adhesive can be further increased through, e.g., thermal post-curing. UV-curable adhesives were investigated with respect to thermal post-curing and accelerated aging [46]. Although the post-curing induced some shrinkage as well as some degradation, it was found that the probability that these devices exhibit dimensional instabilities was reduced due to an increase of the glass transition temperature. We have also performed thermal cycling tests on fiber-to-fiber assemblies. The results are displayed in Figure 19.17. The transmitted power from the incoming to the outcoming fiber is stable up to 100°C once the first temperature ramp-up is passed. This curve shows that a proper selection of the adhesive and a good bond design can lead to a stable optical coupling between room temperature and 100°C.

19.3.2.2.2. Shrinkage. The shrinkage of an adhesive joint is defined as the reduction of its linear dimensions and its volume during polymerization. Typically, the linear shrinkage of epoxies and acrylates is around 0.5% and 1.5% respectively. The shrinkage in optical assemblies results in a displacement between devices arising during the adhesive polymerization and a build-up of shear stresses in the adhesive joint leading to a creep-related misalignment over time.

[11] See Section 19.2.2.3.1 for the influence of T_g on the creep of adhesives.

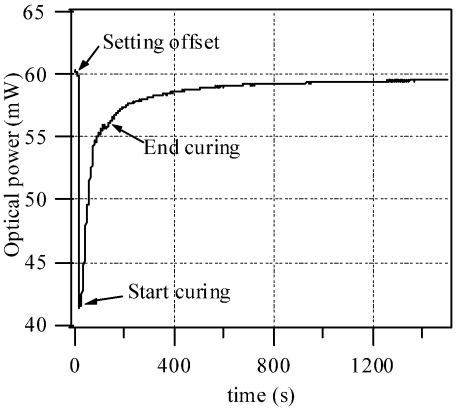


FIGURE 19.18. Shrinkage compensation using a fiber position offset.

The displacement occurring during polymerization can be corrected beforehand with an appropriate component offset. By this means, optimum alignment with minimized insertion losses is obtained after polymerization. Figure 19.18 presents the case of a laser to fiber pigtailing application where the adhesive bonded fiber was slightly offset from maximum optical coupling before starting the polymerization process. A UV epoxy was used in this experiment. The optical losses introduced by the offset are recovered during the epoxy polymerization. It is noteworthy that the shrinkage, or in other words the polymerization process, continues after switching off the UV light. Indeed, cationic species released during the UV curing have not reacted yet when the UV light is switched off but they remain available for further chain reactions.

In order to minimize shrinkage-related creep effects, it is necessary to limit the shrinkage-induced shear stress in the adhesive joint. Similarly to the CTE mismatch-induced thermo-mechanical stresses and assuming that the shrinkage-induced stress does not relax during the polymerization process, it is possible to introduce a maximum shrinkage-related shear stress σ_s:

$$\sigma_s = \frac{\Delta l}{l} G_D \frac{\tanh \beta L}{\beta t_D} = s G_D \frac{\tanh \beta L}{\beta t_D}, \tag{19.12}$$

where β is defined in Equation (19.3) and s is the linear shrinkage (in %) of the adhesive upon polymerization.

As an example, we assume a chip length of 2400 μm and an adhesive thickness of 20 μm. The calculation of the maximum shrinkage-induced shear stresses in the bondline of this chip on submount assembly gives 126 MPa and 204 MPa in the epoxy and acrylate cases (see in Table 19.3 for adhesive material parameters). These stresses are not negligible compared to the CTE mismatch-induced stresses (e.g., 80 MPa for a GaAs chip on a silicon substrate upon soldering, see Section 19.2.2.2.1) and can readily influence the strength of the materials and the creep strain in adhesive bonded assemblies.

19.3.2.2.3. Outgassing. Outgassing of adhesive materials induce organic contamination on optically active parts, e.g., mirrors, lenses, fiber endfaces or laser facets and introduce a change in the response function of optical coating. The NASA has compiled outgassing data of adhesives intended for spacecraft applications. The method is based on the Total Mass Loss (TML) and Collected Volatile Condensable Materials (CVCM) measured in a

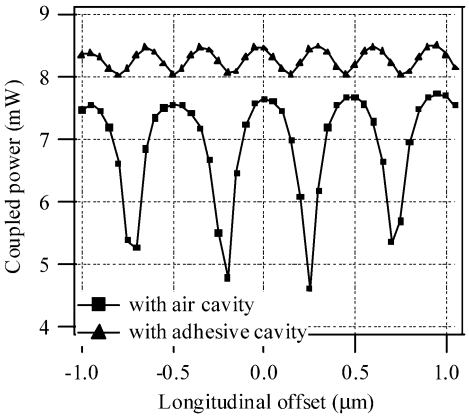


FIGURE 19.19. Fabry-Pérot interferences in the air cavity between a semiconductor laser and a fiber tip.

vacuum environment. Adhesive samples are heated to 398 K (125°C) for 24 hours. This causes the volatile materials to be driven out. The mass loss of the sample is determined from the weights before and after the 398 K exposure, and the percentage loss is calculated to provide the TML. The data can be used as a guide in selecting low-outgassing materials for optical applications.[12]

As a preliminary analysis, we verified that the facets of laser chips were not contaminated by the outgassing of the adhesive during the UV curing step and a subsequent burn-in procedure. Adhesive dots have been dispensed in front of the laser chips. The adhesive dots were UV cured and the assemblies were then submitted to a burn-in step. The analysis of the facet of the lasers was done using a Scanning Electron Microscope. None of the laser chips exhibited an organic contamination of the facets [48].

19.3.2.2.4. Refractive-Index Matching and Photostability. Adhesives usually have a refractive index between 1.4 and 1.6. Therefore, they can be used to reduce Fresnel reflection between optical devices, e.g., between a laser or a polymer waveguide and a glass optical fiber ($n = 1.46$ at a wavelength of 1.55 μm).

Figure 19.19 presents the variation of the coupled power in a laser-to-fiber coupling experiment as a function of the distance between the laser facet and the tip of the fiber. When the medium between the laser die and the optical fiber is air (squares data point), the reflection coefficient at the air–fiber interface and at the air-laser is as high as 4% and 13% respectively. Fabry-Perot interferences build up in this air cavity, generating coupled power variations as large as 35% when the longitudinal distance is varied. These oscillations limit the positioning tolerances of the optical fiber along the optical axis to less than 250 nm. When an adhesive is filling the gap between the laser die and the optical fiber, Fresnel reflections at the adhesive–optical fiber interface fall below 10^{-3} and the Fabry-Pérot interferences are strongly reduced (triangles data point). In this configuration, the influence of the thermal expansion of materials during device operation is not detrimental as far as the longitudinal displacement is concerned.

In addition to the minimization of Fresnel reflections, an adhesive, when used as a filling material between an optical component and an optical fiber, influences the coupling efficiencies and alignment tolerances due to improved mode matching. Figure 19.20 shows

[12] http://outgassing.nasa.gov.

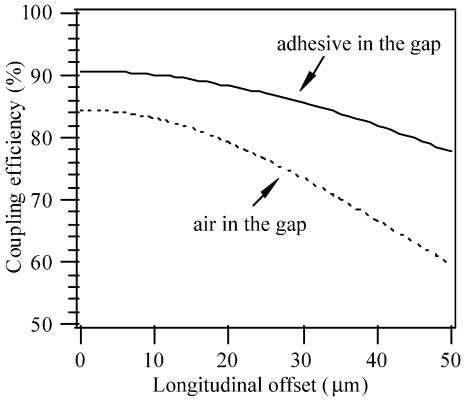


FIGURE 19.20. Comparison of the coupling efficiencies of a single-mode fiber to a 5×5 μm polymer channel waveguide with and without adhesive in the gap obtained in a end-fire coupling configuration.

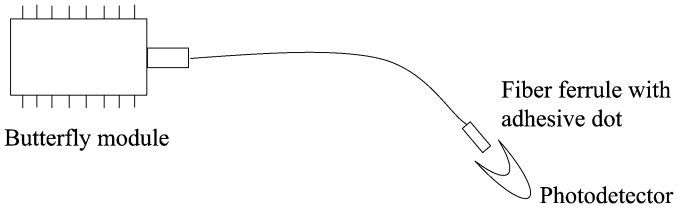


FIGURE 19.21. Schematic representation of the photostability measurement set-up. The light from the module passes through the adhesive dot on the fiber ferrule tip and is detected by the photodetector.

simulations of the optical coupling of a single-mode fiber to 5×5 μm channel waveguides fabricated on a silicon wafer from a fluorinated acrylate polymer with refractive indices of core and cladding of 1.47 and 1.46, respectively. We clearly see that the coupling efficiencies are larger when the gap between the device and the optical fiber is filled with an adhesive material: 91% vs 85% at a longitudinal distance of 1 μm and 77% vs 60% at a longitudinal distance of 50 μm. Therefore, if the photochemical and photomechanical stability is guaranteed, filling the gap between the waveguide and the fiber with an adhesive can be very favorable.

In order to assess the reliability of the gap filling approach, we have performed photostability experiments for various adhesives at 980 nm and under high intensity conditions (from 600 to 900 kW cm^{-2}):

- The organic material to be evaluated was directly deposited on the ferrule tip of a pigtailed module.
- The material is cured according to the polymerization conditions given by the manufacturer.

The module is powered on and light passes through the organic material dot. The transmitted light is detected with a photodetector and recorded versus time (Figure 19.21).

The results of the photostability experiments are summarized in Figure 19.22. After 1000 hours, the optical transmission is still above 0.98 for both adhesives A and B. Similar results were reproduced on 2 other samples for both adhesives. We can then conclude that the adhesives A and B exhibit good photostability properties at 980 nm at intensity levels

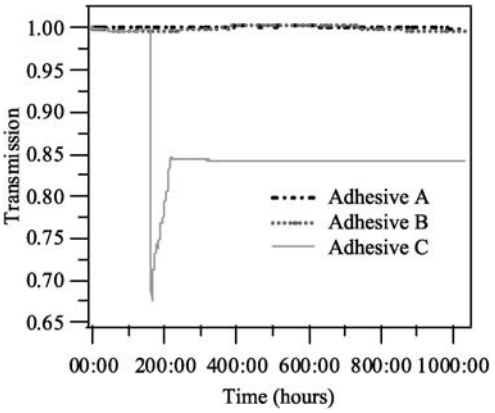


FIGURE 19.22. Evolution of the transmitted optical power over time for three different adhesives. The results are normalized relative to the value of the transmitted power at $t = 0$.

in the range of 600–900 kW cm^2 and could be used as gap filling material under similar stress conditions for low-end products. It is clear however that, in order to evaluate the risk of adhesive-related catastrophic optical damage on the laser chip itself, extended tests with the adhesive between the laser facet and the optical fiber should be performed.

The evolution of the optical transmission with adhesive C presents a steep decrease after 160 hours. The transmission value stabilizes then around 0.85 for the remaining lifetest time. A visual inspection of the adhesive dot at ferrule tip showed that this adhesive material presented cracks and "bubbles." As this behavior was confirmed on 3 other samples, we conclude that adhesive C cannot be used as an index matching material or as a fiber fixing material in 980 nm high power lasers applications.

It is clear that similar test experiments should be reconducted if the operating wavelength is different or if the intensity is higher.

19.3.2.2.5. Effect of Roughness on Moisture Resistance in Optical Assemblies. Moisture absorption is measured in terms of the percentage weight increase of the material caused by water absorption when placed under water or in a highly humid atmosphere for a given period of time. The effect of the adhesive moisture intake is twofold:

- decrease of the adhesive bonding strength,
- release of trapped moisture in sealed packages.

The danger of moisture absorption is significant in hermetically sealed packages when the moisture trapped in the adhesive is released during device operation, building a corroding atmosphere in the package. A corroding atmosphere is detrimental for the material strength of optical fibers [24,39] and electrical interconnects (wires bonds and metallic pads).

Ingress of moisture into the bondline (between the adhesive and the adherend) is the main source of moisture-related adhesive strength degradation and adhesive failure. In order to improve the moisture resistance of the adhesive bond, the properties of the adherend surface should be carefully investigated. In particular, adherend surfaces providing a high density of physical bonds (mechanical interlocks) show a better moisture resistance than smooth surfaces. Figure 19.23(a) and (b) shows the cases of two adhesive dots placed during 24 hours in deionized water. In both cases, the adhesive has been dispensed on silicon. The major difference between both adherends is their surface profile: polished smooth silicon surface in Figure 19.23(a) and non-polished silicon surface in Figure 19.23(b). The

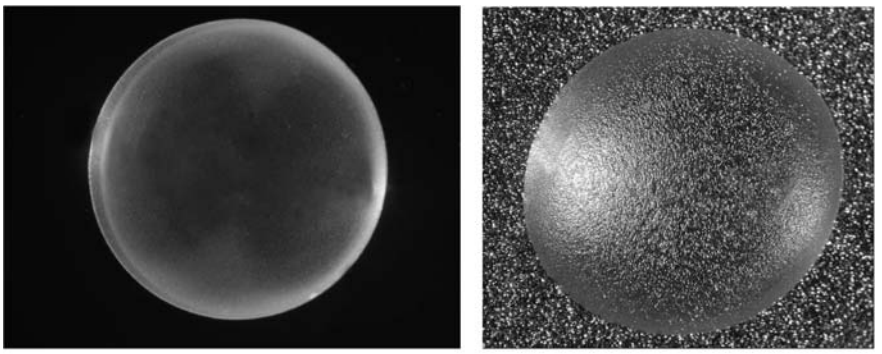


FIGURE 19.23. Adhesive dots after 24 hours dipped in deionized water. The adhesive was dispensed on (a) a polished silicon substrate (b) a non polished silicon substrate. The dots diameter is approximately 2 mm.

adhesive dot of Figure 19.23(a) exhibits a diffusive aspect over the complete dot area. On the other hand, the adhesive dot of Figure 19.23(b) presents a diffusive ring on the external part only of the dot area. Two general conclusions can be drawn from these pictures:

- The moisture ingress is not a volume effect. If it were the case, both adhesive dots would exhibit the same diffusive aspects.
- The location of the structural changes on both pictures is at the adherend–adhesive interface: the moisture ingress is growing from the circumference to the center.
- The substrate in Figure 19.23(b) exhibits a better moisture resistance than the substrate in Figure 19.23(a).

The roughness of the substrates used in these experiments has been measured using Atomic Force Microscopy. The scans performed on both adherends are presented in Figure 19.24. The polished silicon in Figure 19.23(a) and Figure 19.24(a) exhibits no evolved microroughness: the local roughness on a 2.6 μm × 2.6 μm sample is 2 nm and a long-range large (∼30 nm) smooth height variation can be observed. The non-polished silicon in Figure 19.23(b) and Figure 19.24(b) exhibits a local roughness of 8 nm and a long-range, large (∼500 nm) height variation resulting from randomly oriented but regular saw teeth profiles.

The adhesive bonding on this polished silicon substrate mainly relies on molecules adsorption, i.e., secondary bonds that are readily disrupted by moisture. The ingress of moisture in the bond line is slowed down when, in addition to adsorption, the adhesive bonding results from mechanical interlocking.

19.3.3. Dispensing Technologies

Bonding micro-optical elements including optical fibers requires the capability of dispensing adhesive volumes in the nanoliter or even picoliter range. Let's consider a 3 mm long U-groove designed in such a way that the core of the optical fiber lying in the groove is in the plane of the silicon wafer surface. The volume of the U-groove is 23 nl and the volume of the fiber itself is 18 nl, leaving an maximum available volume of 5 nl for the adhesive. Most adhesive dispensers cannot repeatedly deliver such small amount of adhesive. Thus, unless adhesive overflow (in particular on the fiber endface) is not detrimental

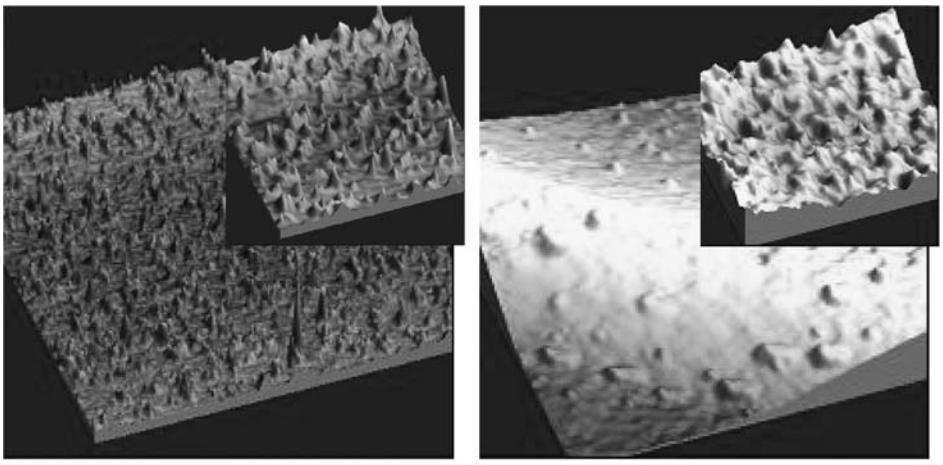


FIGURE 19.24. AFM two-dimensional 2.6 μm × 2.6 μm scans of (a) polished silicon substrate having a local roughness of 2 nm and long-range height variation of 30 nm (b) non-polished silicon substrate having a local roughness of 8 nm but "regular" long-range height variation of 500 nm. The inset pictures are zoom-in over 1 μm × 1 μm scan windows.

for the operation of the final device, special adhesive dispensing techniques or appropriate structures (adhesive reservoir) should be used.

Typical dispensing technologies, e.g., time-pressure, or even truly volumetric systems, screw-based or with piston displacement, are in principle able to generate defined volumes in the range of 2–100 nl. However, the minimum volume ultimately achievable depends strongly of the viscosity and the surface tension of the adhesive as well as the surface properties of the dispenser output capillary. Up to now, sub-nl volume dispensing can only be achieved with inkjet principle based systems, needle transfer or "dipping" techniques.

19.3.3.1. Time Pressure Dispensing Time pressure dispensing is the most widely used dispensing method. This technique uses air pressure applied to the top of a syringe to force material through a needle. The amount of time the air pressure is applied is directly related to the amount of adhesive dispensed.

Several interrelated parameters affect the consistency and volume of the dispensed material: time, pressure, material rheology, and the level of material in the syringe. Indeed, the air pressure applied over the material compresses the entire contents of the syringe before material begins moving through the needle. When the pressure is stopped, however, the compressed material must expand back to its original state. This expansion results in an inability to accurately control the volume of material dispensed. A vision monitoring system can stabilize the dispensing variability by monitoring dot diameters, but it does not address the main origin of the inconsistencies. The dot size limit achievable with time-pressure dispensers is in the range of 200 to 300 μm (2–7 nl) but, in practice, the actual dot dimensions are very sensitive to the rheology of the adhesive, in particular its viscosity and its surface tension. Moreover, such small volumes of adhesive cannot be released with gravity forces and a contact between the hanging droplet and the surface of the submount is required to release the adhesive.

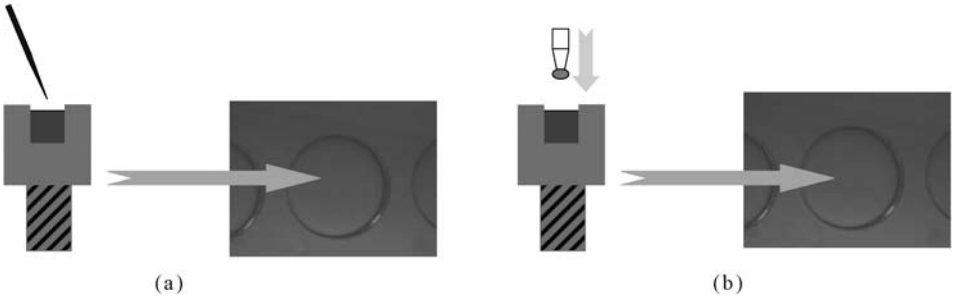


FIGURE 19.25. Basic representations of two sub-nanoliter dispensing techniques: (a) pin-transfer dispensing, (b) dipping.

19.3.3.2. Positive Displacement Piston Positive displacement piston pumps use a piston to force material through a needle. With these pumps, constant low air pressure is applied to a syringe of material. The syringe feeds material to the piston chamber while the piston is in the up stroke position. When the chamber is full, the piston is driven into the needle body, forcing material out the needle. The timing of the piston's up stroke is programmed into the software controlling the equipment and allows just enough time to let the chamber fill. When the piston is in the down stroke, the material feed path is closed so no additional material can escape from the needle. The volume of the material dispensed is determined by two parameters, diameter and piston stroke. Thus, the viscosity of the material plays a minimal role in the volume of material dispensed with piston pumps. Dispensed volumes can be as small as 1 nl with such dispensing systems if a proper piston up-stroke timing is achieved.

19.3.3.3. Ink-Jet Dispensing The ink-jet-principle uses piezo elements to generate pressure waves. The pressure wave travels to the nozzle where a high acceleration ejects a droplet. The volume of the drop is dependent on the adhesive surface tension and on the nozzle size. Adhesive dot diameter down to 70 μm (∼0.1 nl) can be delivered with ink-jet dispensers.

This dispenser can be used only in combination with low viscosity adhesives (up to 300–400 cps). Adhesive dot diameter down to 70 μm (∼0.1 nl) can be delivered with ink-jet dispensers.

19.3.3.4. Pin-Transfer Dispensing The pin-transfer technique (or needle transfer) operates according to the following principle: a needle is first dipped into a reservoir filled with adhesive and then moved to the target substrate [Figure 19.25(a)]. Contact between the needle and the substrate is required to release the adhesive dot. This technique has been recently used to bond graded-index microlenses on a structured silicon substrate in an automated fashion [14]. In that work, the total volume of the target groove is not more than 7 nl. Thus, in order to fix the lenses without having adhesive in the optical active area, small amounts of adhesive well below 1 nl had to be deposited on the substrate.

19.3.3.5. Dipping Technique The intake and release of sub-nanoliter volumes with a pin transfer technique is not trivial because a very small adhesive droplet will have the tendency to climb up the needle with capillary forces. When the consistent release of droplet cannot be achieved with this technique, the dipping method can be used. This technique operates according to the following principle: the bottom surface of the component to be bonded is

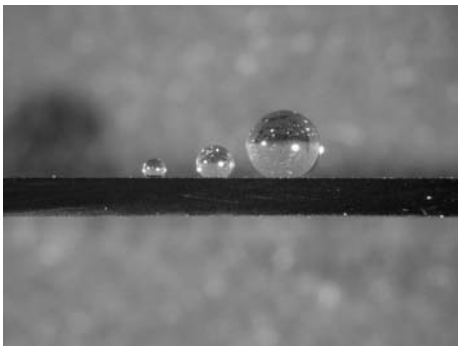


FIGURE 19.26. Assembly of microball lenses on a structured silicon substrate. The diameter of the lenses is 300 μm, 500 μm and 1 mm.

first dipped in a controlled fashion into a reservoir filled with adhesive and then moved to the target substrate [Figure 19.25(b)].

We have used this method together with the pin-transfer dispensing technique to bond microball lenses on a structured silicon substrate. A successful assembly of these lenses is shown in Figure 19.26. The diameter of the lenses was 300 μm, 500 μm and 1 mm respectively. In order to place the microball lenses, round-shaped 70 μm deep grooves had been etched in silicon. The diameter of the grooves varied from 200 μm to 500 μm.

19.4. SOME APPLICATIONS

19.4.1. Laser to Fiber Assembly

This section describes the fiber alignment and fixing procedure for a laser-to-fiber coupling application. The laser chip is a 980 nm laser pump featuring an InGaAs/AlGaAs single quantum well gain structure and a ridge waveguide for optical confinement. The laser die exhibits a vertical far field divergence of 20°.

These lasers are used to pump Er-doped fiber amplifiers. The optical fiber used for the assembly is a single-mode Corning HI1060 for which the mode field diameter at 980 nm is 5.9 μm. The laser is soldered on a cost-effective Al_2O_3 carrier along with a monitor photodiode and a thermistor for temperature measurement. A gold-plated metallic pad in front of the chip provides an ideal bonding surface as surface energy of metals is potentially very large. However, it was not possible to measure the effective surface energy owing to the small dimensions of the pad (400 μm diameter).

The procedure for the fiber alignment step was the following:

- coarse fiber alignment was performed using pattern recognition of the fiber tip and the laser endface,
- a triangulation algorithm was used to achieve an optimal coupling efficiency. The distance between the laser endface and the fiber tip is 2 to 4 microns and the positioning tolerances along the lateral axes are below 1 micron.

The optimal position of the fiber was then stored and the fiber was pulled back to enable adhesive dispensing without damaging the laser chip. The optimal fiber position was then retrieved in a step fashion to avoid a collision between the optical fiber and the laser chip that may result from slight thermal changes or lack of bidirectional reproducibility of the fiber positioning stage.

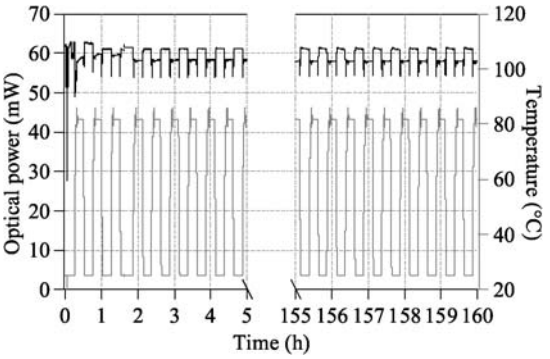


FIGURE 19.27. Thermal cycling tests performed during one week. The assembly went through more than 600 thermal cycles without degradation of the coupling efficiency over time. The upper and bottom lines represent the coupled optical power variation and the temperature cycles respectively. The variation of the coupling efficiency between 25°C and 80°C is 1.5%, resulting from the variation of the Fabry-Pérot air cavity length between the laser facet and the fiber tip as the temperature is increased.

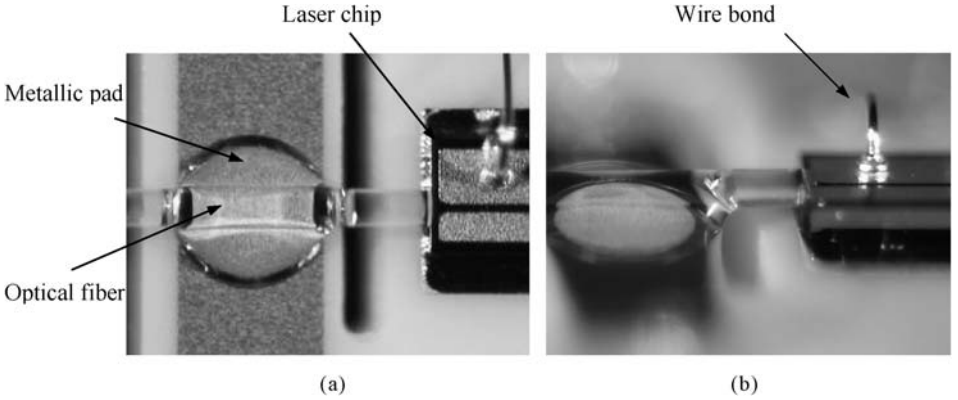


FIGURE 19.28. (a) Top view of the laser-to-fiber assembly. (b) Side-view of the laser-to-fiber assembly. The fiber is embedded in the adhesive dot owing to a planar assembly design.

Once the fiber was properly aligned, a position offset was introduced to compensate for the adhesive shrinkage upon curing (see Section 19.3.2.2.2). Once cured, the assembly went through a stability bake for a better thermal resistance. Figure 19.27 shows the thermal cycling results between 25°C and 80°C over one week (160 hours, more than 600 cycles). No degradation over time is observed during the thermal cycling tests. The variation of the coupling efficiency is 1.5% between 25°C and 80°C and mainly results from the variation of the distance between the fiber tip[13] and the laser endface as the temperature is increased, i.e., a variation of the Fabry-Perot cavity as described in Section 19.3.2.2.4. Indeed, due to the different CTE involved in this assembly, we estimate a cavity length increase of 200 nm. This represents a variation in coupling efficiency of 1% as, over this distance variation, the coupled power goes from a maximum to a minimum of the Fabry-Pérot interferences.

On the top view of the fiber-to laser assembly [Figure 19.28(a)], we see that the flow of the adhesive is well under control as the adhesive joint is limited to the metallic pad.

[13]There is no anti-reflection coating on the fiber tip.

On the side-view of Figure 19.28(b), the fiber is embedded in the adhesive dot owing to a planar chip on carrier assembly configuration.

19.4.2. Planar Lightwave Circuit (PLC) Pigtailing

In a second example, a PLC pigtailing process is described. The PLC's used are 5×5 μm channel waveguides fabricated on a silicon wafer from a fluorinated acrylate polymer with refractive indices of core and cladding of 1.47 and 1.46, respectively.

We first investigated the coupling efficiencies and alignment tolerances that can be achieved with different coupling methods (with/without lenses, with/without adhesive in the gap between waveguide and fiber). The simulation of coupling efficiencies showed that a coupling scheme using lenses gives the largest coupling efficiencies (95%) but tight alignment tolerances (e.g., a lateral fiber shift of 0.85 μm already leads to a decrease in coupling efficiency of 10% this in comparison to 1.4 μm for butt-coupling). Therefore, in order to ease the alignment, an end-fire coupling configuration with slightly lower coupling efficiencies was chosen.

The PLC pigtailing consisted of two main steps:

- The integrated optical polymer chip was first fixed on a submount with a thermal adhesive.
- Subsequently, the fibers were aligned for maximum throughput and finally fixed using a UV-curable adhesive.

In addition to the requirements given in Section 19.2.2.2, the submount should have a high operating temperature as the polymer chip is bonded with a thermal adhesive. Thus, the polymer PEEK (polyether ether ketone) was chosen. The chip bonding step proceeds as follows: three adhesive droplets were dispensed on the submount, the chip was placed onto the submount, and the assembly was put in an oven at a temperature of 120°C for 2 hours.

As the waveguide chip is fully passive, some 'external' light has to be first coupled into the waveguide in order to perform the fiber alignment. The fibers pigtailing setup is described in Figure 19.29. The chip holder was placed between two 3-axis fiber positioning stages having 0.1 μm resolution. The time-pressure adhesive dispenser was mounted on a 3-axis positioning stage for pinpoint adhesive dispensing. A 1550 nm external laser source and detector were used for the active alignment process of the fibers.

Some external light was coupled in the waveguide using the *first* fiber to provide an optical signal for the active alignment of the *second* fiber. The coarse fiber to waveguide in-plane alignment was done using a microscope. The *first* fiber alignment was further improved using an infrared card at the output of the waveguide. Once first light was observed at the output of the waveguide, the *second* fiber could be aligned for maximum throughput.

In order to ease the coarse alignment step of the *second* fiber, the alignment was first performed 100–300 μm away from the waveguide. This procedure was carried out as the laser beam was several tens of microns wide a few hundreds microns away from the waveguide facet due to the waveguide divergence. Once the fiber was properly positioned close to the waveguide optical axis, the fiber was brought 3 to 5 microns away from the waveguide facet and the fiber position was newly optimized. The *first* fiber was now realigned using the reference signal coupled in the waveguide from the *second* fiber in order to measure the maximum achievable optical throughput.

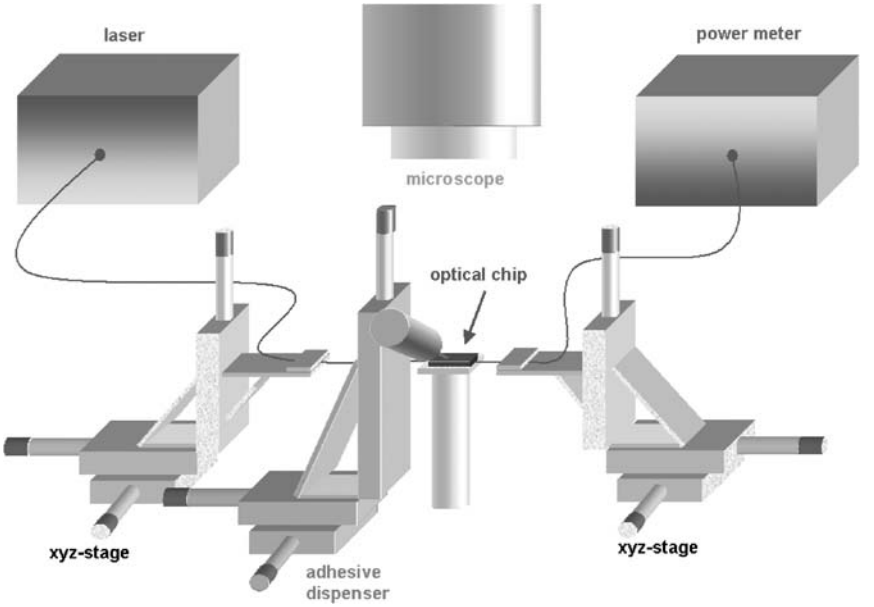


FIGURE 19.29. Setup used for the fiber pigtailing of PLC waveguides.

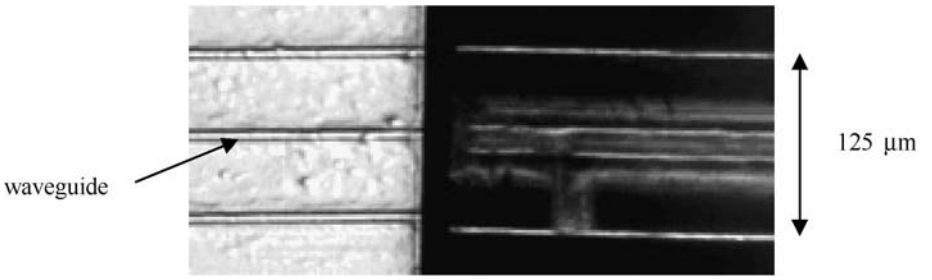


FIGURE 19.30. Image of waveguide (left) and single-mode fiber (right) prior to alignment. The core dimensions are 5×5 µm for the waveguide and 9 µm (diameter) for the fiber.

Figure 19.30 shows an image of the waveguide and the fiber after alignment. Figure 19.31 shows the nice comparison of the coupling efficiencies between simulations and actual experiments carried out before fixing.

Subsequently, the fibers were fixed with a UV-curable adhesive. The *second* fiber was moved away in longitudinal direction and the adhesive was dispensed on the submount. After this dispensing step, the fiber was moved back to its initial position and actively aligned. Finally the adhesive was UV cured. Using this approach, the adhesive also fills the gap between waveguide and fiber which leads to a decrease in the Fresnel losses according to the results of Figure 19.20 and an increase in the alignment tolerances.

The same procedure was applied to the *first* fiber. The current insertion losses (fiber-waveguide-fiber) after fixing were around 4.5 dB compared to a minimum expected loss of around 2.5 dB for the current configuration.

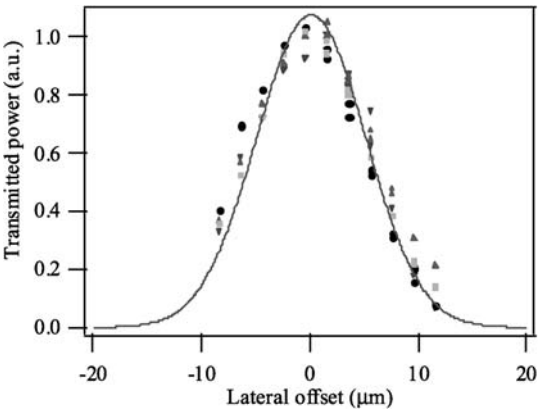


FIGURE 19.31. Influence of a lateral (horizontal) shift on the fiber-coupled power: comparison between simulations (solid line) and experiments (points) displaying a nice correspondence. The lateral alignment tolerance for 10% additional losses is 2.3 μm.

19.5. SUMMARY AND RECOMMENDATIONS

A large variety of active and passive optical devices are currently proposed to supply the telecom and datacom market. The devices are typically fitted into ceramic or metal alloy housings that must provide not only electrical connections and mechanical support but also thermal management and, more critically, optical connections. Optical interconnects are highly directional and, in order to achieve optimal signal transmission, they require extremely precise control of the relative location between components (typ. below 1 μm) and an attachment process that maintains the alignment over time.

Adhesive bonding of optical passive components is often seen as a cheap alternative to laser welding or gold-tin soldering. However, today, there is a mental barrier in using adhesives in optoelectronic packaging. In order to clarify the potential of such bonding approach, extensive reliability tests (thermo-mechanical stability, photostability) must be performed. In order to maximize the adhesion energy and to improve the reliability of adhesive-bonded assemblies, the parts must be free of contaminants and their surface energies should be above 100 mJ m^{-2}. Oxygen plasma cleaning is very effective in removing organic residues and should be done whenever possible. Optical fibers adhesively bonded on silicon substrates have shown to resist pulling load in excess of 2 kg when the silicon surface was oxygen plasma treated.

We have seen that the use of a compliant adhesive material strongly reduces CTE mismatch induced thermal stresses compared with typical solder alloy used in microelectronics and opto-electronics packaging. Provided that internal or external stresses remain at a low level during the entire assembly process, we have estimated that the creep-induced displacement can be limited to values below 1 μm using high T_g adhesives ($T_g > 120°C$). The T_g of a UV adhesive is mainly limited to the actual internal temperature during curing. It is very likely that this temperature does not exceed the required value given in the Telcordia GR1221. Thus, in order to increase the T_g and limit the creep compliance, UV cured adhesives must be thermally post-cured.

The inherent shrinkage of the adhesives during curing can be pre-compensated by applying a component offset before the adhesive polymerization. This offset is then recov-

ered during the curing of the adhesive. The shrinkage is not a problem by itself as long as it is reproducible.

Finally, the refractive index-matching properties of adhesive materials are a unique advantage of adhesive bonding over other fixing technologies including soldering or laser welding. Providing the adhesive is photostable under given stress conditions, the use of an adhesive in the gap between an optical chip and a fiber can improve the optical coupling efficiency and increase positioning tolerances.

ACKNOWLEDGMENTS

The authors would like to thank Jens Kunde, Helmut Knapp and Rainer Bättig for stimulating discussions and valuable suggestions. The authors would also like to acknowledge Yanki Keles for the AFM pictures, Nicolaï Matuschek for the thermal resistivity calculations as well as Sylvain Grossman and Gustavo Aeppli for their technical assistance. Fundings for this work was provided by the Commission pour la Technologie et l'Innovation and by the Cantons of Central Switzerland. The company Bookham (Switzerland) AG provided the laser chips and the modules used in some of the experiments presented in this article. The authors are grateful for their material support.

REFERENCES

1. M. Achouche, V. Magnin, J. Harari, J.-L. Gentner, F. Lelarge, E. Derouin, C. Jany, D. Carpentier, F. Barthe, F. Blache, S. Demiguel, and D. Decoster, New high performance evanescent coupled waveguide UTC photodiodes for >40 Gb/s applications, Proceedings ECOC, Th 3.4.1, Rimini, Italy, 2003.
2. Y. Akasaka, Gain bandwith of optical amplifiers over 100 nm and beyond, Proceedings ECOC, Tu 3.7.1, Rimini, Italy, 2003.
3. T. Akashi, S. Higashiyama, H. Takemori, and T. Koizmi, A silicon optical bench incorporating a tantalum-nitride thin-film resistor, J. Micromech. and Microeng., 14, pp. 283–289 (2004).
4. M.C. Amann, Long-wavelength InP-based VCSELs, Proceedings ECOC, Tu 1.5.1, Rimini, Italy, 2003.
5. E.H. Andrews and A.J. Kinloch, Mechanics of elastomeric adhesion, J. Polymer Sci., 46, pp. 1–14 (1974).
6. W.D. Bascom, R.L. Cottington, and C.R. Singleterry, Dynamic surface phenomena in the spontaneous spreading of oils on solids, in R.F. Gould, Ed., Advances in Chemistry Series, Vol. 43, American Chemical Society, Washington, 1964, pp. 355–379.
7. R. Bättig, H.-U. Pfeiffer, N. Matuschek, and B. Valk, Reliability issues in pump laser packaging, Proc. Lasers and Electro-Optics Society Conference, WW1, Glasgow, 2002.
8. R. Brenot, S. Kerboeuf, N. Bouché, F. Mallecot, V. Colson, O. Gauthier-Lafaye, M. Picq, J.G. Provost, and B. Thédrez, New low-chirp and high power semiconductor amplifier for 10 Gbit/s metropolitan transmission, Proceedings ECOC, Th 3.5.3, Rimini, Italy, 2003.
9. B. Chandran, W.F. Schmidt, M.H. Gordon, and R. Djkaria, The determination and utilization of AuSn solder properties to bond GaAs dice to diamond substrates, ASME Proceedings of the Application of CAE/CAD to electronics systems congress, 1996, pp. 61–66.
10. B.W. Cherry and C.M. Holmes, Kinetics of wetting of surfaces by polymers, J. Colloïd Interf. Sci., 38, p. 174 (1969).
11. B.W. Cherry and S.E. Mudaris, Wetting kinetics and the strength of adhesive joints, J. Adhesion, 2, pp. 42–49 (1970).
12. W.T. Chen and C.W. Nelson, Thermal stresses in bonded joints, IBM J. Research and Development, 23, pp. 179–188 (1979).
13. H.M. Clearfield, D.K. McNamara, and G.D. Davis, Adherend surface preparation for adhesive bonding, in L.-H. Lee, Ed., Adhesive Bonding, Plenum Press, 1991.
14. A. Codourey, A.-C. Pliska, C. Bosshard, B. Sprenger, U. Gubler, A. Steinecker, M. Thurner, and M. Honegger, A robotic system for assembly and packaging of micro-optoelectronic Components, DTIP Conference, Montreux, 2004.

15. L.A. Coldren, Integrated tunable transmitters for WDM networks, Proceedings ECOC, Th 1.2.1, Rimini, Italy, 2003.
16. K.A. Cooper, R. Yang, J.S. Mottet, and G. Lecarpentier, Flip chip equipment for high end electro-optical modules, Proc. 48th ECTC, Seattle, 1998, pp. 176–181.
17. N. El Dahda, A. Shen, F. Devaux, G. Aubin, J.C. Harmand, A. Garreau, B.-E. Benkelfat, and A. Ramdane, Novel InGaAs/InGaAlAs MQW electroabsorption modulator for ultra-fast optical signal processing, Proceedings ECOC, We 2.5.3, Rimini, Italy, 2003.
18. L.R. Dalton, Novel polymer-based, high-speed electro-optic devices, Proceedings ECOC, Tu 3.5.1, Rimini, Italy, 2003.
19. F. Dorgeuille, S. Rabaron, F. Pommereau, C. Artigue, and P. Brosson, Optical amplifier device, United States Patent Application 20020154392 (2002).
20. J. Elwenspoek and H. Jansen, Silicon Micromachining, Cambridge University Press, 1998.
21. F.M. Fowkes, Role of acid-base interfacial bonding in adhesion, J. of Adhesion Science and Technology, 1, pp. 7–27 (1987).
22. K. Fukatsu, T. Takeuchi, K. Shiba, K. Makita, Y. Amamiya, Y. Susuki, and T. Kato, An extremely compact (0.3 cc) 40 Gb/s optical receiver module with ease of use receptacle interface and feedthrough launcher, Proceedings ECOC, Th 3.4.3, Rimini, Italy, 2003.
23. L. Guiziou, P. Ferm, J.M. Jouanno, and L. Shacklette, Low-loss extinction ratio 4×4 polymer thermo-optical switch, Proceedings ECOC, Paper TuL1.4, Amsterdam, 2001.
24. N. Gougeon, M. Poulain, and R.L. Abdi, Strength of silica fibers under various moisture conditions, Photonics Fabrication Europe Conference, Bruges, Proceedings SPIE, 4940 (2002).
25. C. Helming and J. Teunissen, Optical low-cost temperature point sensor, Proceedings SPIE, 4074 (2000).
26. B. Hvolbaek Larsen, L. Pleth Nielsen, K. Zenth, L. Leick, C. Laurent-Lund, L.-U. Aaen Andersen, and K.E. Mattsson, A low-loss silicon oxynitride process for compact optical devices, Proceedings ECOC, We 1.2.6, Rimini, Italy, 2003.
27. V.O. Indisov, V.N. Kuryatov, B.N. Semenov, I.M. Sokolov, and Ya.A. Fofanov, Polarization characteristics of total internal reflection lasers prisms, Part 1, Optics and Spectroscopy, 75, pp. 121–127 (1993). Polarization characteristics of total internal reflection lasers prisms, Part 2, Optics and Spectroscopy, 75, pp. 266–271 (1993).
28. J.N. Israelachvili, Intermolecular and Surface Forces, Academic Press, 1991.
29. R.J Jenkins, M.E. McNie, A.F. Blockley, N. Price, and J. McQuillan, Hollow waveguides for integrated optics, Proceedings ECOC, Tu 1.2.4, Rimini, Italy, 2003.
30. N. Keil, H. Yao, C. Zawadzki, O. Radmer, F. Beyer, M. Bauer, C. Dreyer, and J. Schneider, Polarization and temperature behavior of all-polymer arrayed-waveguide gratings, Proceedings ECOC, Tu 3.5.3, Rimini, Italy, 2003.
31. A.J. Kinloch, Adhesion and Adhesives, Science and Technology, Chapman and Hall, 1987.
32. J. Kunde, M. Thurner, A.-C. Pliska, Ch. Bosshard, A. Codourey, R. Bauknecht, and S. Egger, Comparison of microlens technologies for passive alignment platform applications, Proceedings MOC, V1.6, 2004.
33. J.-H. Kuang, M.-T. Sheen, C.-F. Chang, C.-C. Chen, G.-L. Wang, and W.-Hi. Cheng, Effect of temperature cycling on joint strength of PbSn and AuSn solders in lasers packages, IEEE Trans. Adv. Packaging, 24, pp. 563–568 (2001).
34. J. Lau, C.P. Wong, J.L. Prince, and W. Nakayama, Electronic Packaging: Design, Materials, Process and Reliability, McGraw-Hill, 1998.
35. M. Lee, M. Wong, and Y. Zohar, Characterization of an integrated micro heat pipe, J. Micromech. and Microeng., 13, pp. 58–64 (2003).
36. A. Liu, R. Jones, L. Liao, D. Samara-Rubio, D. Rubin, O. Cohen, R. Nicolaescu, and M. Panaccia, A high-speed silicon optical modulator based on a metal-oxide-semiconductor capacitor, Nature, 427, pp. 615–618 (2004).
37. T.J. Lu, D.F. Moore, and M.H. Chia, Mechanics of micromechanical clips for optical fibers, J. Micromech. and Microeng., 12, pp. 168–176 (2002).
38. M.A. Lyons and D. Dahringer, Electrically conductive adhesives, in A. Pizza and K.L. Mittal, Eds., Handbook of Adhesive Technology, 2nd edn, 2003.
39. M. Mattewson, Environmental effect on fatigue and lifetime prediction for silica optical fibers, Photonics Fabrication Europe Conference, Bruges, Proceedings SPIE, 4940 (2002).
40. M.A. Meyers, R.W. Armstrong, and H.O.K. Kirchner, Mechanics and Materials: Fundamentals and Linkages, John Wiley & Sons, New York, 1999.
41. L.F. Miller, Controlled collapse reflow chip joining, IBM J. Research and Development, 13, pp. 239–250 (1969).

42. R.A. Modavis and T.W. Webb, Anamorphic microlens for laser diode to single mode fiber coupling, IEEE Photon. Tech. Lett., 7, pp. 798–800 (1995).
43. S. Mohrdiek, T. Pliska, R. Bättig, N. Matuschek, B. Valk, J. Troger, P. Mauron, B.E. Schmidt, I.D. Jung, C.S. Harder, and S. Enochs, 400 mW uncooled MiniDIL pump modules, Elec. Let., 39, pp. 1105–1107 (2003).
44. H.S. Morgan, Thermal stresses in layered electrical assemblies bonded with solder, J. Elec. Packaging, 113, pp. 350–354 (1991).
45. A. Norris and J. DeGroot, Silicone materials for optical device applications, Proceedings ECOC, Tu 3.5.6, Rimini, Italy, 2003.
46. I.M. Plitz, O.S. Gebizlioglu, and M.P. Dugan, Reliability characterization of UV-curable adhesives used in optical devices, Proceedings SPIE, 2290, pp. 150–159 (1994).
47. A.C. Pliska, J. Kunde, S. Grossmann, C. Bosshard, T. Pliska, and B. Valk, Automated fiber alignment using machine vision and pattern recognition, Technology Leadership Day, Winterthur, 2003.
48. A.-C. Pliska, J. Kunde, S. Grossmann, Ch. Bosshard, R. Bättig, S. Pawlik, T. Pliska, S. Saintenoy, and B. Schmidt, Low-cost optoelectronic packages: development of a fast alignment technique and a stable bonding process of singlemode optical fibers, EMPC Conference, Brugges, 2005.
49. E.P. Plueddemann, Adhesion through silane coupling agents, in L.-H. Lee, Ed., Fundamentals of Adhesion, Plenum Press, 1991, pp. 279–290.
50. M. Pocha, O.T. Strand, and J.A. Kerns, A silicon microbench concept for optoelectronic packaging, Proc. Surface Mount Int., 1, pp. 377–382 (1996).
51. C.F. Popelar and K.M. Liechti, A distorsion-modified free volume theory for nonlinear viscoelastic behavior, Mechanics of Time Dependant Materials, 7, pp. 89–141 (2003).
52. L.A. Reith, O.S. Gebizlioglu, M. Koza, J. Mann, M. Ozgur, and T. Bowner, The dimensional stability of adhesives, zirconia and silica in optical connector ferrules and their impact on optical performance, Proceedings of Mat. Res. Soc. Symposium, 1998, pp. 65–76.
53. M. Saruwatari and K. Nawata, Semiconductor laser to single-mode fiber coupler, Appl. Opt., 18, pp. 1847–1856 (1979).
54. L.H Sharpe and H. Schonhorn, Advances in Chemistry Series, Vol. 8, American Chemical Society, Washington, 1964, p. 189.
55. C. Strandman and Y. Bäcklund, Passive and fixed alignment of devices using flexible silicon elements formed by selective etching, J. Micromech. and Microeng., 8, pp. 39–44 (1998).
56. A. Stump, U. Gubler, and C. Bosshard, Polymer optical waveguides structured by UV-exposure, EOS Topical Meeting, Engelberg, 2004.
57. E. Suhir, Calculated thermally induced stresses in adhesively bonded and soldered assemblies, Proc. Int. Microelectronics Symposium, Atlanta, 1986.
58. S. Tabata, T. Saito, K. Kawamura, T. Itoh, and T. Hatta, 32×32 bascule optical switch with polymer waveguide, Proceedings ECOC, Tu 3.5.2, Rimini, Italy, 2003.
59. T.G. Tessier, G. Ademon, and I. Turlik, Polymer dielectric options for thin film packaging applications, IEEE Proc. of 39th Elec. Comp. and Tech. Conference, 1989, pp. 127–132.
60. K. Tsuzuki, T. Ishibashi, T. Ito, S. Oku, Y. Shibata, R. Iga, Y. Kondo, and Y. Tohmori, A 40 Gbit/s InP-based n-i-n Mach-Zehnder modulator with a π-voltage of 2.2 V, Proceedings ECOC, We 2.5.2, Rimini, Italy, 2003.
61. M. Uekawa, H. Sasaki, D. Shimura, K. Kotani, Y. Maeno, and T. Takamori, Surface-mountable silicon microlens for low-cost laser modules, IEEE Photon. Tech. Lett., 15, pp. 945–947 (2003).
62. B. Valk, P. Müller, and R. Bättig, Fiber attachment for 980 pump lasers, Proc. Lasers and Electro-Optics Society Conference, ThV2, 2000, pp. 880–881.
63. S.S. Voyutski, Autohesion and Adhesion of High Polymers, Interscience, New York (1963).
64. E. Zielinski and H.P. Mayer, Optohybrids in high capacity communication systems, Symp. Opto-Microelectronics Devices and Circuits, Stuttgart, 2002.
65. W.H. Wong, E.Y.B. Pun, and K.S. Chan, Rare-earth doped polymer waveguide amplifiers, Proceedings ECOC, Th 4.2.7, Rimini, Italy, 2003.

20

Electrically Conductive Adhesives: A Research Status Review

James E. Morris[a] and Johan Liu[b,c]

[a]*Department of Electrical and Computer Engineering, Portland State University, Oregon, USA*
[b]*SMIT Center and Department of Microtechnology and Nanoscience, Chalmers University of Technology, Goteborg, Sweden*
[c]*SMIT Center, Shanghai University, China*

20.1. INTRODUCTION

20.1.1. Technology Drivers

There are two primary categories of Electrically Conductive Adhesive (ECA):

- Isotropic Conductive Adhesive (ICA).
- Anisotropic Conductive Adhesive (ACA).
- ACAs are available as paste (ACP) or film (ACF).

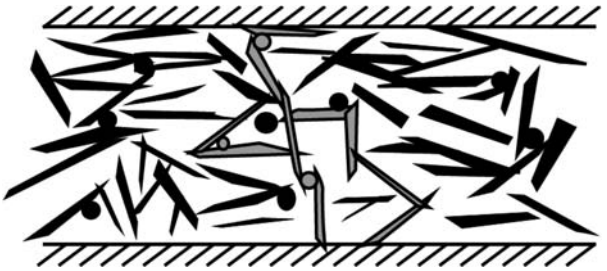
Both types conduct through metal filler particles in an adhesive polymer matrix.

ECAs have been used for electronics packaging applications for decades in hybrid, die-attach and display assembly. ICAs have been used extensively for die attach, and in automotive electronics. ACF technology is employed with almost every liquid crystal display. But there has been growing interest from the electronics industry over the past decade or so in other kinds of electronics packaging applications. While toxicity issues and environmental incompatibility of the lead in tin-lead solders triggered that greater interest at the outset, it has been the other evident advantages which continue to drive further research. ECAs can offer the following additional potential advantages:

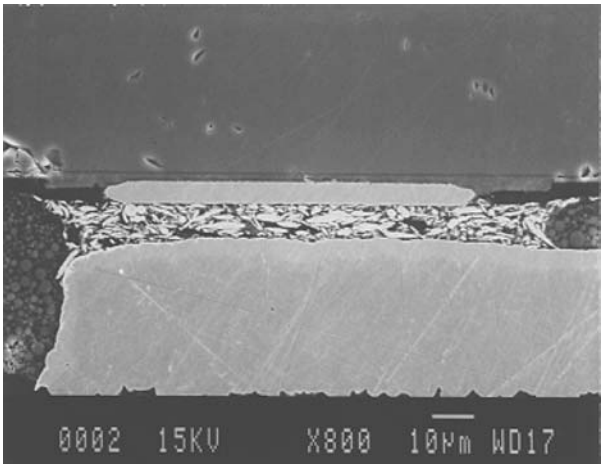
- Fine-pitch capability, especially when using ACAs for flip-chip.
- Elimination of underfilling with ACA bonding.
- Low temperature processing capability.
- Flexible, simple processing and hence low cost.

In addition, solder failures due to the formation of voids or brittle intermetallics are controlled by diffusion rates at small dimensions, and time-to-failure scales as the square of the dimensions, limiting solder reliability lifetimes at ultra-fine pitch [1].

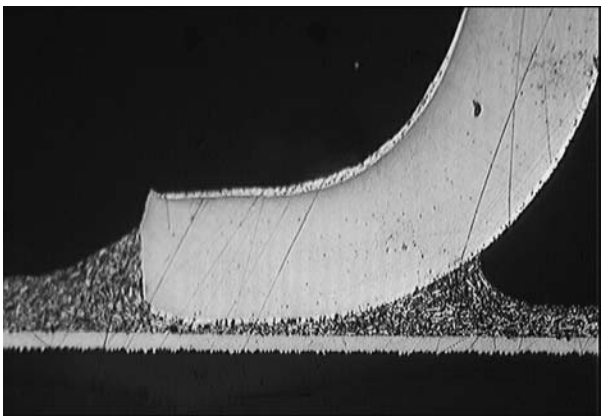
ACF has long been the interconnect of choice in the LCD display industry, and ACP is now finding applications in flex circuits and surface mount technology (SMT) for chip-



(a)



(b)



(c)

FIGURE 20.1. ICA contact joints: (a) schematic, (b) flip-chip on FR-4, (c) SMT on FR-4.

scale package (CSP), ASIC, and flip chip attachment for cell phones, radios, PDAs, laptop PCs, and cameras. ICAs are used extensively in die-attach, for many years for small passive chip attachment in automotive electronics, and in varied consumer products, e.g., in Matsushita/Panasonic's stud bump joints. More recent applications include RFID tags, potentially for the antenna as well as chip connections.

This chapter presents an overview of the current status of understanding of conductive adhesives in various electronics packaging applications, and of some fundamental issues relevant to their continuing development. It is organized with initial discussions of basic ECA concepts of structure-related properties, and how these are affected by material selection and processing, followed by general properties and reliability considerations. In each section, there is material common to both ICAs and ACAs, and parallel treatments of topics specific to each.

20.1.2. Isotropic Conductive Adhesives (ICAs)

Ag is usually used as the filler material due to its high conductivity and simple processing for ICA applications. Figure 20.1 shows the microstructure of an ICA joint, (a) in schematic form, (b) for a flip chip component, and (c) for SMT attachment (both on FR-4 substrates). The metallic filler content is high enough (25–30 volume percent) to cause direct metallic contact from bump or lead to circuit board pad.

20.1.3. Anisotropic Conductive Adhesives (ACAs)

Polymer based metal-plated spheres or Ni fillers are mainly used for ACA applications. In an ACA joint, the filler particles normally constitute between 5–10 volume percent, and do not cause any direct metallic contact. It is only after the application of pressure during curing that electrical conduction becomes possible in the normal (z-axis) direction, as is illustrated in Figure 20.2. The Ni particles are usually matched to soft Au contacts, and the contact deforms to define the contact area. By contrast, the polymer beads, usually Au-plated, deform to establish bump and pad contacts. As there is no direct contact between the particles, ACA technology is very suitable for fine pitch assembly, and is starting to find applications in flip-chip technology.

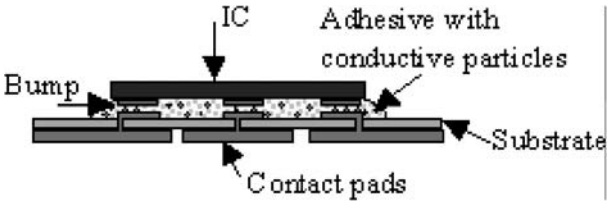
20.1.4. Non-Conductive Adhesive (NCA)

The complementary NCA technology is mentioned only briefly here, for completeness. In this case, the contacts themselves are in direct metal-to-metal contact, held together by the NCA. The concept relies upon surface asperities for the direct contacts.

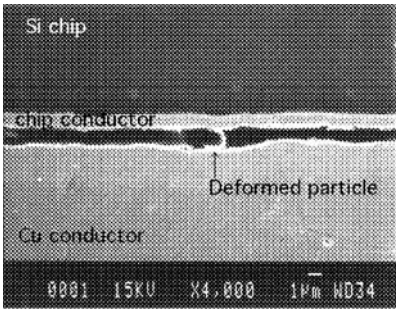
20.2. STRUCTURE

20.2.1. ICA

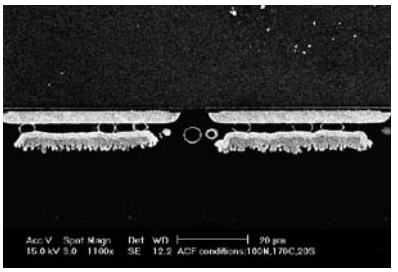
As the proportion of metal in the ICA polymer matrix is increased (Figure 20.3), the resistance drops only slightly until the "percolation threshold" is reached, when the first continuous metal path is established through the composite material [2]. The resistance continues to drop more slowly as multiple parallel paths are developed with the continued addition of more metal filler. Ideally one would like to use the minimum quantity of filler necessary to pass the threshold, but in practice manufacturing tolerances require a design target composition significantly beyond the threshold. Very small contact volumes increase the statistical spread of resistivities already inherent in a percolation structure. Minimal filler content is a requirement for both economic reasons (since Ag is expensive) and to maximize the proportion of polymer adhesive. Both issues can be addressed by the use of



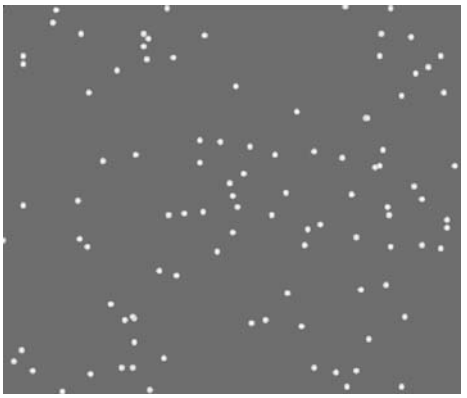
(a)



(b)



(c)



(d)

FIGURE 20.2. ACA flip-chip joining: (a) schematic, (b) Ni hard particle filler, (c) metal coated polymer filler, (d) vertical view showing random filler dispersion.

metal flakes (or rods) as filler instead of spheres. The benefits can be readily understood by considering the extreme case of flakes of zero thickness, which would clearly establish a percolation threshold at zero metal content, i.e., at 100% adhesive with zero filler cost. Practical commercial materials achieve substantial reduction in the threshold composition by the use of flakes, by virtue of the increased connectivity which accompanies the increased surface-to-volume ratio, compounded by bimodal particle size distributions (Figure 20.4). The efficiency of bimodal particle distributions has been demonstrated [3] and either flakes or powders can be used for the smaller particles [4].

Theoretical simulation has been carried out to optimize the electrical performance of a conductive adhesive joint using a bimodal distribution of metal fillers using computer modeling with finite-element analysis. Two classes of metal fillers were used, i.e., nanoscale and micro-scaled particles. The goal was to decrease the metal loading to improve

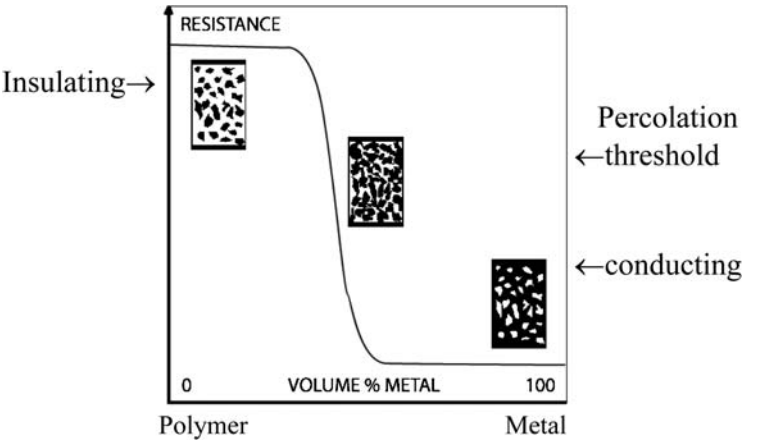


FIGURE 20.3. Percolation threshold [2].

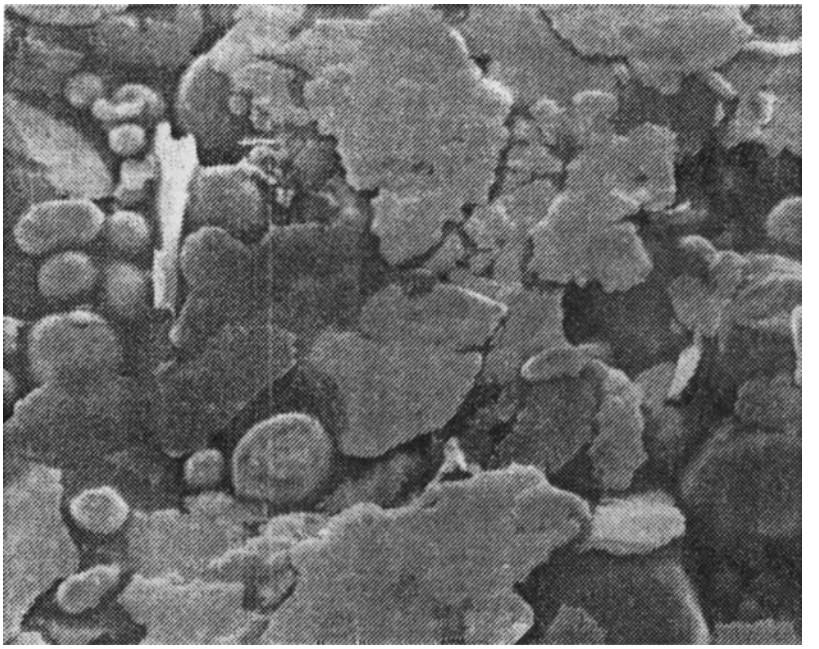


FIGURE 20.4. ICA bimodal filler distribution (Ag flakes and powder) with surface layering [4,6].

the mechanical performance for specified electrical properties. It has been shown that it is possible to decrease the total metal loading with good electrical conductivity using a bimodal filler distribution [5], but that the nano-particles increase resistivity for a given total filler content, due to mean free path limits and increased numbers of contacts.

In addition, the flakes at the material surfaces seem to be aligned parallel to the surface to a depth of a few flake thicknesses (Figure 20.4). The effect seems to be universal, with squeegee or syringe dispensing, and also appears inside air bubbles, probably due to surface tension [6]. One also sees particle alignment inside the material, parallel to the

contacts, if pressure is applied [7], as seen in Figure 20.1(b). Particulate orientations are important for predictions of both adhesive strength and electrical resistance. Flow modeling of the ICA under stencil, print, or dispensing, and/or positioning pressure is much more difficult than underfill flow modeling [8] due to filler sizes and shapes. Recently, a dynamic model of the effects of compression has demonstrated flake alignment quite dramatically [9].

There are various novel approaches to the improvement of electrical connectivity at a given metal content, including magnetic alignment of nickel filler rods [10], the use of polymer particles to force z-axis alignment of flakes [11], and electric fields [12]. Anything that increases the packing density/efficiency of the particles also increases the connectivity and reproducibility of the structure, permitting the design composition to move closer to the percolation threshold, and a reduction in metal filler.

Modeling of the material structure for prediction of the electrical properties has included

- Distributions of uniformly sized spheres [13].
- 2D [14] and 3D [15] rectangular particles with limited size distributions and x, y, z orientations.
- Rectangular particles rotated about the x, y, z axes in 1° increments [16].

The extension to bimodal flake representations requires a dramatic increase in computing power. The difficulty lies in the development of an efficient algorithm for the random placement of the particles, which must not impinge upon the space occupied by another. As the structure fills up, this process becomes more and more time consuming. A potential energy technique has proved effective [16], but the most recent advances have been by the use of compression algorithms applied to initially well separated particles [9,17,18].

20.2.2. ACA

The first structural issue for ACAs is the distribution of particles, and how many are captured and compressed between the contacts. A software package has been developed to calculate the average resistance of the ACA contacts for pastes or randomly loaded ACFs [19], with statistical distribution. Input parameters include particle size and distribution, particle loading, pad size, and pad spacing. The program also performs the inverse calculations, i.e., it determines the particle loading requirements for a specified minimum number of particles per pad (i.e., minimum conductance).

It is implicit in the program above that the particle distribution is well behaved, but for pastes, the realities of flow around and over pads as pressure is applied will dramatically alter the distribution. So the program has only been validated against ACFs, and studies on the entrapment of particles between pads continue. There have been efforts made to model ACP flow [20], and the problem should be more tractable than underfill flow modeling [8], due to the smaller particles, but there appears to be no result yet suitable for inclusion into a particle distribution model.

There is, of course, a finite probability that there are no particles in the joint. Williams et al. tried to estimate this by assuming the particles are placed in the bonding area according to the Poisson distribution [21]:

$$P(n) = \frac{e^{-\mu}\mu^n}{n!}, \tag{20.1}$$

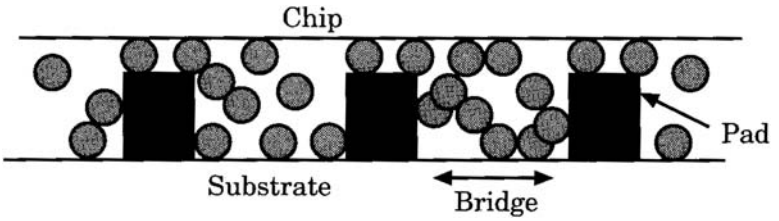


FIGURE 20.5. Bridging during ACA bonding (Courtesy of Mannan et al.) [24].

where P is the probability of an open joint, n is the number of particles per pad, and μ is the average number particles on a pad, given by:

$$\mu = \frac{3Af}{2\pi r^2}, \tag{20.2}$$

where A is the total bonding area, f is the volume fraction of the particles, and r is the radius of the particle. So, for a typical volume fraction ranging from 3 to 15 vol%, chip area (100 mm^2) and pad size (100 μm^2), the probability for an ACA open circuit is

$$P(0) = e^{-\mu} = e^{-\frac{3Af}{2\pi r^2}}, \tag{20.4}$$

which gives 10^{-13} to 10^{-3} for typical parameters, i.e., extremely small. However, in reality, there is always a crowding effect which must be taken into account. In this case, the particle distribution can be described using a binominal distribution model [21]:

$$P(n) = C_n^N (1-s)^{N-n} s^n, \tag{20.5}$$

where N is the maximum number of particles that can be contained in an area A, C_n^N is the binominal coefficient, and s is equal to f/f_m where f_m is the volume fraction corresponding to maximum packing [22]. In the limit when $f \ll 1$, Equations (20.4) and (20.5) give identical results for $P(0)$.

Bridging is possible due to there being too many particles between contacts, or insufficient contact separation, as illustrated in Figure 20.5 [23,24]. The probability for bridging is given as

$$p = 1 - \left[1 - \left(\frac{6f}{\pi}\right)^{\frac{d}{4r^2}}\right]^{\frac{hl}{4r^2}}, \tag{20.6}$$

where f is the volume fraction of the particles, h is the pad height, l the pad length, d the spacing between the pads, and r is the radius of the particle. Bridging probabilities are also very small, as seen for various cases in Figure 20.6, where it is clearly shown that the lowest combined probability for bridging and skipping occurs in the volume fraction between 7 and 15% depending on what model we use. This volume fraction range is also generally used for commercial ACA materials today.

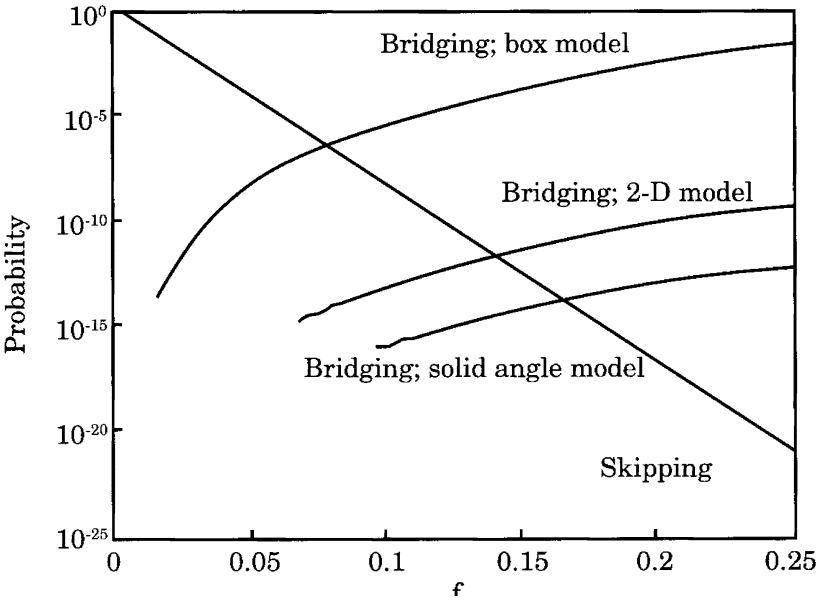


FIGURE 20.6. Probability of particles bridging gap as a function of volume fraction of particles during ACA bonding (Courtesy of Mannan et al.) [24].

20.2.3. Modeling

The goal of any physical modeling exercise is to demonstrate that the physical processes are well understood. Validation is achieved by agreement of the model predictions with experimental data. It is clear for both ICAs and ACAs, that successful comparisons of electrical modeling with experimental results require accurate structural modeling first. For ICAs, there have been some superficial efforts at structural modeling in the past, and as a result, comparisons of electrical models with experiment were either strictly qualitative [14,15] or subject to parameter adjustment to achieve a fit [16]. For ACAs, one would expect less difficulty in structural characterization, and the general form of the match of theory to experiment for the resistance variation with pressure [25] is convincing, even recognizing the need for fitting parameters. For ICAs, similar results are achievable for "dry" test systems [10], i.e., without epoxy, but cannot be correlated with actual ICA internal pressures because these are unknown. Real progress in realistic ICA structural modeling has bee made only recently [9,18].

20.3. MATERIALS AND PROCESSING

20.3.1. Polymers

The common use polymers in ECAs are thermosetting epoxies, with sufficient thermoplastic mixed in to allow for softening and release for rework under moderate heat. Up until now, the polymers used in ICAs have been adapted from those developed for other purposes. Recent research has provided clear pointers to the properties required of the new generation of materials, but the engineering trade-offs make implementation difficult, and the universal solution has yet to appear on the market. Certainly, the need for mechanical

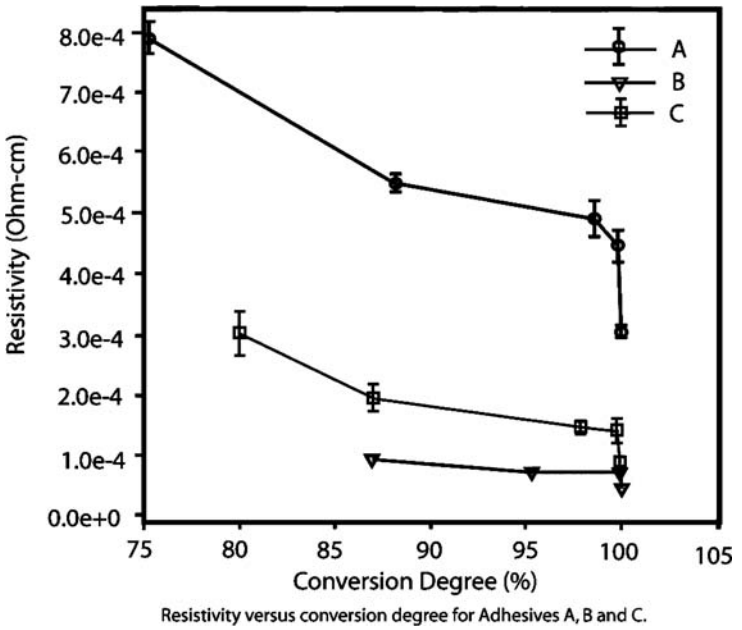


FIGURE 20.7. Resistance dependence on cure as modeled by Equation (20.7) [4,6,27].

energy dissipative materials is well known, and the search for new materials must focus on this property, but without compromising on others. Polymer adhesion is a fundamental property which must be understood, with fundamental contact angle wetting experiments for ICA base polymers a first step.

The cure process has been modeled successfully by very simple mathematical expressions [26,27], which do however require accurately determined parameters from experimental Differential Scanning Calorimetry (DSC) data. For reaction rate

$$d\alpha/dt = kf(\alpha), \tag{20.7}$$

where $\alpha =$ degree of cure, $k = A\exp -(E/kT)$ is the chemical rate constant, and $f(\alpha)$ is a function of the reactant concentration, given by

$$f(\alpha) = (1-\alpha)^n \tag{20.8}$$

for an n-th order model, the degree of cure is calculated as:

$$\text{1st order} \quad d\alpha/dt = k(1-\alpha), \quad \alpha = 1-\exp(-kt), \tag{20.9}$$

$$\text{2nd order} \quad d\alpha/dt = k(1-\alpha)^2, \quad \alpha = 1-(1+kt)^{-1}, \tag{20.10}$$

etc. More complex treatments employ a linear combination with the auto-catalyzed model

$$d\alpha/dt = kf(\alpha) = (k_1 + k_2\alpha^m)(1-\alpha)^n. \tag{20.11}$$

The success is demonstrated by the observation of the sudden decrease in ICA resistance (Figure 20.7) at the predicted point of 100% cure [6,27] according to Equation (20.9).

The assumption is that the resistance drop is due to physical shrinkage of the polymer matrix approaching complete cure, but this particular property, i.e., physical shrinkage and the development of an internal pressure to squeeze filler particles together, has not been measured, and does not appear to have been modeled. The measurement of a dimensional shrinkage with cure should not be a major problem, nor should the measurement of internal pressures.

Prior to the initiation of cure, however, the carrier solvents which permit the ICA paste (or ACP) to be printed must be expelled. Failure to do so degrades reliability, adhesion, and impact strength. The problem arises when partially cured polymer traps the evaporating solvent in bubbles. Brief exposure to vacuum prior to cure visibly decreases ICA paste volume as gas escapes from the surface [28], but the more practical technique is a pre-cure heat soak, e.g., for about 20–30 minutes at 100–120°C [29] which achieves the same result and marked reliability improvement.

ICAs typically cure at around 150°C for 20–30 minutes. This is considered too long to be competitive with solders [30], but there are much faster "snap-cure" alternatives.

20.3.2. ICA Filler

The metal filler of choice has been Ag, but the warning that Ag is accompanied by electromigration problems keeps on coming up. Certainly the existence of the electromigration problem has been documented [31], but it does not seem to be a problem in practice. It is evident in uncured material [12], and it has been suggested that commercial additives to the polymer seal the silver surface, defeating migration tendencies. There also appears to be a field threshold, and moisture is a requisite, but systematic study is required to establish the boundaries to the effect. In addition, the diffusion and clustering of metals in polymers is well established [32,33], and should be investigated for Ag in appropriate polymers as the limiting zero-field phenomenon. Sencaktar et al. have recently correlated electromigration with Ag surface pitting [34].

There have also been materials reported using low melting point alloys [35,36], or Sn-coated Ag particles [37]. The intent is for the particles to form metallurgical bonds during the polymer cure, to achieve lower contact resistances. The greater rigidity of the metallic network could be a problem if the contacts fracture under mechanical stress, but apparently the Sn inhibits Ag migration [37].

20.3.3. ACA Processing

ACA assembly typically requires a thermode for the bonding process, which takes place in about 30 seconds. Many parameters can affect the bonding quality during the ACA bonding process:

- Curing temperature and time of the ACA.
- Bonding temperature and time.
- Temperature ramp rate.
- Alignment accuracy.
- Pressure value, pressure distribution and pressure application rate.
- Bump height and uniformity.
- Board planarity and stiffness.

If the thermal ramp rate is too high, no particle contact will occur as the adhesive will be cured during the bonding process, before the pad and bump can both reach the particles.

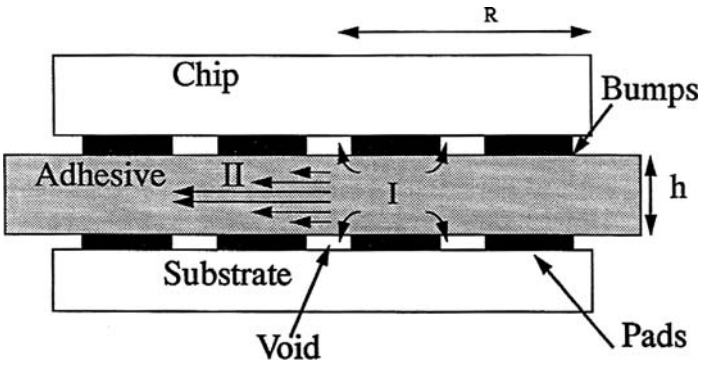


FIGURE 20.8. Schematic drawing of the ACA flow during bonding (Courtesy of Mannan et al.) [24].

The effect of bump to pad alignment accuracy is mainly related to the performance of the bonder. A flip-chip bonder that offers a 5 μm alignment accuracy is usually good enough. The question is how much tolerance is acceptable for the ACA bonding. As has been discussed earlier, the particle location will affect the electrical conduction performance of the joint, and bad alignment will have the same effect.

ACA flow during bonding has also been modeled by Mannan et al. [24]. A typical situation of the particle flow is shown in Figure 20.8. Assuming Newtonian flow

$$\tau_{xy} = \eta \delta\gamma/\delta t, \tag{20.12}$$

where τ_{xy} is the shear stress, $\delta\gamma/\delta t$ is the strain rate, and η is the viscosity. By solving the Navier-Stokes equation in this condition, one can obtain the following equation that describes the pressure under the chip during the bonding due to the flow of the ACA:

$$P = 2F/\left(\pi R^2\right)\left(1 - r^2/R^2\right), \tag{20.13}$$

where F is the bonding force, r is the distance from the chip center, and R is half of the edge length of the chip. Pressure variations cause planarity problems during the bonding. In reality, the ACA resin probably follows a power law type of flow as

$$\tau_{xy} = \eta_0 (d\gamma/dt)^n, \tag{20.14}$$

where n is the power law index and η_0 is the viscosity of the fluid. In this case, Mannan et al. have successfully obtained the following equation to predict the time for the ACA to be squeezed out between the chip and substrate:

$$Tp = (2n+1)/(n+1)\left\{2\pi\eta_0^{(n+3)}/\left[F(n+3)h_0^{(n+1)}\right]\right\}^{1/n}\left[(h_0/h)^{(n+1)/n} - 1\right], \tag{20.15}$$

where h is the ACA height. One of the important issues here is that the viscosity of the ACA will increase dramatically when it starts to get cured. This can be described as

$$\eta_0 = \eta_{00} e^{-b(T-T_0)}, \tag{20.16}$$

where b and η_{00} are constants and T_0 is an arbitrary reference temperature. However, measuring the viscosity of the ACA material close to the curing point remains a problem.

20.4. ELECTRICAL PROPERTIES

20.4.1. ICA

The electrical resistance of the ICA has four distinct components [2]. Three of these are obvious; they are the metal "intra-particle" resistance, the "inter-particle" contact resistance, and the "contact" resistance between the surface particles and the lead or contact pad. The fourth component comes from the meandering "percolation" path of the continuous metallic connection(s) through the material. Figure 20.9 illustrates the principle, showing how some metal particles carry current, while others do not.

(1) *Percolation theory* is well developed for the elementary system of uniform conducting spheres (or cubes) in a perfectly insulating medium [38]. There are no analytical solutions, and the theoretical results are deduced by the averaging of multiple Monte Carlo randomized simulations. Conducting particles are randomly assigned to sites on a specified regular array. For any realistic random system, the site separations must be much less than the particle sizes, and the problem becomes how to fit new particles into the structure at high particle concentrations (as required here) with reasonable efficiency.

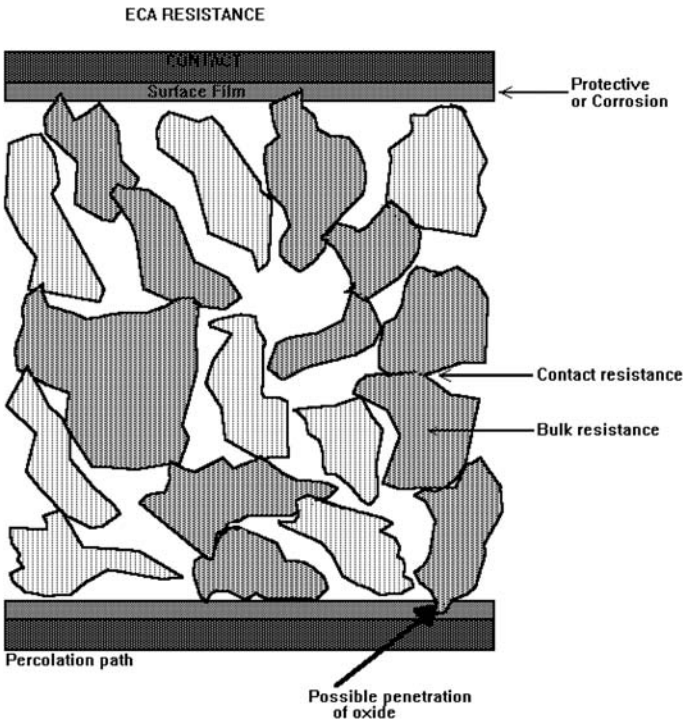


FIGURE 20.9. Schematic of metal particles/flakes in an ICA joint, showing sources of resistance [2].

The percolation modeling literature commonly includes finite intra-particle resistivities, but inter-particle resistances are seen less often. This is partly because the most common systems of interest would assume electron tunneling between particles, and the exponential dependence of the tunneling probability on separation introduces a strong parametric dependence on a poorly characterized variable. Percolation models of the electrical resistance of ICA systems have included both intra-particle and inter-particle resistances, but so far with only gross approximations, e.g., with the simplifying assumption of uniform tunneling thickness [13–16]. (To accommodate 1 nm variations in a 1–10 nm tunneling separation range between particles requires an underlying simulation grid with a 1 nm pitch in a brute force approach, i.e., a substantial increase in resolution over that otherwise required for micron sized particles. It would be more efficient, however, and just as valid to superimpose the tunnel gap distribution on contacting islands distributed on a coarser grid.) The structural modeling requirements have been outlined above.

Electrical modeling requires the addition of the conduction processes discussed below to each of the elements: intra-particle, inter-particle, and contact, with the structural model itself providing the percolation component. Existing models confirm the effects of surface layering at a qualitative level [15], and size effects [13,39], i.e., the increase and decrease in effective conductivity respectively for limited ICA sample dimensions parallel and perpendicular to current flow. In particular, the "short sample" percolation threshold is less than the bulk isotropic value, if the dimension parallel to current flow approaches or falls below the percolation "coherence length," as for practical flip-chip bump-to-pad interconnect. As contact separation decreases to filler particle size (the ACA situation), the threshold obviously tends to 0%. On the other hand, for a short dimension perpendicular to current flow, as, for example, in many experimental test "tracks" along a substrate surface, the percolation threshold increases due to the intersection of percolation paths with the surface. So pad-to-pad "contact" structures tend to possess lower resistivities than the nominal bulk value, and "track" configurations higher resistivities. Flake alignment along the surface tends to produce the opposite effects, increasing the number of inter-particle contacts in z-axis current flow, and short-circuiting the isotropic interior of the "track." Figure 20.10 shows the track conductance variation with thickness. At low thicknesses the track flakes are all layered, but as thickness increases, the proportion of internal disorder increases [40]. Figure 20.11 shows the resistance of a z-axis contact as it is mechanically thinned. The sharp drop in resistance corresponds to the removal of the aligned surface layer [41].

As models improve, one expects eventual quantitative agreement. It is probably satisfactory to continue to ignore the effect of finite polymer conductivity, but this assumption should be checked for new materials. One prediction of percolation theory which has never been validated in these systems is the frequency dependence of the conductivity in terms of the coherence length [42], which appears to be masked by the particle skin effect (see below).

It is noted here that there is evidence of percolation chains dropping in and out of the conduction paths during thermal cycling. This effect is demonstrated by reproducible hysteresis in resistance versus temperature plots during thermal cycling of ICA structures where only a limited number of percolation paths are expected to exist, i.e., for very small contact areas [43].

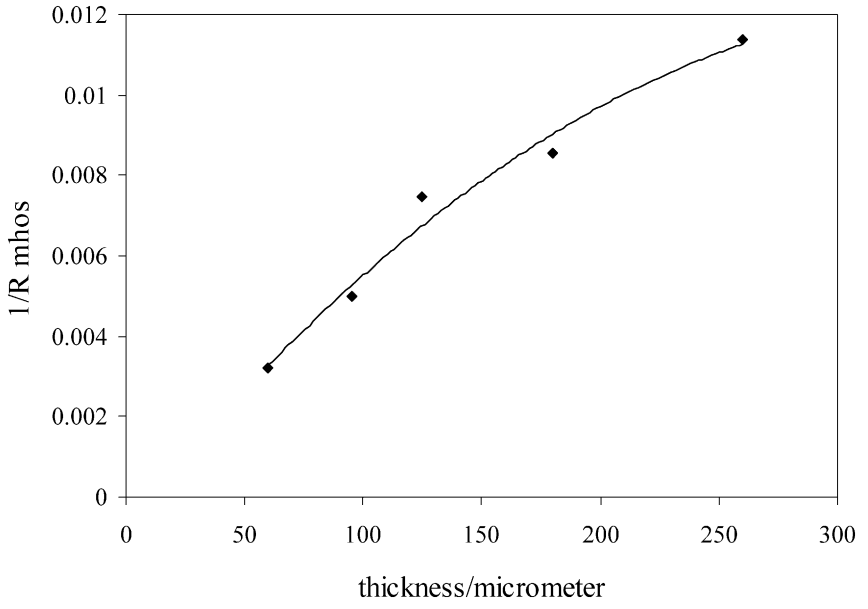


FIGURE 20.10. Variation of ICA resistance with track thickness [40].

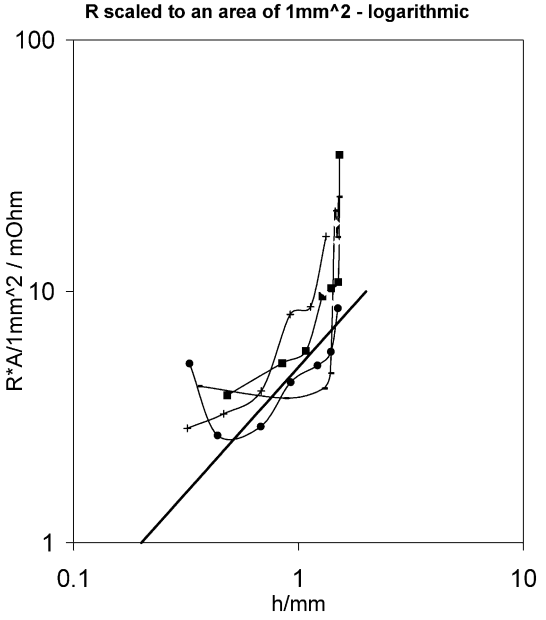


FIGURE 20.11. Log-log resistance variation with thinned thickness. The line of slope for constant bulk resistivity is shown [41].

(2) *Inter-particle conduction* is generally assumed to take place directly from metal to metal, or by tunneling through an insulating layer, whether of intervening polymer or of surface contaminants, or by conduction through a surface oxide film, which

for Ag would be a conductive degenerate semiconductor. There is a range of tools available to characterize conduction mechanisms, including frequency effects, non-ohmic high field behavior, temperature dependencies, etc., and the absence of any observation of negative temperature coefficient of resistance (TCR) or non-ohmic behavior is sufficient to eliminate most other conduction mechanisms from contention. Silver is typically tarnished, and presumably would oxidize within the polymer, even if initially "clean," but it is not clear what effect the surface lubricants identified on flake surfaces [44] would have on this process. An XPS surface analysis of the flake surfaces should be carried out to distinguish between the presence of Ag oxide and the oxygen content of the lubricant and/or polymer residue.

No matter which of the mechanisms apply at the gap itself, the contact area is accepted as being typically small, of diameter, d, 10 nm or less [6]. Clearly there will be some constriction of current flow between particles at the contact, and the Holm theory specifies this contact resistance to be $\rho/2d$ [2]. With the constriction resistance proportional to d^{-1}, and the tunneling or oxide resistances proportional to d^{-2}, one should be able to identify the dominant contribution from a pressure dependence, which has been theoretically matched to experimental data for a d^{-2} dependence, but not for an ICA [45]. The internal pressure exerted by the curing process must be quantified for further progress to be made. Returning to the differences between the contact conduction mechanisms, the TCR will be zero (or slightly negative) for tunneling and positive for the oxide. Unfortunately, it is difficult to find data on the exact electrical properties of the oxide, which are subject to local formation conditions, so it is not known whether the TCR would be less or greater than the metal particle TCR. It would appear from the formula that the constriction resistance would have the same TCR as the metal particles', but the derivation does not include the mean free path reduction which will be associated with contact dimensions less than the bulk value in the particles; this extension to the theory is necessary.

The inter-particle conduction mechanism remains undetermined, and the points made above, which reflect those in the literature, are very general. What is needed is a comprehensive basic study of the metal-polymer interface, to investigate charge transfer and band effects in the polymer(s), time and temperature effects, and how the process proceeds during curing. In addition, actual contact points need to be located and isolated, and the potential distribution plotted across the boundary from one particle to the other. Electrical noise measurements are also often a useful diagnostic tool, and there is noise data in the literature [46,47], but as is often the case, the interpretation is ambiguous. Wong has studied the role of the flakes' surface lubricant, and has achieved reduction of overall resistance by replacing the traditional stearic acid with shorter chain alternatives [36,48]. Benson, on the other hand, has shown that the lubricant breaks down during cure, and leaves a carbon residue on the flake surface [49], which is expected to control the inter-particle resistance.

Frequency dependencies are easily determined and can be definitive in the identification of some conduction mechanisms. In the ICA case, ac measurements were expected to short circuit the tunnel gaps between particles, with the corner frequencies providing the means to separate out the particle resistance from the inter-particle gap contribution. The method was validated by ac measurements before cure. Resistivities below the percolation threshold decreased with frequency

to limiting values similar to those above it [50]. (This experiment suggests the use of impedance spectroscopy as a manufacturing quality test for ICAs, as for solder paste.) When applied to cured ICAs, however, no such effect was observed (see below), so either there is no tunneling gap (i.e., conducting oxide or metal–metal contact) or the tunnel resistance is just much less than the particle resistance [2,4].

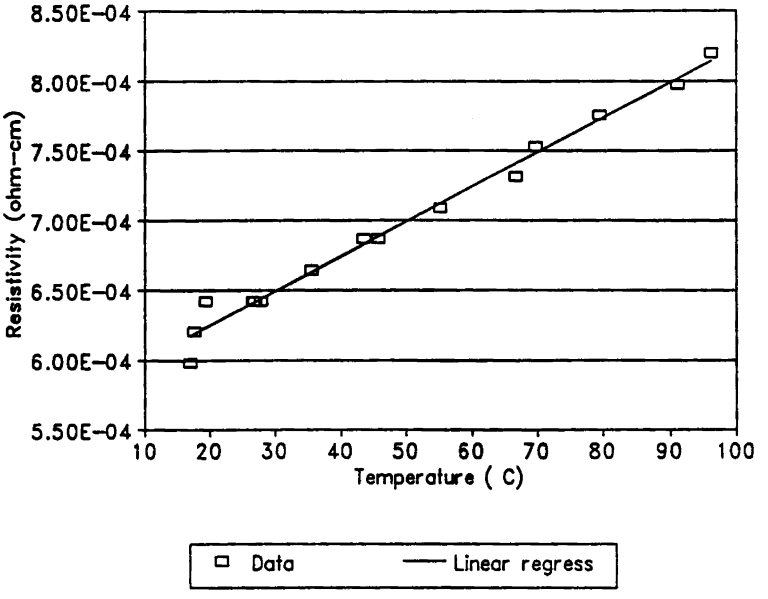
(3) *Intra-particle resistance* accounts for a substantial proportion of the measured ICA resistance, and in some cases essentially all of it. This conclusion is based on temperature coefficient of resistance TCR measurements on a variety of commercial ICAs, where the TCR values range downward from the bulk metallic value, but are always positive [2,4,42], and hence clearly indicative of the dominance of the metallic component of the resistivity. (The other possible contributions to resistance all have zero or negative TCRs.) Within the limits of experimental accuracy, the data are consistent with a model of the intra-particle metallic resistance in series with a zero TCR contact resistance, but cannot be totally conclusive (Figure 20.12). Thermal testing needs to be extended to much lower temperatures to resolve this point.

For ten micron diameter flakes one micron thick, and micron-sized smaller particles, the electron mean free path (mfp) is not going to be reduced significantly from the bulk value, and no accounting is needed for size effects in the particles. (Note that this would not apply to the nano-particle ICA variant [51,52], where the mfp is limited by the particle dimensions.) But the nature of the surface could be important for the assessment of mfp limitation for constriction resistance (with rough surfaces limiting the mfp by random "diffuse" scattering, and with "specular" reflections from smooth surfaces having no such mfp effect).

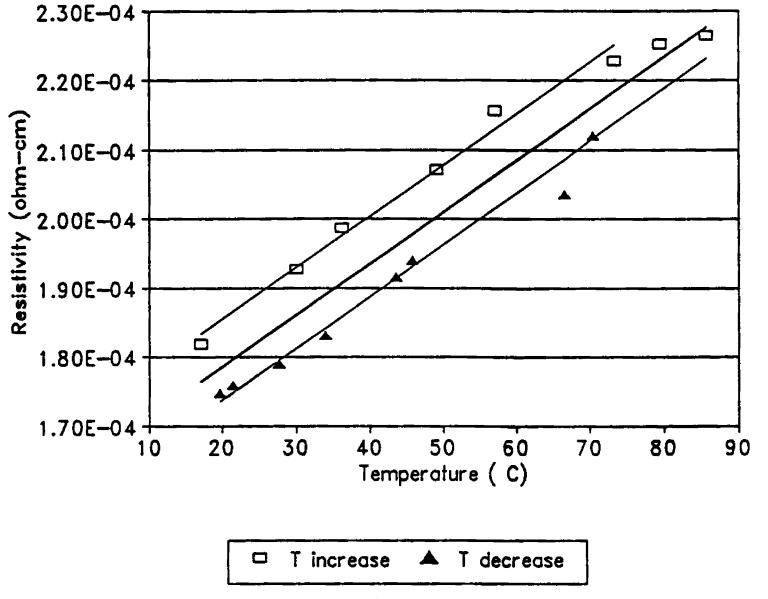
The ac measurements mentioned above were run on the same ICA materials which gave TCRs identical to the bulk value for Ag, and so it is not surprising that the ac characteristics were in total accordance with the predictions of skin effect resistance and inductance for Ag [4]. These experiments should be duplicated for materials with greater inter-particle resistances.

At frequencies where skin effect is dominant, the lower resistance advantage enjoyed by solder disappears, as the effective cross-sectional area shrinks with the skin depth for solder and ICA alike [41].

(4) *Contact resistance* can be isolated from the bulk composite resistivity by the combination of three-terminal measurement with the more common four-terminal (see below) [26]. It is the contact resistance which has been shown to be the source of electrical reliability problems [53–55]. The oxidation (corrosion) of Cu or the Sn in Sn/Pb contact pads or Pb coatings has been demonstrated, and explains the greater long term stability of noble metal contacts (Au or Pd under thermal cycling and 85/85 stress testing). More recently galvanic corrosion has been identified between dissimilar metals [56,57]. One would fully expect to see the effects of interfacial diffusion in the longer term (albeit limited by low process temperatures), and the formation of brittle inter-metallics; although these would probably have no discernible effect on mechanical reliability, the electrical impact could be significant, given the limited number of low diameter contact points to the percolation paths. The solution to these problems would be to eliminate the dissimilar materials, and the use of silvered contacts with Ag-based ICAs would seem logical. However, Ag on Cu introduces much the same problems, and requires a barrier layer (of Ni, for example).



(a)



(b)

FIGURE 20.12. ICA resistance versus temperature. The slopes match the TCR of Ag [4]. (a) Brass stenciled CT-5047-02 thermoset sample (TCR = 0.0039/°C). (b) CSM-933-65-1 screen printed thermoplastic sample (TCR = 0.0038/°C).

The high-frequency ICA data of Li et al. [4] have been extended by Wu et al. [58] and by Dernevik et al. [59,60] Li and Wu focused mainly on the MHz region, and Dernevik on the GHz region. Wu reported that ICA joints can change their high frequency properties during bending.

20.4.2. Electrical Measurements

Resistivity measurements must be made on genuinely isotropic samples of suitable size, unless size or layering effects are the actual object of the measurement. On the other hand, the interconnect application will actually include both, as described above, so will measurements along a long thin sample, but with the opposite effects [2]. The measurement of small resistances with sufficient sensitivity to detect early corrosion, etc., is difficult and usually accomplished by depositing the long specimen just mentioned or by daisy-chaining multiple interconnect samples. In future, the impedance transformer should see increased application to solve this problem [2].

Printed ICA "tracks" are also commonly used for 4-terminal and 3-terminal measurements (Figure 20.13) which provide separation of bulk and contact resistances, an essential step to interpretation of any resistance data. Finite geometries can lead to errors, however, if care is not exercised (Figure 20.14). The same sort of problem can be experienced with z-axis samples (and with ACA testing) due to finite track resistances comparable to the sample's (Figure 20.15) [41]. However, self-consistent results have been demonstrated [61].

20.4.3. ACA

We move now to conduction through an individual particle between two contact pads.

The electrical conductivity in an ACA joint has been estimated by Williams et al. [23]. Assuming both hard and soft spheres, plastic deforming lands and spheres, the joint contact resistivity (ρ) has been estimated as:

$$\rho = \frac{A\rho_B\left(\sqrt{\dfrac{6\pi n\kappa}{\sigma A}} - \dfrac{1}{R_B}\right)}{4\pi n R_B} \tag{20.17}$$

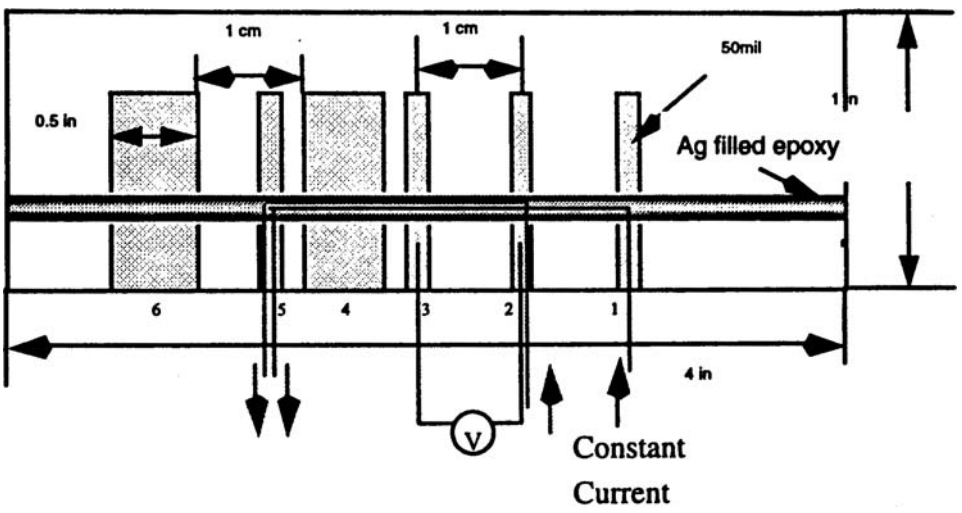


FIGURE 20.13. Four-terminal and three-terminal measurements [2,26]. The test current is injected at the right hand terminal for the usual 4-terminal measurement. Injection into the voltage test terminal includes that contact to the sample. The difference of the two measurements yields the contact resistance.

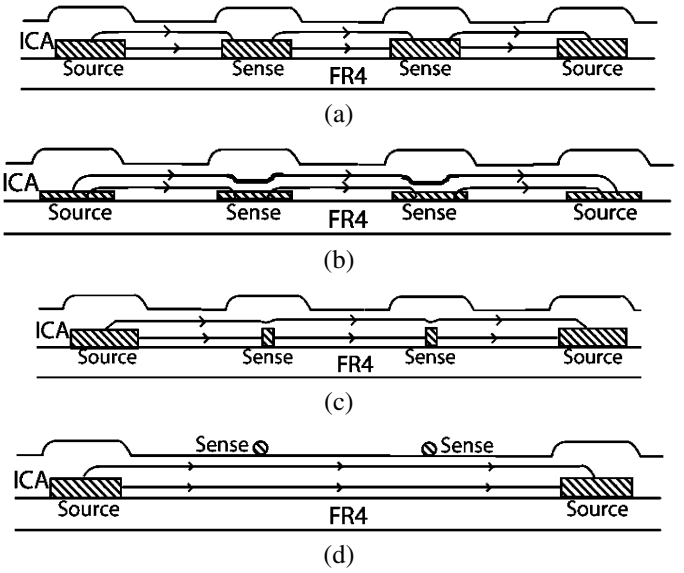


FIGURE 20.14. 4-point measurement using ICA track across proto-board current lines [41]. (a) Sense lines short ICA, (b) thinned contacts, (c) trimmed contacts, (d) surface point contacts.

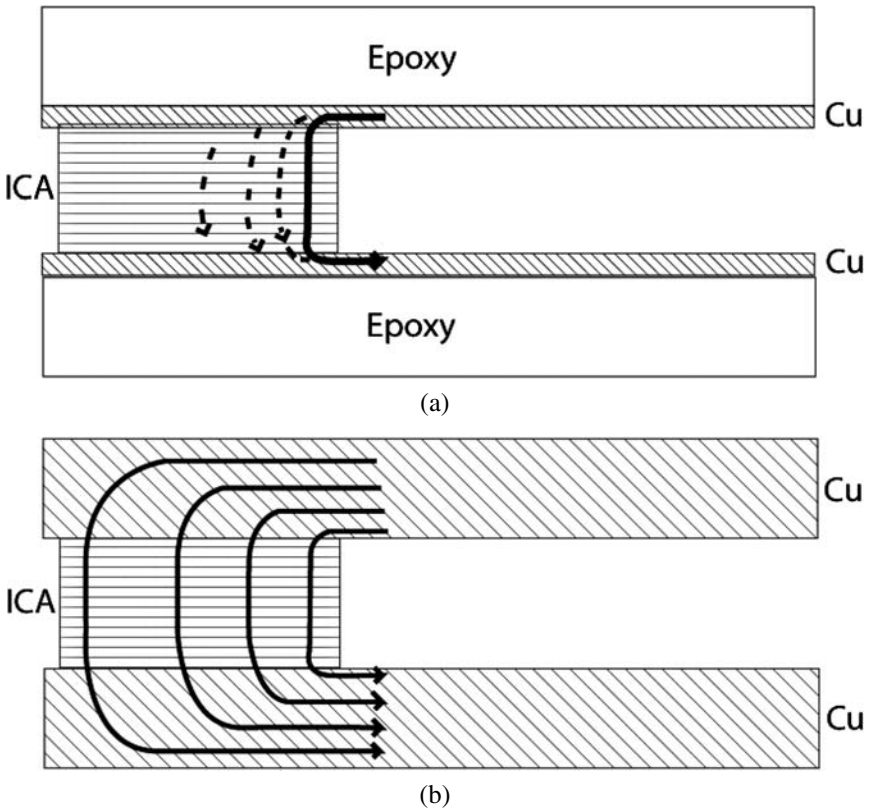


FIGURE 20.15. Current crowding effects due to contact track thickness [41]; (a) thin PWB track, (b) thick Cu plates.

for a single particle, where ρ_B is the resistivity of the sphere material, n is the number of contacts within the contact area A, κ is the shear yield stress on a single contact sphere of radius R_B, and σ is the pressure applied to the joint. Using Equation (20.7) one can calculate the electrical conductivity through an ACA joint. For instance, for 25 micron radius Ni alloy particles with $A = 0.1$ mm^2, $\kappa = 1$ GPa, $\rho_B = 6.1 \times 10^{-8}$ Ω m and $\sigma =$ 1 MPa we obtain joint conductivity of 4×10^5 Ω^{-1} m^{-1} with 50 conducting particles and 1×10^5 Ω^{-1} m^{-1} for 2 particles.

Shi et al. [25] have calculated the resistance of a solid spherical contact particle analytically as the sum of the bulk resistance and constriction resistance components, and then extended the result to include distributed particle sizes and the influence of pressure on the particle contact area. Both elastic and plastic deformation models were included. Yim and Paik [62] also developed an analytical model, but also, like Oguibe et al. [63], performed finite element modeling of both solid sphere and coated polymer particle systems. There is direct comparison with experiment and good agreement is claimed. Meanwhile, Fu et al. have found that the particle location in an ACA joint can affect the electrical conductivity [64]. Generally speaking, a particle in the center of the joint contributes much more to the electrical performance than a particle close to the edge of the joint, which helps to explain the large scatter observed in single joint resistance values.

The high frequency properties of ACAs in packaging have also been studied by several groups [65–67]. ACA can offer similar high frequency flip-chip packaging performance up to 20 GHz [65] as solder. Significant transmission loss is due to the silicon material itself. It is characteristic of experimental comparisons of ECA joints with solder at high frequencies that the results owe more to the board lines and connections than to either the solder or ACA joints, which both contribute comparably negligible loss to the overall system. Kim [66] has used electrical equivalent circuit modeling of the ACA flip-chip interconnect in conjunction with experiments to extract the high frequency effect of the ACA interconnecting material itself. Fu et al. [67] used a physical modeling approach to understand the high frequency performance of the ACA joint. They found that at extremely high frequency, the capacitive displacement bypasses the filler contacts, and the high frequency performance is largely independent of the particle size and number. It is therefore doubtful if arrayed structures can improve high frequency ACF performance.

20.5. MECHANICAL PROPERTIES

20.5.1. ICA

The vast majority of the published ICA data is comprised of electrical resistance and adhesive strength measurements, both presented in the context of reliability testing. But while there has been some parallel effort to interpret the electrical properties in terms of both structure related and physical failure models, there has been little similar effort to understand the mechanisms of adhesion at a comparably fundamental level. Obviously, such a study requires a systematic approach to the measurement of adhesive strengths for a matrix of metal surface and polymer combinations. It is clear that surface cleanliness plays a crucial role in effective adhesion (and one which may be overlooked in the desire to apply ICAs as drop-in replacements for solder), so surface treatments must be included as a secondary variable, with consideration of roughening effects also included. While this could be regarded as an empirical study, the fundamental goal of determining the relative contri-

butions of various adhesion mechanisms (e.g., chemical, mechanical, electrical) should not be lost sight of.

Plasma cleaning of the adherent surfaces would seem to be a logical step, but so far preliminary data shows no improvement in adhesive strength with either Ar or O_2 plasma treatments, despite the demonstrated removal of organic contaminants and oxides [12,59]. Wolter et al. [68,69] have demonstrated that it is the polar component of surface energy which is increased by plasma treatments. Experimental studies consistently show that the mechanical component of adhesion dominates [70,71], with best results from surface roughening (which may be accomplished by high energy plasmas). (A simple NCA shows good electrical stability, provided the contact surfaces are roughened [40].)

Published ICA adhesive and shear strengths are on the same order as those of solder, usually a little less [72], occasionally higher [69], but anyway adequate. The problem has been the drop test failure rate, leading to the widespread adoption of the NCMS (National Center for Manufacturing Science) criteria [73] as a de facto standard. In general, it is the larger devices which are most at risk, and indeed current commercial materials seem adequate for smaller devices such as SMT passives, which have been in mass production with ICA attachment for some years, e.g., in automotive electronics. Improved understanding of this particular phenomenon is currently leading to the development of ICA materials specifically designed to address the drop test problem. The key lies in the imaginary component of the complex Young's modulus, which represents energy dissipation in the material, as opposed to the energy storage of the simple deformation represented by the conventional form of Hooke's Law. The complex modulus is therefore directly analogous to the complex dielectric constant, and dynamic stress–strain relationship actually includes a phase shift. When one examines drop test survival data, the success rate correlates with the (imaginary) dissipation modulus rather than with adhesive strength [74,75]. One way to design materials with high dissipation modulus is to select polymers with glass transition temperature T_g below the operating range, i.e., below room temperature, in general [76]. However, operating polymers above T_g carries its own penalties, e.g., higher temperature coefficients of expansion, so the next step is to develop adhesive polymer blends with high mechanical absorption without those disadvantages.

Figure 20.16 shows the improvement in drop test results for a commercial ICA with the addition of a pre-cure heat soak to the processing schedule [77]. The additional step enables the material to survive an additional 20 4-foot drops and at least 10 more from 5 feet, greatly exceeding the NCMS standard. Improvement in resistance, adhesion, and drop test results are all tentatively attributed to the elimination of bubbles at the contact interfaces, where they reduce the effective contact area.

20.5.2. ACA

Several research groups have now studied the deformation effect on the electrical conduction development during the ACA assembly [58,62,64,78]. The key work has been done by Wu et al. [58] where they consider two types of particle situations as shown in Figure 20.17 (rigid and hard particle) and in Figure 20.18 (soft particle).

In addition, Wu et al. have also successfully deduced an equation that describes the relationship between the resistance and bonding pressure. These are shown in Figures 20.19 and 20.20 for the rigid and deformable particles respectively.

Wang et al. also focused on quantification of criteria for a good ACA flip-chip joint, as this is one of the most important issues during the ACA bonding process [79]. The

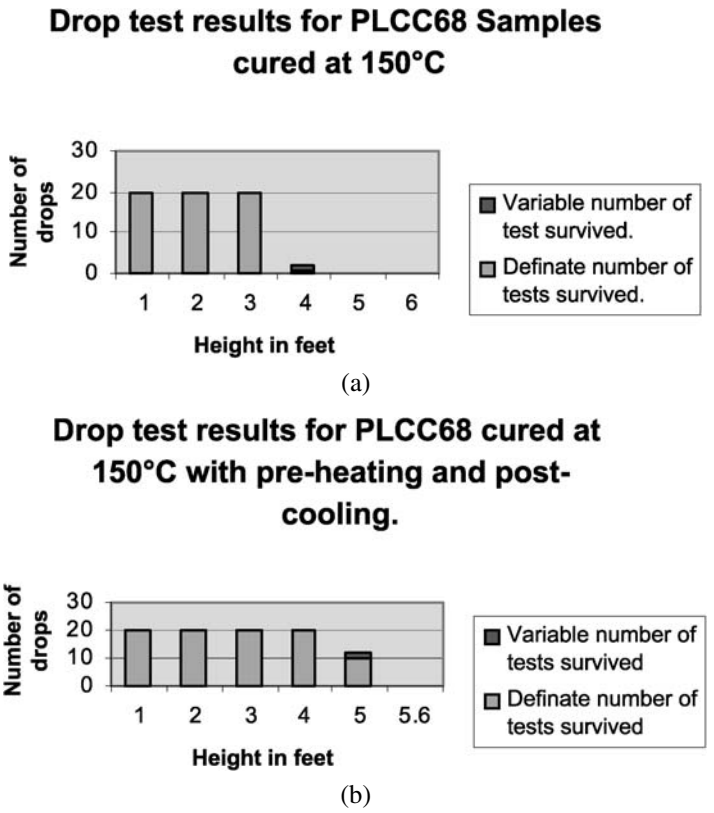


FIGURE 20.16. ICA drop test improvement from (a) to (b) with pre-heating before cure [77]. Each sample device was dropped 20 times from 1 foot, then 20 times from 2 feet, then 20 times from 3 feet, etc., until failure.

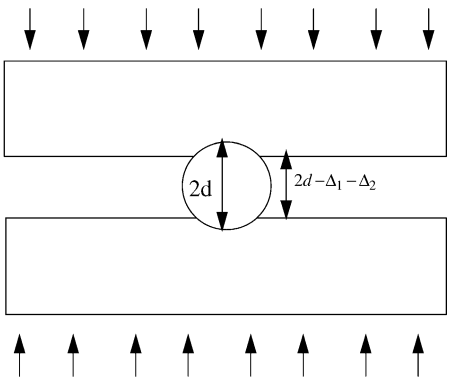


FIGURE 20.17. Schematics of a rigid particle system (courtesy of Wu et al.) [58].

purpose of their work was to study the relationship between bonding pressure and the deformation of conducting particles in ACAs. It has been shown that the deformation of filler particles plays an important role in both bonding quality and the reliability of ACA joints. The deformation degree as a function of bonding pressure during ACA film flip-

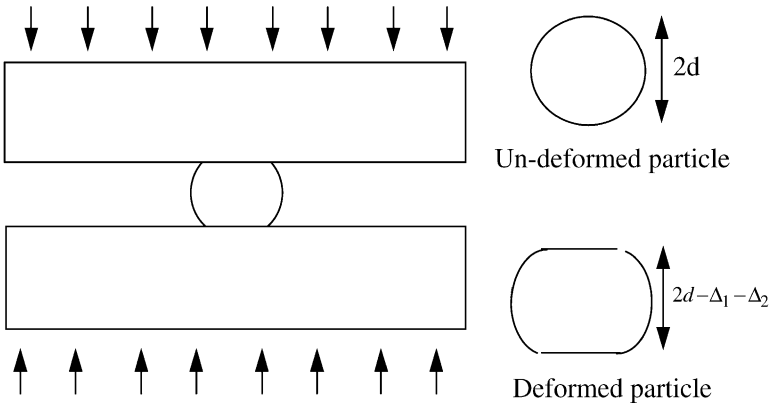


FIGURE 20.18. Schematics of a deformable particle system (courtesy of Wu et al.) [58].

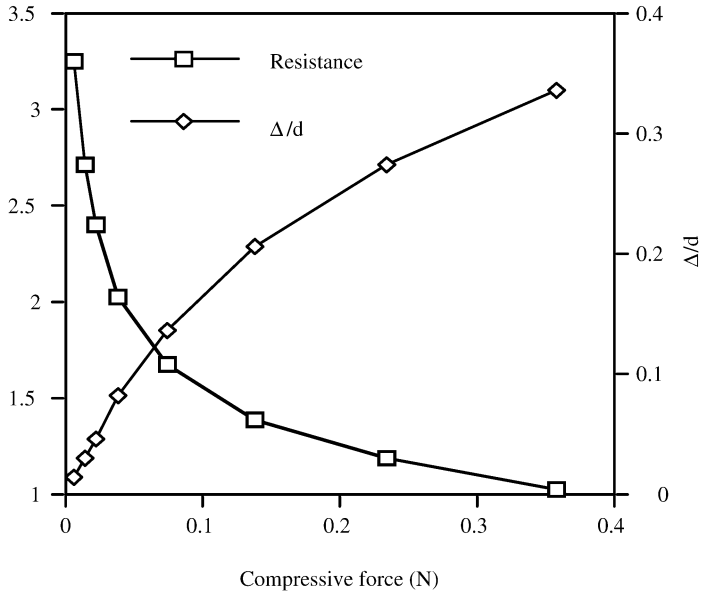


FIGURE 20.19. Force–resistance–deformation relationship for a deformable particle system (courtesy of Wu et al.) [58].

chip bonding is shown in Figure 20.21. As can be seen, deformation and bonding pressure exhibit a linear relationship.

As the mechanical properties of a thin layer adhesive material differs from its bulk material properties, it is important to know the ACA material properties for mechanical model and simulation purposes. Because of this, Young's modulus and Poisson's ratio have been determined by Zribi et al. [80] for an ACF over the temperature range of 15 to 60°C (Figure 20.22).

A group at Chalmers University has recently carried out theoretical simulation in order to explain the microscopic mechanism of the electrical contact conduction through the metal fillers for ACAs. By comparing with experiments performed by Wang et al. [79], it was concluded that the deformation of the metal filler is plastic even at rather low external

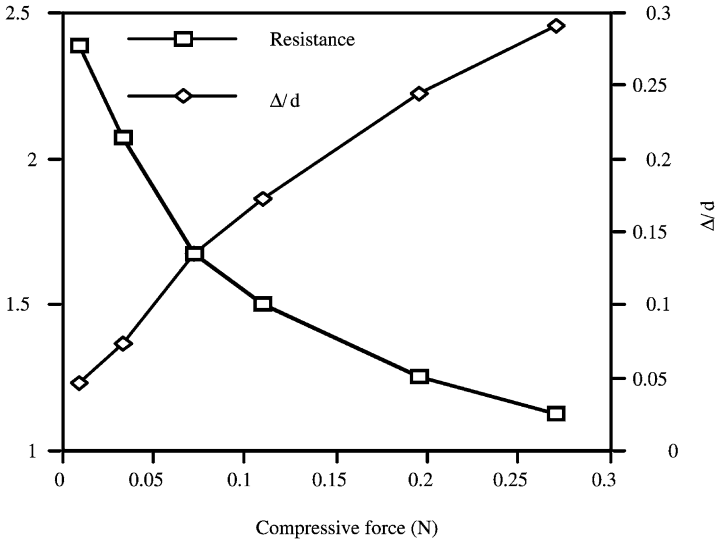


FIGURE 20.20. Force–resistance–deformation relationship for a rigid particle system (courtesy of Wu et al.) [58].

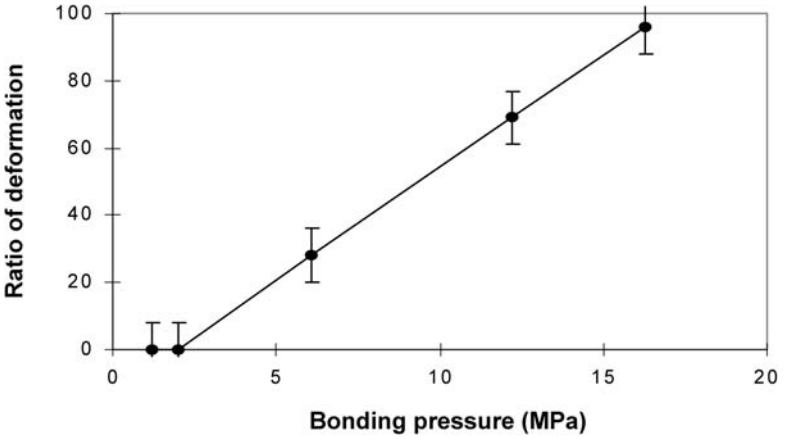
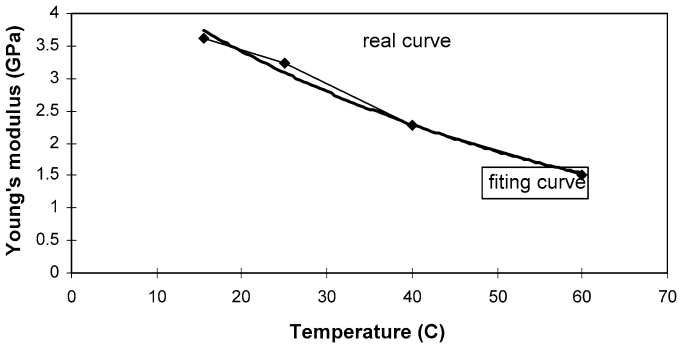


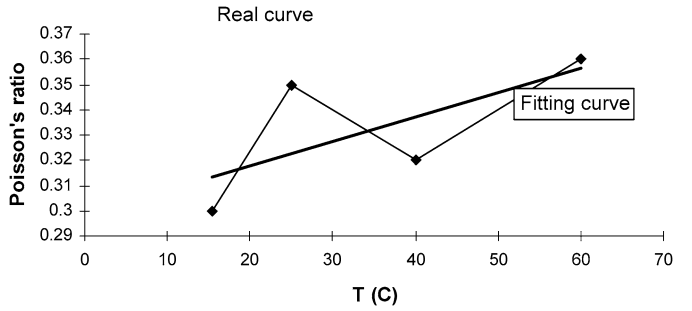
FIGURE 20.21. The relationship between nickel particle deformation and bonding pressure during ACA flip-chip bonding [79].

loads. Further theoretical simulation reveals two aspects of the conductance characteristics: the conductance is improved by increasing the external load, but the dependence of the conductance becomes stronger on the spatial position of the metal filler [65].

The consequence of the bonding pressure during the ACA bonding on the stress generation is evident. This stress generation is probably the reason for catastrophic failure. Wu et al. have simulated this using finite element modeling. They have found that both for rigid and deformable particles, significant stress is built up in the interface between the two contacts [58]. Wang et al. used a similar approach to calculate the stress after various degrees of deformation [65,69]. It is clearly shown in Figure 20.23 that the highest stress point is in the edge of the ACA particle.



(a)



(b)

FIGURE 20.22. (a) Young's modulus, and (b) Poisson's ratio of ACA film vs. temperature after cure [80].

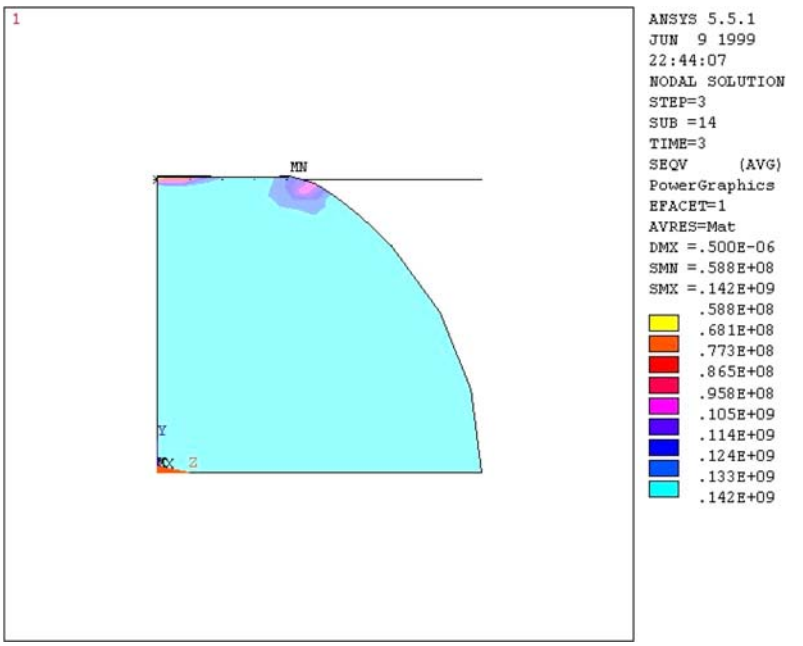


FIGURE 20.23. Stress profile at the reduction in height of 0.5 μm [20].

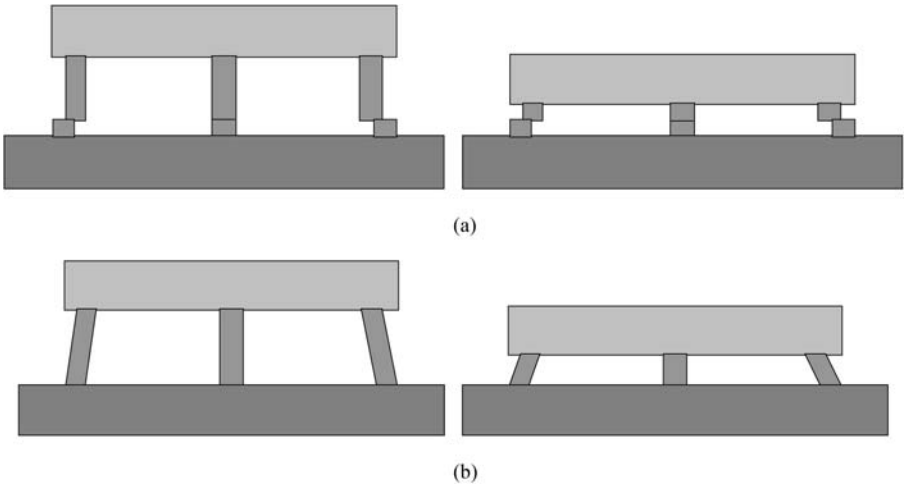


FIGURE 20.24. Deformation mechanism of the ACA joint due to thermally induced stress [81]. (a) Strain caused by thermal mismatch cannot change with bump height in ACA joint. (b) Schematic diagram showing that thermal mismatch strain reduces as bump height increases for solder joints.

Lai et al. have proposed a similar failure mechanism of the ACA flip-chip joint structure [81] as can be seen in Figure 20.24(a). In this case, the particles are exposed to large shear stress due to the expansion of the substrate. Therefore, it seems that the stress concentration is the most critical parameter that governs the reliability of the ACA joint. Therefore the failure mechanism of ACA joint differs from the traditional solder joint failure mechanism where plastic strain is the most critical parameter that governs the joint reliability, see Figure 20.24(b).

The most complete ACA joint stress analysis has been done by Wu et al. [82]. It is shown that the residual stress is larger on a rigid substrate than on a flexible substrate after bonding.

As an ACA bonded module may be used in connection with soldering technology for a final product, we need to know if the ACA material can withstand the soldering temperature. Sugiyama et al. reported that the ACA material from Sony can withstand the soldering profile three times and without causing reliability problems [83]. Törnvall reported different results in terms of soldering effect where he showed that an ACA flip-chip module cannot withstand a normal soldering process [84]. More research is therefore needed to clarify this matter further.

Oxidation is an ACA joint failure mechanism that was identified at an early stage [85]. The idea was that the particle is oxidized causing electrical performance decrease. For a parabolic oxide growth law, and assuming the resistance change, ΔR, is ohmic, then the change with time, t, is given for contact area, A, by

$$\Delta R = (2Dt)^{1/2}\rho_{\text{oxide}}/A, \tag{20.18}$$

where D is the diffusion constant of oxygen in the metal filler, and ρ_{oxide} is the volume resistivity of the oxide.

20.6. THERMAL PROPERTIES

20.6.1. Thermal Characteristics

The thermal performance of an adhesively assembled chip is of vital interest as power dissipation in the chip increases. Power dissipations have been simulated by Sihlbom et al. for both ICA and ACA flip-chip joints [86]. They concluded that the ACA flip-chip joint is more effective in transferring heat to the substrate from the powered chip than the ICA joint, because the adhesive thickness is so much thinner than the ICA joint.

20.6.2. Maximum Current Carrying Capacity

Dernvik et al. have also studied the effect of maximum current carrying transmission capacity through the ACA adhesive joints at 3.2 GHz [59]. The copper bridge structure was subjected to a maximum transmission of 25 W of average pulsed power for 10 minutes, with a pulse length of 10 μs and peak power of 250 W at duty cycles of 1, 5 and 10%. The result indicated that bonding pressure has a strong influence of the joint quality. At 150 N, no electrical transmission loss is observed, but for 75 N bonding force, some deterioration was observed after the final power exposure; however, the effect was small, at about 0.5 dB.

In a similar dc study of three ICAs, Morris et al. [12] found that failure correlated directly with temperature rise, which in turn correlated with the joint resistance, and hence to power dissipation. The breakdown mechanism was clearly polymer degradation, accompanied by the emission of noxious fumes. The onset of breakdown shows thermal runaway (Figure 20.25), as measured at the adhesive surface. Comparisons of surface temperatures with estimates of internal temperatures (from resistance values and the previously mea-

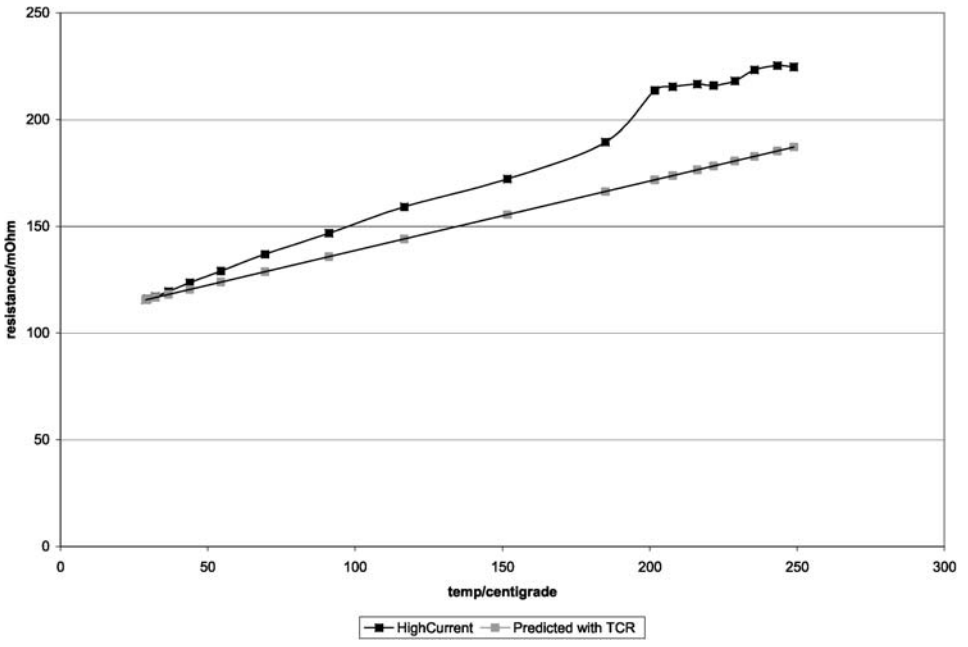


FIGURE 20.25. ICA resistance and surface temperature, as current is increased to thermal runaway and destruction [40]. The straight line corresponds to internal temperatures estimated from the resistance and the TCR.

sured temperature coefficient of resistance (TCR)) suggest that the internal temperature at this point is a good 60–80°C higher than the 200°C shown. While two of the ICAs failed like this at 9 A/mm^2, the other showed no ill effects whatsoever.

Bauer et al. has studied the power dissipation performance of NCA joints under dc conditions. It was found that the NCA joints still show ohmic behavior up to 25 A/mm^2 [87].

20.7. RELIABILITY

20.7.1. ICA

(1) *Cure schedule* control is undoubtedly very important for joint reliability. It seems that the electrical resistance of the joint is related to the curing degree, especially for non-noble metal surfaces, as can be seen in Figure 20.26. Figure 20.26(a) shows the contact resistance vs. curing time for an epoxy conductive adhesive cured at 150°C, and following 1000 hours of damp heat treatment at 85°C, 85%RH [88]. The corresponding curing degree varies between 65% and 90%, determined by DSC. Below a critical curing degree (77% for this adhesive), the electrical resistance of the joint increases significantly, because an incompletely cured epoxy can absorb a significant amount of moisture, which in turn causes oxidation/hydration of the Sn37Pb bonding surface in Figure 20.26(a) [89] and less crosslinking/shrinking of the polymeric matrix. If a noble metal, for instance Au or Pd is used as the bonding surface, no electrical resistance change is observed, despite the fact that the curing degree can be very low [Figure 20.26(b)]. Once a critical curing degree is reached (72%), it seems that the shear strength of the joint on the Sn37Pb bonding surface can be maintained at a constant level as is illustrated in Figure 20.27(a). However, on the noble metal bonding surface, the shear strength of the joint is almost independent on the curing degree in the range from 67% to 92% [Figure 20.27(b)]. In summary, it can be said that a minimum curing degree appears to be required to provide a certain level of mechanical and electrical performance in the adhesive system. Once this is achieved, increasing curing times do not result in significant improvement. At full cure conditions, however, the electrical resistance and the mechanical strength of conductive adhesives are

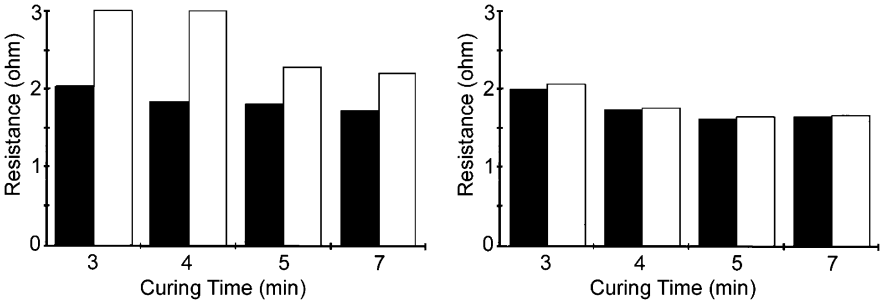


FIGURE 20.26. Series contact resistance of 10 epoxy-based ICA mounted chip components in series, before and after 1000 hours of 85°C and 85%RH [88]. (a) Sn37Pb plated chips and boards; (b) Ag/Pd plated chips and Au plated boards.

also guaranteed [90]. These above results also indicate that for conductive adhesive joining, noble metal surfaces are preferred to non-noble metal surface, due to the absence of galvanic corrosion.

(2) *Moisture effects* on the polymer degradation in conductive adhesives have been studied by Khoo and Liu [91]. Moisture sorption effects may be reversible or irreversible, and are usually small enough to make detection of the molecular changes during absorption/adsorption very difficult. Figure 20.28(i) shows the FTIR spectra of an ACA (but ICAs would be similar) after curing and subsequent 41 hours conditioning at 85°C and 85%RH. The difference spectrum (a)–(b) represents the changes occurring due solely to exposure to the moisture conditions at 85°C and 85%RH. The most obvious real changes are the negative bands at 868, 916, 1345, 3005 and 3058 cm^{-1}, implying decreasing epoxy functionality, and thus further progress of the cure reaction. The new bands at 3560 and 3350 cm^{-1} may both be attributed to hydroxyl groups, of which the former are free groups, which could be formed on further curing, or as an oxidation product resulting from thermo-oxidation/degradation processes. The latter are attributed to hydrogen-bonded hydroxyl groups, indicating the type of bonding of the adsorbed water to the epoxy resin. The slight rise at about 1640 cm^{-1} indicates the presence of absorbed water in the epoxy resin. Finally, new ester linkages indicative again of further curing are indicated by the presence of a broad absorption region between 1000 and 1300 cm^{-1}.

Figure 20.28(ii) compares the molecular events happening on further exposure to these same conditions. The figure shows the difference spectra (a) after 41 hours conditioning, (b) after 162 hours conditioning, (c) after 821 hours conditioning, all at 85°C and 85%RH. It is observed firstly that the subtracted spectra all show the same profile, indicating that the subtraction procedure has been consistent and that the subtraction spectra are valid, but a closer scrutiny also shows that the bands at 3560, 3350, 1640, 1573 cm^{-1} are increasing in intensity with increasing conditioning time. Both the increases at 3350 and 1640 cm^{-1} follow the increasing adsorption of water in the epoxy. The steadily growing absorption at 1573 cm^{-1} is tentatively attributed to unsaturated vinyl structures ("C=C") which are formed as a result of degradation actions. Moisture degradation is felt to occur by hydrolysis of the ester linkages ("R−(C=O)−OR"). Such hydrolytic attack breaks the polymer chain creating two new end groups, a hydroxyl and a carbonyl. Although

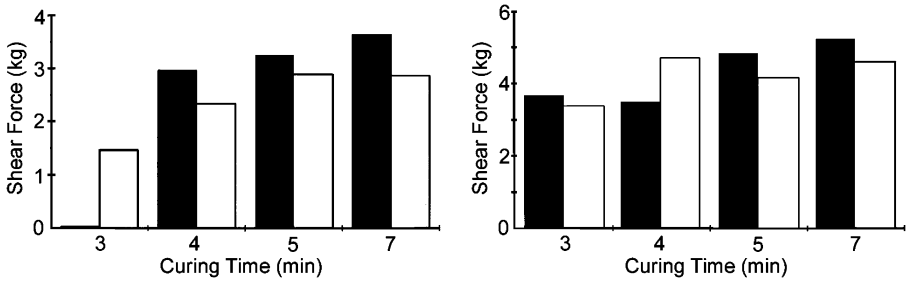
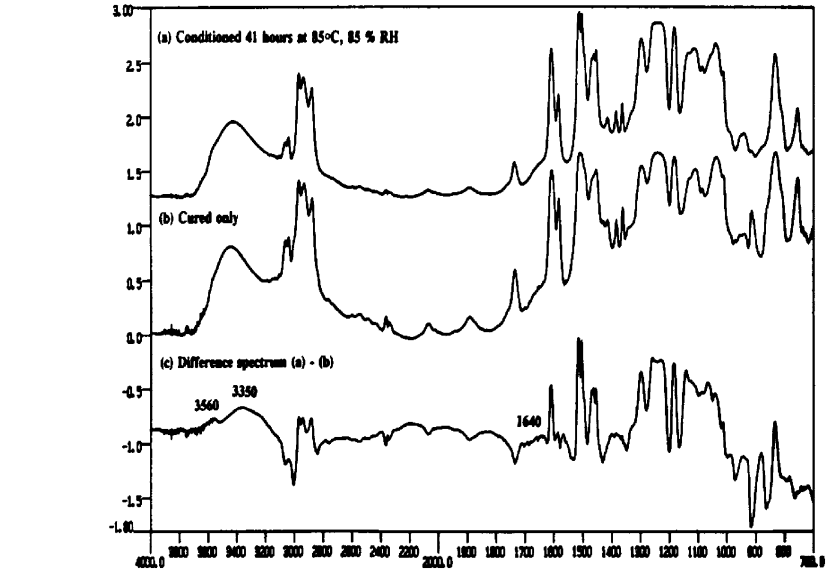
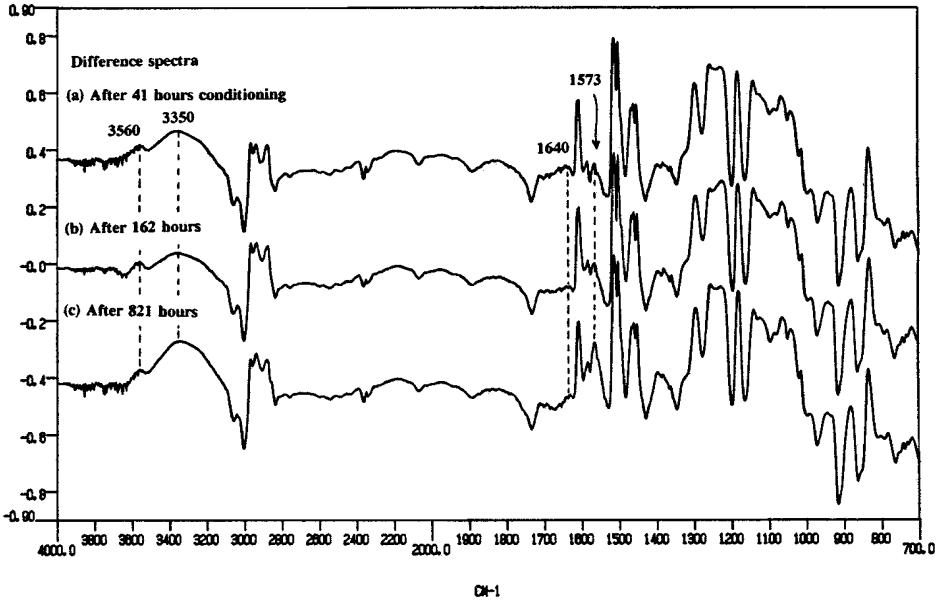


FIGURE 20.27. Average shear strength of 10 epoxy-based ICA-mounted chip components in series, before and after 1000 hours of 85°C and 85%RH [89]. (a) Sn37Pb plated chips and boards; (b) Ag/Pd plated chips and Au plated boards.



(i)



(ii)

FIGURE 20.28. FTIR results for cured ACAs [91]. (i) FTIR Spectra (a) after 41 hours of 85/85 testing, (b) before the testing, and (c) the difference spectrum. (ii) Difference spectra for (a) 41 hours, (b) 162 hours, and (c) 821 hours at 85/85.

it is difficult to see a new emerging carbonyl group in this figure, the presence of the band at 3560 cm^{-1}, which indicates free hydroxyls, supports the suggestion of degradation reactions occurring with increasing exposure to heat and humidity at 85°C and 85%RH. Hence, in conclusion, it can be said that on exposure of the

cured adhesive to 85°C and 85%RH, both moisture adsorption and further curing can be observed. After a certain time, however, further curing will not be observed, but instead, degradation effects may be seen.

(3) *Galvanic corrosion* at the contact interfaces has been demonstrated by Lu et al. [56,57]. They correlated the degree of contact corrosion (as indicated by resistance increases) to the electrochemical series, and demonstrated the requirement for moisture in the process. With this understanding of the process, it was shown that resistance drift could be inhibited by the addition of corrosion inhibitors, oxygen scavengers, and/or sacrificial anode material to the polymer matrix [92–94]. Figure 20.29 illustrates the classic result. The bulk ICA resistance varies little, decreasing somewhat with further cure. The contact resistance changes little for the Ag filler on a Au surface, but corrosion produces major increases for the Ag/Cu combination.

(4) *Thermomechanical* cycling is the litmus test for all package interconnect, especially for soldered flip-chip, and there are many examples of ICA data in the literature. Polymer cure temperatures are usually less than even the lowest solder reflow temperatures, so initial thermomechanical stress is lower at room temperature. In addition, the ICA's polymer base provides greatly increased creep properties in comparison to its solder competitors [95], so it may relax more readily to a zero stress state. It is therefore not surprising that ICAs out-perform solder on mechanical cycling tests by an order of magnitude (Figure 20.30) [96]. In an SMT application, however, the thermoplastic properties of the polymer lead to the accumulation of plastic strain, which initiates cracking [29].

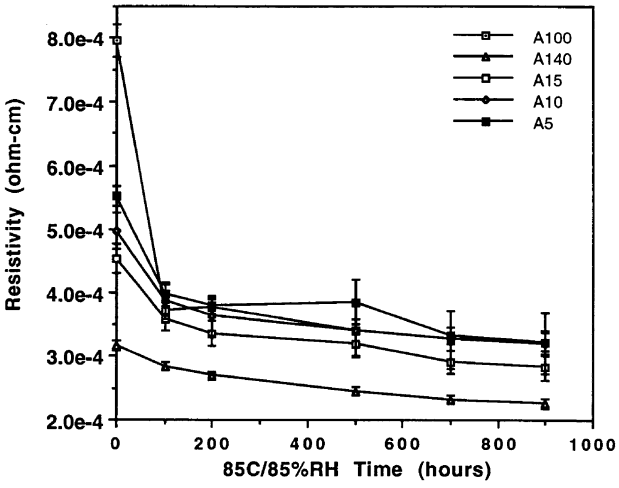
20.7.2. ACA

The bonding pressure, if not homogeneously distributed, can cause uneven deformation of the particles. There are many possible reasons for uneven pressures, but the main reasons are improper electrical routing on the substrate, unadjusted pressure bonding head, etc. It is also clear that on rigid and thick substrates, electrical routing is not so critical, but it is crucial on thin and flexible circuitry.

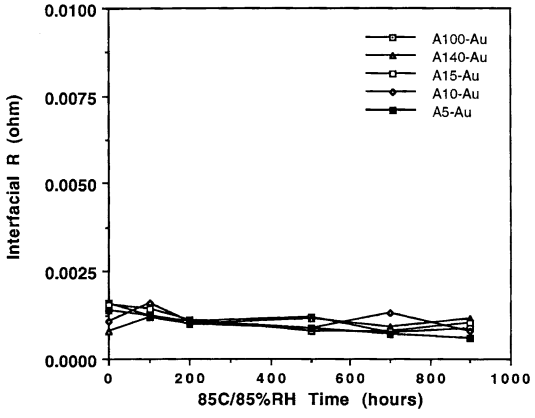
Electrical failures during thermal cycling are observed at both low and high temperatures, and the different types of failure (Figure 20.31) can be attributed to different combinations of particle size and chip geometry, as can be seen in Figure 20.32 [97]. The best case is that the particles are deformed uniformly and metallurgical bonding between the particles and contacts is achieved (Type 1). If the particles are just in contact with the bonding surface, we will have a situation that electrical reliability is not good at high temperature due to the fact that epoxy will expand more than the metal particles (Type 2). If the particles are not the same in size and for unbumped dies and for chips with non-uniform bump heights, one obtains a situation that the smaller particles are not deformed and will shrink more than the big particles causing problems at low temperature (Type 3). The final case is when there is a uniform height of the bump and chip surface but very a large particle size variation. This will cause electrical opens at both low and high temperature (Type 4).

The dynamic pressure on the particle necessary to cause separation from the contact is, for the deformed (polymer particle) case, given by [97]:

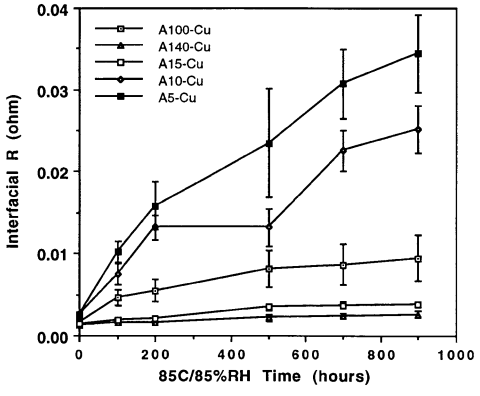
$$P_i = N\left(\frac{1}{b^2}\right)\sqrt{\frac{\rho}{AB}}\frac{l}{\cos\alpha}, \tag{20.19}$$



(a)



(b)



(c)

FIGURE 20.29. Post-cure resistance changes for an Ag-filled ICA [4,6,27]. (a) Bulk resistance; (b) contact resistance to Au; (c) contact resistance to Cu.

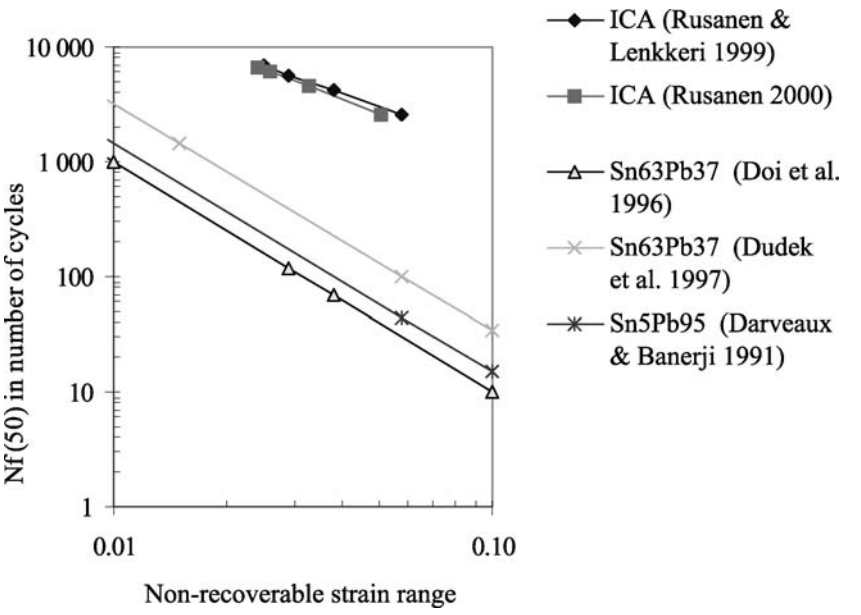


FIGURE 20.30. Strain cycling joint lifetimes according to the Coffin-Manson equation [96].

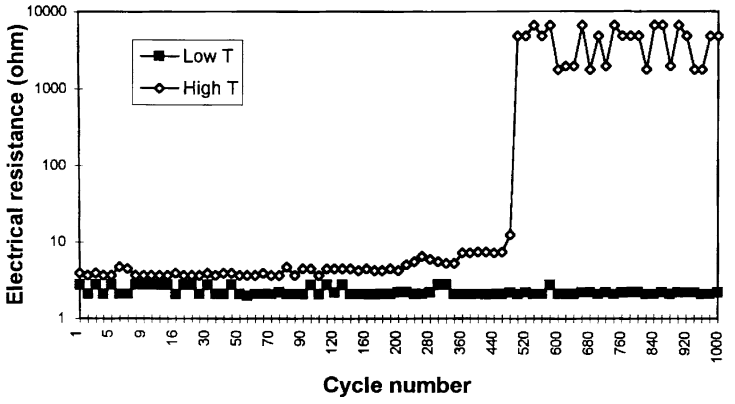
where P_i is the stress necessary for decohesion, N is the bonding force, ρ is the curvature of the particle, b, l and AB are geometric length, and α is the decohesion angle. For the undeformed (Ni sphere) case, a similar equation can be obtained [97]:

$$P_i' = N'\left(\frac{1}{b^2}\right)\sqrt{\frac{\rho'}{A'B'}}\frac{l}{\cos\alpha'}. \tag{20.20}$$

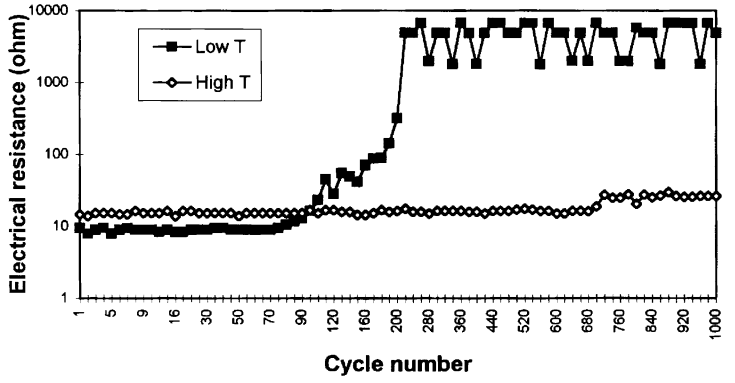
From Equations (20.19) and (20.20) it is clearly seen that $P_i > P_i'$. Therefore, for the deformed case, a larger stress is necessary to cause decohesion and, for the same stress level, it is easier to obtain failure with an undeformed particle.

As shown analytically by Hu et al., the electrical resistance decreases with increasing bonding pressure, until the polymer coating fractures. This has been demonstrated experimentally [14,98].

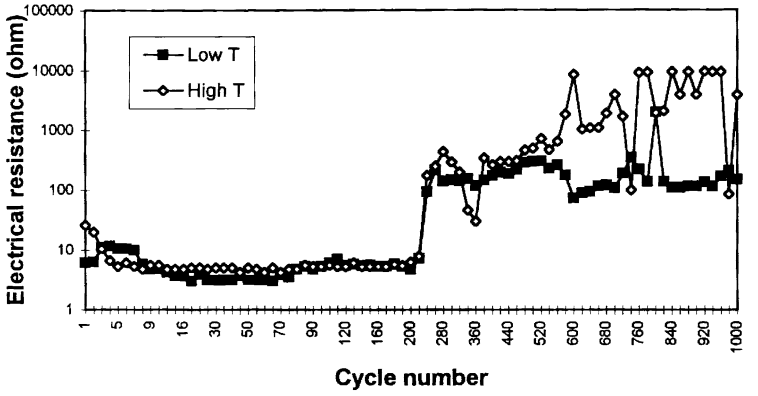
Another important issue is the substrate hardness, geometry and material. A soft substrate material may deform during the bonding due to the softening of epoxy at the bonding temperature, which is above the glass transition temperature (T_g) of the epoxy matrix. On an FR-4 substrate, it has been observed that the electrical conductivity and reliability of a joint depends on the depth of the glass fiber in the substrate where the bump exerts pressure on the pad [99]. Deeper fibers mean thicker layers of soft epoxy that deform more readily during the bonding. Therefore, insufficient particle deformation will be obtained at that point. Shorter distances to the fiber bundles resist substrate deformation better, and hence yield better electrical conductivity and reliability. See Figures 20.33 and 20.34. One approach to reduce the amount the pad can sink into the substrate is to use a relatively large pad area in comparison to the bump area to reduce the pad pressure for a given bonding force.



(a)



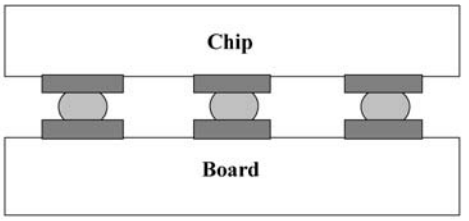
(b)



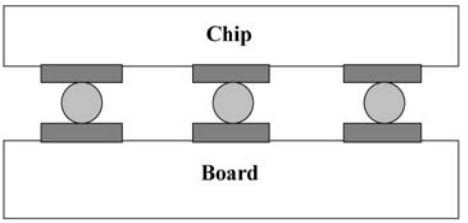
(c)

FIGURE 20.31. High temperature cycling [97]. (a) Type 2 joint failures occur at higher temperatures; (b) Type 3 joint failures occur at lower temperatures; (c) Type 4 joint failure occur in both high and low temperature regions.

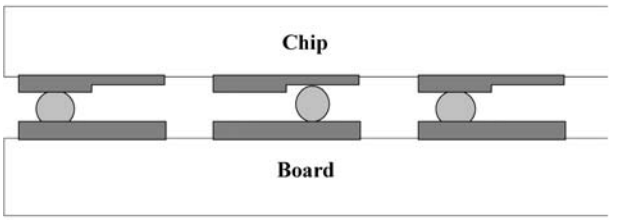
Despite the complex processing and manufacturing conditions, under optimum conditions, good reliability data on ACA flip-chip joint has been reported [100]. Cumulative fails after the temperature cycling test from −40 to +125°C for 3000 cycles with a hold



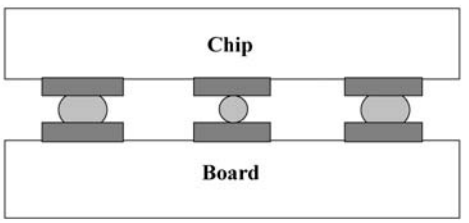
Type 1, the most reliable structure.



Type 2, unstable at high temperature.



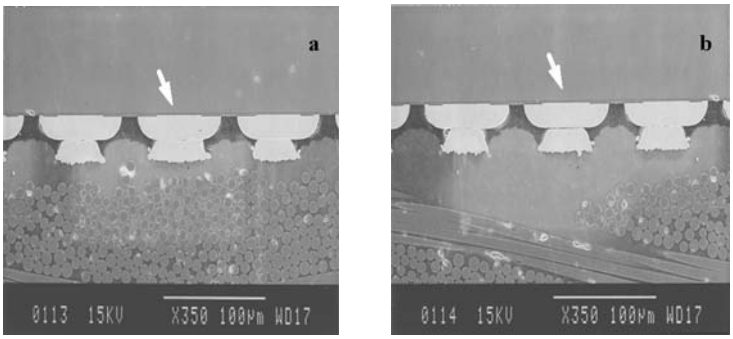
Type 3, unstable at low temperature.



Type 4, unstable when temperature changes.

FIGURE 20.32. Schematics of four types of ACA joints caused by different bump geometries, variations in filler size, and differences in bonding pressure [97]. The classifications correspond to Figure 20.31.

time of 15 minutes at hold temperature are shown in Figure 20.35. However, the number of fails is dependent on the definition of the failure. Figure 20.35 shows three interpretations of the same cumulative failure data, based respectively on the different criteria: $>20\%$ of contact resistance increase; >50 mΩ; >100 mΩ. When the criterion was defined at 20% of resistance increase, all joints failed after 2000 cycles. This definition might be too harsh for those joints having initial contact resistances of only several mΩ. The 20% increase means only variations of a few milliohms are allowed. In some cases, the limitation is still within the margin of error of the measurement. Therefore, it is reasonable that the crite-



(a) (b)

FIGURE 20.33. The case on the left is located close the glass fibers (a) and has an excellent resistance (5 mΩ) and reliability, whereas the latter (b) has a high contact resistance (14 mΩ) and poor reliability (49 mΩ after 1000 cycles of TC test) due to its location far from the fibers [99].

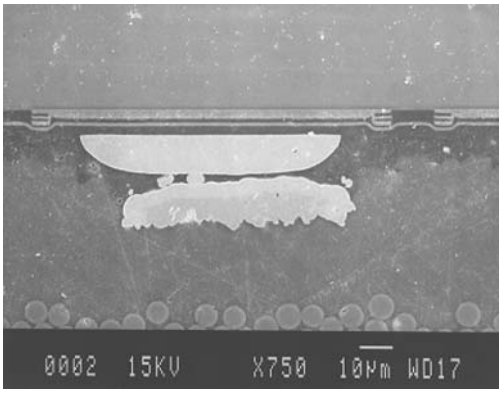


FIGURE 20.34. Pad sinking led to the bad bonding quality [99].

rion be defined according to the production requirements. However, if we define failure as 50 or 100 mΩ, the mean times between failures (MTBF) become 2500 and 3100 cycles, respectively.

In order to predict the real service life under different environmental conditions, a low-cycle fatigue testing machine was used to perform low-cycle fatigue experiments on ACA made joints in both dry and humid environments at different temperatures [80]. The final goal was to predict the real service life using the fatigue life data generated under different plastic strain loads. The plastic strain is a function of the temperature cycling interval, frequency and temperature ramp rate, etc. The daisy chain resistance curves versus the number of cycles are shown in Figure 20.36. A fairly significant decrease of the load level was noticed during the test. The load level ranged for both samples between 1 N and 0.7 N. According to the test data, failure occurred in the humid environment much faster than in the dry environment, which was physically expected. In fact in the humid atmosphere, the joints are subjected to very severe operating conditions (expansion of the adhesive due to moisture uptake, creep of the adhesive under exhaustive conditions).

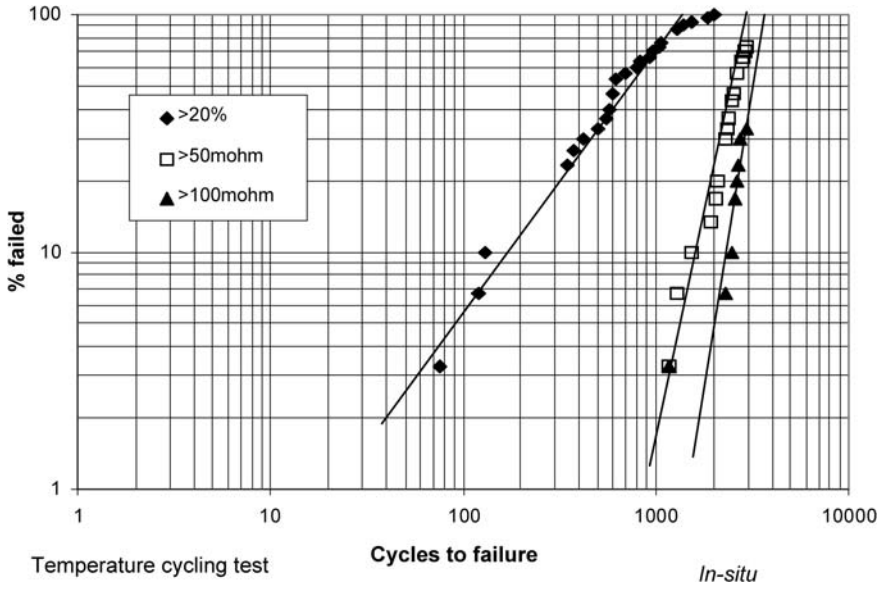


FIGURE 20.35. Cumulative failure of an epoxy-based ACA joint in after a temperature cycling test from −40 to +125°C with a dwell time of 15 min at hold temperatures [100].

There are two features of ACA flip-chip joining technology, which are different from other microelectronic interconnection techniques, such as flip-chip soldering, surface mounting, etc. These are:

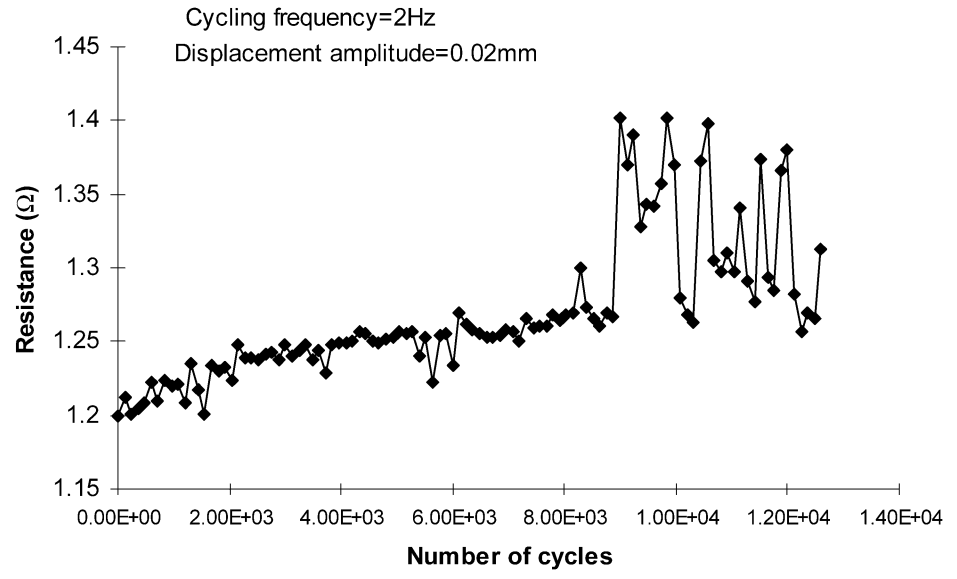
- Application of bonding force.
- Simultaneous bonding of all bumps of a chip.

As the conductivity of ACA joints is directly determined by the mechanical contact between the terminals of chips and the electrodes on chip carriers, the bonding force plays a critical role in the electric performance. High bonding pressure is certainly favorable to an intimate contact, and thus to a low contact resistance. In addition, because the bonding of all bumps of a chip is performed simultaneously, uniform conductivity of all joints in the chip requires the bonding situation of every joint to be completely the same. In other words, every joint in the chip must have:

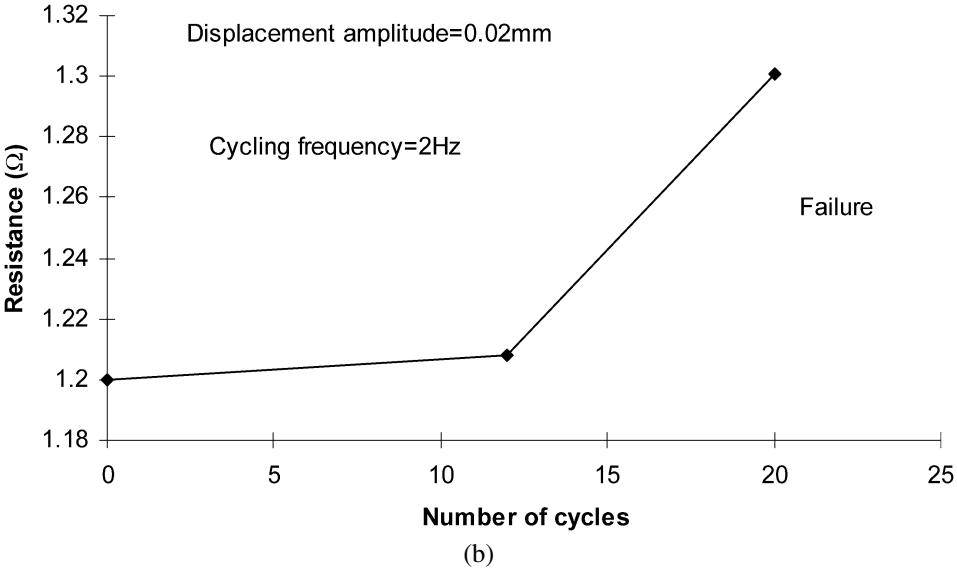
- Same bonding pressure.
- Same number of conducting particles.
- Same particle size.
- Same bump and pad geometry.

In practice, these requirements are hardly ever met. Many factors can affect the bonding situation, including:

- Distribution of bonding pressure on bonding tool.
- Alignment.
- Variation of conductive particle size.
- Distribution of particles in bonding area.
- Pad planarity.



(a)



(b)

FIGURE 20.36. Electrical resistance of a daisy chain versus the number of cycles of low-cycle fatigue testing. (a) Dry environment; (b) 85°C and 85%RH [80].

- Bump planarity.
- Variation of substrate thickness.

Although we understand significant amount of the ACA joining technology, we still need to understand the following issues:

- High frequency behavior, its coupling with semiconductor devices, cross-talk between the particles especially at the range beyond 20 GHz.

- The maximum current carrying capability at high frequency and also after exposure to various environmental tests.
- The planarity effect of the substrate on the ACA joint reliability. This remains one of the most important issues to be understood before we can use ACA in real high volume for low-cost applications.
- Life time prediction models of ACA joints. For this, we need to understand further the effect of UV-degradation of the polymer chain, the corrosive gas effect and high temperature storage effect.

20.7.3. General Comments

As conductive adhesives are made of polymers and metal fillers, light may break the polymer chain. How this works is still unclear as there is no report on this work.

A similar situation is noted as to how various gases affect the polymer chain stability in conductive adhesive applications, although it is clear that polymer chains may be affected and broken by the corrosive gases.

One of the major important questions that need to be addressed is the estimation and prediction of the real service life of a conductive adhesive joint. As conductive adhesives consist of a metal part and an epoxy part, it is a composite material. Therefore, it is unlikely that the acceleration laws used for prediction of the real service lives of pure metals and pure polymers can be directly used for conductive adhesives.

20.8. ENVIRONMENTAL IMPACT

The environmental impact of ECAs has been studied by several research groups [101–103]. Segerberg et al. [102] compared use of conductive adhesive joining with soldering for SMT applications and concluded that it is really dependent on the mining condition of silver when determining the environmental load of the conductive adhesives. If silver is mined in a pure silver mine, then the environmental load is much smaller compared to soldering. On the other hand, if it is mined as a by-product in a copper mine, then conductive adhesive joining technology will have a much larger environmental load index compared to soldering.

Westphal et al. [101] came to the conclusion in their study that conductive adhesives are generally better in terms of environmental loading compared to solder. More work is needed to clarify this topic.

20.9. FURTHER STUDY

Readers are directed to references [104] and [105] for introductory concepts and a more comprehensive coverage respectively. Recent reviews of ACAs include references [106,107]. In addition, many of the points made here are expanded on in the on-line course at www.cpmt.org [108]. This work has been an update of an earlier review [109].

REFERENCES

1. J. Kivilahti (personal communication).
2. J.E. Morris, Conduction mechanisms and microstructure development in isotropic, electrically conductive adhesives, in J. Liu, Ed., Chapter 3 of Conductive Adhesives for Electronics Packaging, Electrochemical Publications, Ltd., UK, 1999, pp. 36–77.

3. R.P. Kusy, Influence of Particle size ratio on the continuity of aggregates, J. Appl. Phys., 48(12), pp. 5301–5305 (1977).
4. L. Li, H. Kim, C. Lizzul, I. Sacolick, and J.E. Morris, Electrical, structure and processing properties of electrically conductive adhesives, IEEE Trans. Compon. Packag. Manuf. Technol., 16(8), pp. 843–851 (1993).
5. Y. Fu, J. Liu, and M. Willander, Conduction modeling of conductive adhesive with bimodal distribution of conducting element, International Journal of Adhesion and Adhesives, (Dec.), pp. 281–286 (1998).
6. L. Li, Ph.D. dissertation, Basics and applied studies of electrically conductive adhesives, State university of New York at Binghamton, 1995.
7. J. Constable, T. Kache, S. Muehle, H. Teichmann, and M. Gaynes, Continuous electrical resistance monitoring, pull strength, and fatigue life of isotropically conductive adhesive joints, IEEE Trans. Compon. Packag. Technol., 22(2), pp. 191–199 (1999).
8. P. Li, E. Cotts, Y. Guo, and G. Lehmann, Viscosity measurements and models of underfill mixtures, Proc. 3rd International Conference on Adhesive Joining and Coating Technology in Electronics Manufacturing, Binghamton, NY (Adhesives'98), Sept., 1998, pp. 328–333.
9. M. Mundlein and J. Nicolics, Modeling of particle arrangement in an isotropically conductive adhesive joint, Proc. 4th International Conference on Polymers and Adhesives in Microelectronics and Photonics, Portland, Oregon, 2004, PP-3.
10. E. Sancaktar and N. Dilsiz, Anisotropic alignment of nickel particles in magnetic field for electrically conductive adhesives application, J. Adhesion Sci. Technol., II(2), pp. 155–166 (1997).
11. T. Inada and C.P. Wong, Fundamental study on adhesive strengh of electrical conductive adhesives (ECAs), Proc. 3rd International Conference on Adhesive Joining and Coating Technology in Electronics Manufacturing, Binghamton NY (Adhesives'98), Sept., 1998, pp. 156–159.
12. J.E. Morris, C. Cook, M. Armann, A. Kleye, and P. Fruehauf, Electrical conduction models for electrically conductive adhesives, Proc. 2nd IEEE International Symposium Polymeric Electronics Packaging (PEP' 99), Gothenburg, Sweden, 1999, pp. 15–25.
13. E. Sancaktar and N. Dilsiz, Thickness dependent conduction behavior of various particles for conductive adhesive applications, Proc. 3rd International Conference on Adhesive Joining and Coating Technology in Electronics Manufacturing, Binghamton, NY (Adhesives'98), Sept., 1998, pp. 90–95.
14. L. Li and J.E. Morris, Electrical conduction models for isotropically conductive adhesives, J. Electronics Manuf., 5(4), pp. 289–298 (1996).
15. L. Li and J.E. Morris, Electrical conduction models for isotropically conductive adhesive joints, IEEE Trans. Compon. Packag. Manuf. Technol. Part A., 20(1), pp. 3–8 (1997).
16. P. McCluskey, J.E. Morris, V. Verneker, P. Kondracki, and D. Finello, Models of electrical conduction in nanoparticle filled polymers, Proc. 3rd International Conference on Adhesive Joining and Coating Technology in Electronics Manufacturing, Binghamton, NY (Adhesives'98), Sept., 1998, pp. 84–89.
17. G.G.W. Mustoe, M. Nakagawa, X. Lin, and N. Iwamoto, Simulation of particle compaction for conductive adhesives using discrete element modeling, Proc. 49th IEEE Electronic Components and Technology Conference, San Diego, 1999, pp. 353–359.
18. M. Mundlein, G. Hanreich, and J. Nicolics, Simulation of the aging behavior of isotropic conductive adhesives, Proc. 2nd International IEEE Confer. Polymers and Adhesives in Microelectronics and Photonics (Polytronic 2002), Zalaegerszeg, Hungary, 2002, pp. 68–72.
19. L. Li and J.E. Morris, Structure and selection models for anisotropic conductive adhesives films, Proc. 1st International Conference on Adhesive Joining and Coating Technology in Electronics Manufacturing (Adhesives'94), Berlin, Germany, 1994.
20. A.O. Ogunjimi, S.H. Mannan, D.C. Whalley, and D.J. Williams, Assembly of planar array components using anisotropic conducting adhesives—A benchmark study, Proc. 2nd International Conference on Adhesive Joining and Coating Technology in Electronics Manufacturing (Adhesives in Electronics, '96), Stockholm, 1996, pp. 270–284.
21. D.J. Williams and D.C. Whalley, The effects of conducting particle distribution on the behavior of anisotropic conducting adhesives: non-uniform conductivity and shorting between connections, Journal of Electronics Manufacturing, 3, pp. 85–94 (1993).
22. R. Herczynski, Distribution function for random distribution of spheres, Nature, 255, pp. 540–541 (1975).
23. D.J. Williams, D.C. Whalley, O.A. Boyle, and A.O. Ogunjimi, Anisotropic conducting adhesives for electronic interconnection, Soldering and Surface Mount Technology, (14), pp. 4–8 (1993).
24. S.H. Mannan, D.J. Williams, D.C. Whalley, and A.O. Ogunjimi, Models to determine guidelines for the anisotropic conducting adhesives joining process, in J. Liu, Ed., Chapter 4 of Conductive Adhesives for Electronics Packaging, Electrochemical Publications Ltd., UK, 1999, pp. 78–98.

25. F. Shi, M. Abdulla, S. Chungpaiboonpatana, K. Okuyama, C. Davidson, and J. Adams, Electrically anisotropic conductive adhesives: a new model for conduction mechanism, Proc. 5th International Symposium Advanced Packaging Materials, Braselton, GA, 1999, pp. 163–168.
26. D. Klosterman, L. Li, and J.E. Morris, Materials characterization, conduction development, and curing effects on reliability of isotropically, IEEE Trans. Compon. Packag. Manuf. Technol. Part A, 21(1), pp. 23–31 (1998).
27. L. Li and J.E. Morris, Modeling cure schedules for electrically conductive adhesives, in J. Liu, Ed., Chapter 5 of Electrically Conductive Adhesives: A Comprehensive Review, Electrochemical Press, UK, 1999, pp. 99–116.
28. J.E. Morris and S. Probsthain, Investigations of plasma cleaning on the reliability of electrically conductive adhesives, Proc. 4th International Conference on Adhesive Joining and Coating Technology in Electronics Manufacturing (Adhesives in Electronics), Espoo, Finland, 2000, pp. 41–45.
29. M.G. Perichaud, J.Y. Deletage, D. Carboni, H. Fremont, Y. Danto, and C. Faure, Thermomechanical behavior of adhesive jointed SMT components, Proc. 3rd International Conference on Adhesive Joining and Coating Technology in Electronics Manufacturing, Binghamton, NY (Adhesives'98), Sept., 1998, pp. 55–61.
30. O. Rusanen and J. Laitinen, Reasons for using lead-free solders rather than isotropically conductive adhesives in mobile phone manufacturing, Proc. 4th International Confer. Polymers and Adhesives in Microelectronics and Photonics, Portland, Oregon, 2004, pp. 2–4.
31. Y. Wei, Ph.D. dissertation, Electronically conductive adhesives: conduction mechanisms, mechanical behavior and durability, Clarkson University, 1995.
32. J.E. Morris and J.H. Das, in J.E. Morris, Ed., Electronics Packaging Forum, Vol. 3, IEEE Press, 1994, pp. 41–71.
33. J.E. Morris and J. Das, Metal diffusion in polymers, IEEE Trans. Compon. Packag. Manuf. Technol.- Part B: Adv. Pkg., 17(4), pp. 620–625 (1994).
34. E. Sencaktar, P. Rajput, and A. Khanolkar, Correlation of silver migration to the pull-out strength of silver wire embedded in an adhesive matrix, Proc. 4th International Confer. Polymers and Adhesives in Microelectronics and Photonics, Portland, Oregon, 2004, RT2-1.
35. K.S. Moon, J. Wu and C.P. Wong, Improved stability of contact resistance of low meting point alloy incorporated isotropically conductive adhesives, IEEE Trans. Compon. Packag. Technol., 26(2), pp. 375–381 (2003).
36. C.P. Wong and Y. Li, Recent advances on Electrical Conductive Adhesives (ECAs), Proc. 4th International Conference Polymers and Adhesives in Microelectronics and Photonics, Portland, Oregon, 2004, PL-1.
37. K. Suzuki, Y. Shirai, N. Mizumura, and M. Konagata, Conductive adhesives containing Ag-Sn alloys as conductive filler, Proc. 4th International Conference Polymers and Adhesives in Microelectronics and Photonics, Portland, Oregon, 2004, MP3-1.
38. P. Smilauer, Thin metal films and percolation theory, Contemp. Phys., 32, pp. 89–102, (1991).
39. G.R. Ruschau, S. Yoshikawa, and R.E. Newnham, Percolation constraints in the use of conductor-filled polymers for interconnects, Proc. 42nd Electronic Components and Technology Conference, Atlanta GA, May 1992, pp. 481–486.
40. J.E. Morris, F. Anderssohn, E. Loos, and J. Liu, Low-tech studies of isotropic electrically conductive adhesives, Proc. 26th International Spring Seminar on Electronics Technology, Stara, Lesna, Slovak Republic, 2003, pp. 90–94.
41. J.E. Morris, F. Anderssohn, S. Kudtarkar, and E. Loos, Reliability studies of an isotropic electrically conductive adhesive, Proc. 1st International IEEE Conference Polymers and Adhesives in Microelectronics and Photonics, 2001, Potsdam, Germany, pp. 61–69.
42. R. Zallen, The Physics of Amorphous Solids, Chapter 4, Wiley, New York, 1983.
43. J.E. Morris, S. Youssof, and X. Feng, Electrically conductive adhesives for pin-through-hole applications, J. Electronics Manuf., 6(3), pp. 219–230 (1996).
44. C.P. Wong, D. Lu, and Q. Tong, Lubricants of silver fillers for conductive adhesive applications, Proc. 3rd International Conference on Adhesives Joining and Coating Technology in Electronics Manufacturing, Binghamton, NY (Adhesives'98), Sept., 1998, pp. 184–192.
45. E. Sancaktar and N. Dilsiz, Pressure dependent conduction behavior of various particles for conductive adhesive applications, Proc. 3rd International Conference on Adhesive Joining and Coating Technology in Electronics Manufacturing, Binghamton, NY (Adhesives'98), Sept., 1998, pp. 334–344.
46. U. Behner, R. Haug, R. Schutz, and H.L. Hartnagel, Characterization of anisotropically conductive adhesive interconnections by 1/f noise measurements, Polymers in Electronics Packaging Conference, Norrkoping, Sweden, 1997, pp. 243–248.

47. L.K.J. Vandamme, M. Perichaud, E. Noguera, Y. Danto, and U. Buehner, 1/f noise as a diagnostic tool to investigate the quality of isotropic conductive adhesive bonds, IEEE Trans. Compon. Packag. Technol., 22(3), pp. 446–454 (1999).
48. Y. Li, K. Moon, H. Li, and C.P. Wong, Conductive improvement of isotropically conductive adhesives, Proc. 6th IEEE CPMT Conference High Density Microsystem Design and Packaging and Component Failure Analysis (HDP'04), Shanghai, 2004, pp. 236–241.
49. J. Miragliotta, R.C. Benson, and T.E. Phillips, Measurements of electrical resistivity and raman scattering from conductive die attach adhesives, Proc. 8th International Symposium Advanced Packaging Materials, Stone Mountain, GA, 2002, pp. 132–138.
50. H. Kim, State University of New York at Binghamton, unpublished data (1992).
51. S. Kottaus, R. Haug, H. Schaefer, and B. Guenther, Investigation of isotropically conductive adhesives filled with aggregates of nano-sized Ag-particles, Proc. 2nd International Conference on Adhesive Joining and Coating Technology in Electronics Manufacturing (Adhesives in Electronics '96) Stockholm, 1996, pp. 14–17.
52. B. Guenther and H. Schaefer, Porous metal powders for conductive adhesives, Proc. 2nd International Conference on Adhesive Joining and Coating Technology in Electronics Manufacturing (Adhesives in Electronics, '96), Stockholm, 1996, pp. 55–59.
53. J. Liu, K. Gustafsson, Z. Lai, and C. Li, Surface characteristics, reliability and failure mechanisms of tin, copper and gold metallisations, Proc. 2nd International Conference on Adhesive Joining and Coating Technology in Electronics Manufacturing (Adhesives in Electronics, '96), Stockholm, 1996, pp. 141–153.
54. L. Li, J.E. Morris, J. Liu, Z. Lai, L. Ljungkrona, and C. Li, Reliability and failure mechanism of isotropically conductive adhesives, Proc. 45th Electronic Components and Technology Conference, Las Vegas, NV, May 1995, pp. 114–120.
55. H. Botter, R.B. Van Der Plas, and A.A. Junai, Factors that influence the electrical contact resistance of isotropic conductive adhesive joints during climate chamber testing, Int. J. Microelec. Pkg., 1(3), pp. 177–186 (1998).
56. D. Lu, C.P. Wong, and Q.K. Tong, Mechanisms underlying the unstable contact resistance of conductive adhesives, Proc. 49th Electronic Components and Technology Conference, San Diego, June 1999, pp. 324–346.
57. D. Lu and C.P. Wong, Conductive adhesives with improved properties, Proc. 2nd International Symposium Polymeric Electronics Packaging (PEP'99), Gothenburg, Sweden, 1999, pp. 1–8.
58. S. Wu, K. Hu, and C.P. Yeh, Contact reliability modeling and material behavior of conductive adhesives under thermomechanical loads, in J. Liu, Ed., Chapter 6 of Conductive Adhesives for Electronics Packaging, Electrochemical Publications Ltd., UK, 1999, pp. 117–150.
59. M. Dernevik, R Sihlbom, Z. Lai, P. Starski, and J. Liu, High frequency measurements and modeling of anisotropic, electrically conductive adhesive flip-chip joint, Advances in Electronics Packaging, 1997, EEP-Vol. 19-1, ASME, 1997, pp. 177–184.
60. R. Sihlbom, M. Dernevik, Z. Lai, P. Starski, and J. Liu, High-frequency measurements and simulation on the Sequential Build-Up Boards (SBU), Proc. 1st International Symposium on Polymeric Electronics Packaging (PEP'97), Norrköping, Sweden, 1997, pp. 131–139.
61. A. Kulkarni and J. Morris, Reliability life time studies on isotropic conductivity adhesives, Proc. 3rd International IEEE Conference Polymers and Adhesives in Microelectronics and Photonics, Montreux, Switzerland, 2003, pp. 333–336.
62. M.J. Yim and K.W. Paik, Design and understanding of Anisotropic Conductive Films (ACF's) for LCD packaging, IEEE Trans. Compon. Packag. Manuf. Technol., Part A, 21(2), pp. 226–234 (1998).
63. C. Oguibe, S. Mannan, D. Whalley, and D. Williams, Conduction mechanisms in anisotropic conducting adhesive assembly, IEEE Trans. Compon. Packag. Manuf. Technol., Part A, 21(2), pp. 235–242 (1998).
64. Y. Fu, Y.L. Wang, X. Wang, J. Liu, Z. Lai, G.L. Chen, and M. Willander, Experimental and theoretical characterization of electrical contact in anisotropically conductive adhesive, IEEE Trans. Compon. Packag. Manuf. Technol., Park B, Advance Packaging, 23(1), pp. 15–21 (2000).
65. R. Sihlbom, M. Dernevik, M. Lindgren, P. Starski, Z. Lai, and J. Liu, High frequency measurements and simulations on wire-bonded modules on the Sequential Build-Up boards (SBU's), IEEE Trans. Compon. Packag. Manuf. Technol., Part A, 21(3), pp. 478–491 (1998).
66. J. Kim, Proc. 2nd International Academic Conference, March 17–19, Atlanta, USA, 2000.
67. Y. Fu, J. Liu, and M. Willander, Electromagnetic wave transmission through lossless electrically conductive adhesive, Journal of Electronics Manufacturing, 9(4), pp. 275–281 (1999).
68. A. Paproth, K.-J. Wolter, T. Herzog, and T. Zerna, Influence of plasma treatment on the improvement of surface energy, Proc. 24th International Spring Seminar Electronics Technology, Romania, 2001, pp. 37–41.

69. T. Herzog, M. Koehler, and K.-J. Wolter, Improvement of the adhesion of new memory packages by surface engineering, Proc. 2004 Electronic Components and Technology Conference, Las Vegas, 2004, Vol. 1, pp. 1136–1141.
70. S. Liong, C.P. Wong, and W.F. Burgoyne, Jr., Adhesion improvement of thermoplastic isotropically conductive adhesives, Proc. 8th International Symposium on Advanced Packaging Materials, Braselton, GA, 2002, pp. 260–270.
71. L.L.W. Chow, J. Li, and M.M.F. Yuen, Development of low temperature processing thermoplastic intrinsically conductive polymer, Proc. 8th International Symposium on Advanced Packaging Materials, Braselton, GA, 2002, pp. 127–131.
72. R. Luchs, Application of electrically conductive adhesives in SMT, Proc. 2nd International Conference on Adhesive Joining and Coating Technology in Electronics Manufacturing (Adhesives in Electronics, '96), Stockholm, 1996, pp. 76–83.
73. M. Zwolinski, J. Hickman, H. Rubin, Y. Zaks, S. McCarthy, T. Hanlon, P. Arrowsmith, A. Chaudhuri, R. Hermansen, S. Lau, and D. Napp, Electrically conductive adhesives for surface mount solder replacement, Proc. 2nd International Conference on Adhesive Joining and Coating Technology in Electronics Manufacturing (Adhesives in Electronics, '96), Stockholm, 1996, pp. 333–340.
74. Q. Tong, S. Vona, R. Kuder, and D. Shenfield, Recent advances in surface mount conductive adhesives, Proc. 3rd International Conference on Adhesive Joining and Coating Technology in Electronics Manufacturing, Binghamton, NY (Adhesives'98), Sept., 1998, pp. 272–277.
75. S. Vona, Q. Tong, R. Kuder, and D. Shenfield, Surface mount conductive adhesives with superior impact resistance, Proc. 4th International Symposium and Exhibition on Advanced Packaging Materials, Processes, Properties and Interfaces, Braselton, GA, 1998, pp. 261–267.
76. S. Luo and C.P. Wong, Thermo-mechanical properties of epoxy formulations with low glass transition temperatures, Proc. 8th International Symposium on Advanced Packaging Materials, Braselton, GA, 2002, pp. 226–231.
77. S.A. Kudtarkar and J.E. Morris, Reliability studies of isotropic conductive adhesives: drop test for thermally cycled and 85/85 tested samples, Proc. 8th International Symposium on Advanced Packaging Materials, Braselton, GA, 2002, pp. 144–150.
78. Y.L. Wang, G.L. Chen, J. Liu, and Z. Lai, The contact characterization of conductive particles in anisotropically conductive adhesive using FEM, Proc. 2nd IEEE International Symposium on Polymeric Electronics Packaging, Göteborg, Sweden, 1999, pp. 199–206.
79. X. Wang, Y.L. Wang, G.L. Chen, J. Liu, and Z. Lai, Quantitative estimate of the characteristics of conductive particles in ACA by using nano indenter, IEEE Trans. Compon. Packag. Manuf. Technol. Part A, 21(2), pp. 248–251 (1998).
80. A. Zribi, K. Persson, Z. Lai, J. Liu, Y. Kang, S. Yu, and M. Willander, Effect of the bump height on the reliability of ACA made joints for flip-chip electronic packaging, 2nd International ATW on Flip-Chip Technology, Braselton, GA, 1998.
81. Z. Lai, R.Y. Lai, K. Persson, and J. Liu, Effect of bump height on the reliability of ACA flip-chip joining with FR4 rigid and polyimide flexible substrate, Journal of Electronics Manufacturing, 8(3&4), pp. 217–224 (1998).
82. C.M.L. Wu, J. Liu, and N.H. Yeung, Reliability of ACF in flip-chip with various bump height, Soldering and Surface Mount Technology, 13/1, pp. 25–30 (2001).
83. T. Sugiyama, The latest bare chip bonding by ACF, Sony Chemicals, seminar documentation at Chalmers University of Technology, 2000.
84. M. Törnvall, Proceedings of a Swedish National Seminar on Environmentally Compatible Materials Research for Electronics Packaging, IVF.
85. J. Liu and R. Rörgren, Joining of displays using thermo-setting anisotropically conductive adhesive joints, Journal of Electronics Manufacturing, 3, pp. 205–214 (1993).
86. A. Sihlbom, R. Sihlbom, and J. Liu, Thermal characterization of electrically conductive adhesive flip-chip joints, Proc. of the IEEE/CPMT Electronic Packaging Technology Conference, Dec. 8–10, Singapore, 1998, pp. 251–257.
87. A. Bauer and T. Gesang, Electrically conductive joints using Non-Conductive Adhesives (NCAs) in surface mount applications, in J. Liu, Ed., Chapter 12 of Conductive Adhesives for Electronics Packaging, Electrochemical Publications Ltd., UK, 1999, pp. 313–341.
88. C. Khoo, J. Liu, M. Ågren, and T. Hjertberg, Influence of curing on the electrical and mechanical reliability of conductive adhesive joints, Proc. 1996 IEPS Conference, Austin, Texas, 1996, pp. 483–501.
89. J. Liu, K. Gustafsson, Z. Lai, and C. Li, Surface characteristics, reliability, and failure mechanisms of tin/lead, copper, and gold metallizations, IEEE Trans. Compon. Packag. Manuf. Technol. Part A, 20(1), pp. 21–30 (1997).

90. J. Liu, P. Lundström, K. Gustafsson, and Z. Lai, Conductive adhesive joint reliability under full-cure conditions, EEP-Vol. 19-1, Advances in Electronics Packaging, Vol. 1, ASME 1997, pp. 193–199.
91. C. Khoo and J. Liu, Moisture sorption in some popular conductive adhesives, Circuit World, 22(4), pp. 9–15 (1996).
92. Q.K. Tong, D. Markley, G. Fredrickson, R. Kuder, and D. Lu, Conductive adhesives with stable contact resistance and superior impact performance, Proc. 49th Electronic Components and Technology Conference San Diego, 1999, pp. 347–352.
93. D. Lu and C.P. Wong, Development of conductive adhesives for solder replacement, IEEE Trans. Compon. Packag. Technol., 23(4), pp. 620–626 (2000).
94. H. Takezawa, T. Mitani, T. Kitae, H. Sogo, S. Kobayashi, and Y. Bessho, Effects of zinc on the reliability of conductive adhesives, Proc. 8th International Symposium on Advanced Packaging Materials, Braselton, GA, 2002, pp. 139–143.
95. O. Rusanen, Ph.D. dissertation, Adhesives in micromechanical sensor packaging, University of Oulu, VTT Publication 407, 2000.
96. O. Rusanen, Modeling of ICA creep properties, Proc. 4th International Conference on Adhesive Joining and Coating Technology in Electronics Manufacturing (Adhesives in Electronics), Espoo, Finland, 2000, pp. 194–198.
97. Z. Lai and J. Liu, Anisotropically conductive adhesive flip-chip bonding on rigid and flexible printed circuit substrates, IEEE Trans. Compon. Packag. Manuf. Technol., Part B: Advanced Packaging, 19(3), pp. 644–660 (1996).
98. Z. Lai and J. Liu, The effects of bonding force on ACA flip-chip reliability, IMAPS ATW on Flip-Chip Technology, Braselton, 1999.
99. J. Liu, A. Tolvgård, J. Malmodin and Z. Lai, A reliable and environmentally friendly packaging technology—flip-chip joining using anisotropically conductive adhesive, IEEE Trans. Compon. Packag. Technol., 22(2), pp. 186–190 (1999).
100. J. Liu and Z. Lai, Reliability of ACA flip-chip joints on FR-4 substrate, InterPACK'99, International, InterSociety, Electronic Packaging Technical/Business Conference and Exhibition, June 13–17, Hawaii, USA, Maui, 1999, pp. 1691–1697.
101. H. Westphal, Health and environmental aspects of conductive adhesives-the use of lead-based alloys compared with adhesives, in J. Liu, Ed., Chapter 18 of Conductive Adhesives for Electronic Packaging, Electrochemical Publications Ltd., UK, 1999, pp. 415–424.
102. T. Segerberg, Life Cycle Analysis—A comparison between conductive adhesives and lead containing solder for surface mount application, IVF report, 1997.
103. A. Tolvgård, J. Malmodin, J. Liu, and Z. Lai, A reliable and environmentally friendly packaging technology—flip chip joining using anisotropically conductive adhesive, Proc. 3rd International Conference on Adhesive Joining and Coating Technology in electronics Manufacturing, Binghamton, NY, 1998, pp. 19–26.
104. L. Li and J.E. Morris, An introduction to electrically conductive adhesives, Int. J. Microelectronic Packaging, 1(3), pp. 159–175 (1998).
105. J. Liu, Ed., Conductive Adhesives for Electronics Packaging, Electrochemical Publications Ltd., Port Erin, Isle of Man, UK, 1999, ISBN No. 0901150371.
106. G. Dou, D.C. Whalley, and C. Liu, Electrical conductive characteristics of ACA bonding: a review of the literature, current challenges and future prospects, Proc. 6th IEEE CPMT Confer. High Density Microsystem Design and Packaging and Component Failure Analysis (HDP'04), Shanghai, 2004, pp. 264–276.
107. J. Liu and Z. Mo, Reliability of interconnects with conductive adhesives, in S. Dongkai, Ed., Lead-free Solder Interconnect Reliability, ASM, Materials Park, OH, 2005, pp. 249–276.
108. J.E. Morris and J. Liu, An Internet course on conductive adhesives for electronics packaging, Proc. 50th Electronic Components and Technology Conference, Las Vegas, May, 2000, pp. 1016–1020.
109. J. Liu and J.E. Morris, State of the art in electrically conductive adhesives, Workshop Polymeric Materials for Microelectronics and Photonics Applications, EEP-Vol. 27, ASME, 1999, pp. 259–281.

21

Electrically Conductive Adhesives

Johann Nicolics and Martin Mündlein

Institute of Sensor and Actuator Systems, Gusshausstraße 27-29, A-1040 Wien, Austria

Abstract

Electrically conductive adhesives are being used in electronic packaging for several decades. A brief review of the dynamic development of conductive adhesives under the influence of the miniaturization, the adaptation of environmental friendly manufacturing processes is presented.

With respect to the importance of isotropically conductive adhesives (ICA), a new contact model to analyze the principle influences of e.g., particle size, particle geometry, and filler content on the percolation threshold is introduced. With this model the arrangement of the particles within a contact is calculated by considering different types of forces (elastic, friction, adhesion, and inertia). Taking into account the electrical properties of the filler particles, the electrical contact behavior including its changes due to aging is investigated.

Finally, typical applications of isotropically conductive adhesives are presented. One example shows how the thermal requirements for attaching a GaAs heterojunction power transistor can be fulfilled using an adhesive with an extremely high filler content (thermal conductivity: >60 W/mK). In another case it is demonstrated how extreme thermomechanical requirements resulting from a thermal expansion mismatch of parts of a sealed IR sensor housing can be corresponded using an adhesive with a comparatively low glass transition temperature. A further example shows a packaging concept of a miniaturized, biocompatible multichip module. For mounting both, narrowly spaced SMDs and bare chips, an isotropically conductive adhesive has been applied.

21.1. INTRODUCTION AND HISTORICAL BACKGROUND

The industrial application of electrically conductive adhesives (ECA) for interconnecting and mounting electronic components on circuit carriers is comparatively new. However, a series of inventions related to various mixtures of conductive and non-conductive substances were made already in the first half of the last century. An early patent of an electrically conductive adhesive describes electromechanical applications already in 1926 [1]. In 1933 Hans Schuhmann claims a *conducting varnish* (consisting of a mixture of about 50% oil varnish, about 40% lithopone, and 10% of soot) which can be applied for shielding purposes in the radio industry [2]. Only two years later it turned out that conducting layers consisting of a binding agent mixed with a finely divided conductor lose their conductivity during operation or do not at all posses this conductivity even if applied in the form in

which they are on the market [3]. In 1944 H.J. Loftis proposed a *molded electrically conductive body* consisting of Bakelite as the resin filled with graphite or metal powder and an additive of an alkaline earth metal salt providing hygroscopic properties to preserve a lubricating effect for the use as commutator brush [4]. An early description of an *electrically conductive adhesive* more similar to the modern ECA can be found in the patent of N.H. Collings et al. in 1948 on the base of a self-setting adhesive and, e.g., finely divided silver powder [5].

The first use of ECAs in electronic technology is known only in 1956: in the U.S. patent [6] *an electrically conducting cement comprising a thermosetting binding medium* is described *for fixing a semi-conducting crystal* (germanium) *to a metal base or holder*. Thereafter, the number of publications reporting experimental and theoretical investigations of electrically conductive adhesives increased continuously. Twenty years later C. Mitchel and H. Berg summarized the state-of-the-art in 1976. They reported that conductive adhesives due to their increasing reliability become more and more acceptable for many different applications. Also the most significant advantages compared to other interconnection techniques, such as low processing temperature and the possibility to use a large variety of low-cost surface fillers are already mentioned [7]. Especially for mounting chips with larger size on substrates with different coefficient of thermal expansion adhesives are superior to eutectic solder alloys due to their significantly higher flexibility. Conductive adhesive bonding started to become popular with the fabrication of plastic-foil keyboards and the use of flexible substrate materials [8].

More recently, in 2000, Alan J. Heeger, Alan G. MacDiarmid and Hideki Shirakawa are awarded the Nobel Prize in Chemistry for their discovery of electrically-conducting polymers whereby initial experiments with silvery shining films of polyacetylene lead back to the beginning of the 1970s [9].

Along with the development of liquid crystal displays (LCDs) in the 80s and their continuing miniaturization a special kind of ECA—the anisotropic conductive adhesive (ACA) has been invented for components with high-lead count and fine pitch interconnections [10,11]. ACAs allow to produce an excellent low-ohmic contact only perpendicular to the mating metalization planes whereby one-particle contact bridges between these metalizations are formed. However, due to their very low loading of conductive filler particles, ACAs remain insulating in the lateral direction. As such, no high-precision printing process is needed. In the meanwhile, ACAs are successfully applied especially for high-resolution displays in portable devices such as notebooks and PDAs (Personal Digital Assistant). Although, the capability of ACAs for comparatively high current load with even up to several Amperes per square millimeter were investigated [12], due to their low filler content their main application fields will probably remain the interconnection of components with low current consumption.

Whereas ACAs have a volume resistivity like an insulator, another type of ECA with a significantly higher degree of conductive filler has a low volume resistivity, comparable with that of conductors (Figure 21.1). They are called isotropic conductive adhesives (ICA, [13]). The goal of this chapter is to develop a deeper understanding of the contact formation process and the influence of application related parameters on the electrical contact resistance of ICAs.

The ICA technology became more interesting in the last years due to several reasons. One of the major driving forces is the increasing packaging density allowing to develop a huge manifold of complex but nevertheless light portable electronic equipments like e.g., cell phones, digital cameras, and notebooks. This development was mainly based on three key factors:

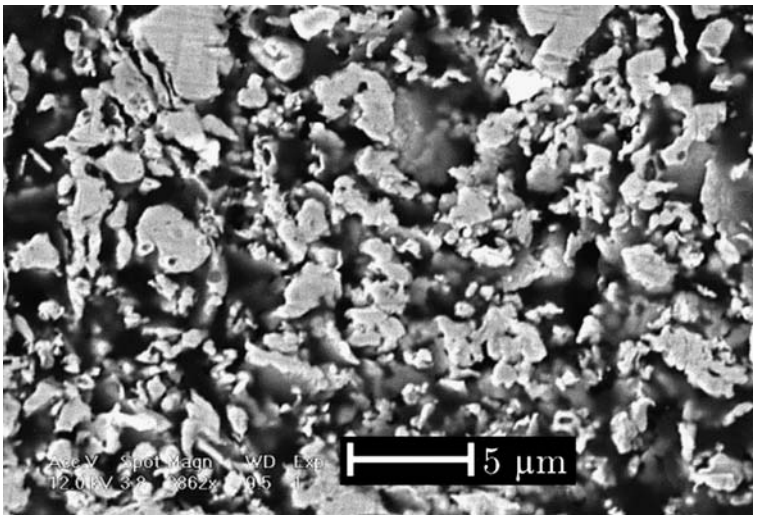


FIGURE 21.1. SEM cross-sectional view of an ICA.

– an inconceivable fineness of structures in the semiconductor level coming close to 0.1 µm and allowing a high integration level,
– a decrease of package size coming down from small outlined (SO) packages over quad flat pack (QFPs), ball-grid arrays (BGAs), and chip-scale packages (CSPs) to direct or bare chip attach techniques like flip-chips (FCs), and
– new printed circuit board manufacturing (PCB) techniques like microvia multilayer technology with plasma or laser structured vias with even less than 0.1 mm diameter allowing to realize a three-dimensional wiring net with extremely high density.

Although, soldering technology has matured through many decades, with the mentioned advances in microelectronics technology a limitation of solder paste become more and more evident in its use for fine pitch components [14]. With current solder paste technology very fine pitch interconnection become hard to handle due to soldering defects such as bridging and solder balling [15–17]. This means to connect the components on the circuit boards under these boundary conditions improved or new interconnection technologies are needed. One of these new techniques could be conductive adhesive bonding. Among these technical demands new legislative reforms make an adaptation of the existing technologies necessary. From the 1st July 2006 on member states of the European Union have to ensure that new electrical and electronic equipment does not contain lead, mercury, cadmium and other named materials. This is regulated in the directive 2002/95/EC of the European Parliament and of the Council on the restriction of the use of certain hazardous substances in electrical and electronic equipment [18]. Under these restriction the prohibition of lead is most serious for the packaging technology because the widely applied solder technology commonly uses solder pastes consisting of a tin lead alloy.

Due to their composition conductive adhesives are innately lead free. Additionally, they allow to avoid flux residues without cleaning. They need a low processing temperature but allow operation at high temperatures. Adhesives are preferable whenever high elasticity is needed to compensate for thermal expansion mismatch between substrate and component body [19–22]. However, compared to soldering, electrically conductive adhesive technology is still in its infancy [23,24]. Two critical issues of conductive adhesives for

surface mount applications are the contact resistance shift during lifetime and a low impact performance [25]. As a long-term effect, the contact resistance of ECAs on ignoble metal finishes can increase dramatically during exposure to elevated temperature and humidity. The low impact performance is observed by separation of components from the board due to impact of significant shock during manufacturing and lifetime of the product. The ability of ECAs to withstand this stress can be determined with the standardized drop test [26]. Another drawback of one-component adhesives compared to soldering is also frequently cited: a lower productivity due to the long curing times. Results obtained by applying a new curing method which allows to reduce the duration of exposition to temperature to an extent comparable to that of soldering processes are also presented.

21.2. CONTACT FORMATION

Let us consider an insulating polymer mixed with a specific amount of conductive particles between a pair of parallel electrodes representing the mating pads of a contact to be formed. It is obvious that these mixture keeps non conductive as long as the particle content is small enough (below its percolation) and the particles are smaller than the gap between the electrodes. If the amount of particles is increased clusters of particles being in contact with each other are formed, whereby the mean size of these clusters increases with the growing amount of particles. From the moment when the first cluster touches both electrodes an electrical connection is formed. If the amount of filler is further increased additional clusters merge successively to one big cluster which leads to an improvement of the electric contact between the two mating pads (Figure 21.2).

The amount of filler particles needed to get the particle-adhesive mixture electrically conductive depends on microscopic geometrical parameters like shape, distribution, and orientation of the particles. The principle influences of these parameters can be analyzed using the percolation theory.

21.2.1. Percolation and Critical Filler Content

The percolation theory is a branch of probability theory dealing with the properties of randomly distributed media in general [27]. One main idea of the percolation theory

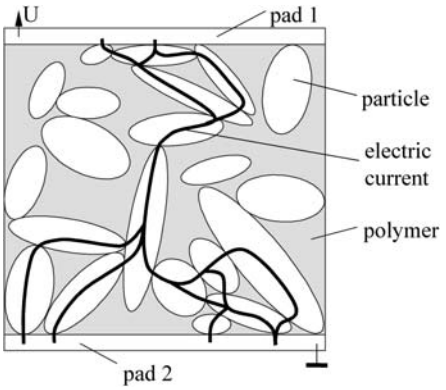


FIGURE 21.2. Schematic cross section of an isotropically conductive joint.

is to predict the macroscopic behavior of a randomly distributed composite material in dependence on its microscopic properties which are expressed purely by statistical data. In case of an isotropically conductive adhesive this composite material is the mixture of the non-conductive polymer and the conductive filler particles. In the simplest case (infinite conductivity of particles and infinite resistivity of the filler) the macroscopic behavior of the joint is described by the probability for a connection path between the two mating pads through the filler particles. Microscopic properties are amount, orientation, shape, size or size distribution of the filler particles. The percolation theory shows that there exists a distinct threshold particle content in excess to which the probability for the formation an *infinite* cluster of particles and therewith for an electrical conductance is bigger than zero and which rapidly increases with rising particle content. This threshold is called the percolation threshold [28].

The percolation theory is valid for infinite systems which means that the particle size is negligible with respect to the distance between the electrodes. Hence, this theory is unable to consider the influence of the particle size on the relation between filler content and percolation probability. However, how important the parameter "particle size" is can be seen by considering an ACA forming a contact with a filler content much below any percolation threshold. In many practical applications of ICAs the relation between the path length (distance between the mating pads in perpendicular direction) and the particle size is only one or two orders of magnitude which causes deviations between ideal (*infinite*) systems as considered by the percolation theory and real ones. For this reason it is important to distinguish between infinite and finite systems. The percolation threshold in an *infinite* system would correspond to the so called *critical filler content* in a real joint. In real (= *finite*) systems only clusters with a limited number of particles are necessary for an electrical connection between the mating borders. The smaller a contract system with respect to the particle size the mellower is the transition between the non conductive and the conductive state in the percolation curve or more precisely: in the conduction probability-versus-filler content function (a comparison with the ideal percolation curve can be seen in Figure 21.5). Nevertheless, here the terminology *percolation curve* is used defining the *percolation threshold* as the point on which the probability of an interconnection is 50% for real systems. (For definitions, symbols and terminology used in this chapter refer to the Section Notations and Definitions.)

21.2.2. ICA Contact Model

Using the percolation theory, it is possible to show principle influences on the properties of isotropically conductive adhesives depending on the geometric properties of the particles. But in order to get a deeper understanding of the processes inside a conductive adhesive joint additional factors have to be considered. Such factors are the volume electrical conductance of the filler particles, the contact resistance between the particles, and the particles and the pads, the influence of surface films on the contact resistance, the mechanical force between the particles, and the effects of oxidation and galvanic corrosion on the electric resistance of the joint [29].

For this purpose, a two-dimensional (2D) model of a complete joint including a numeric simulation was developed which is capable of calculating the particle alignment and density distribution as well as the voltage and current distribution inside the model area. On this base the total DC-resistance of the modeled joint was calculated. The model is based on the following considerations and assumptions [30]:

- the modeled system is two-dimensional,
- the considered area is rectangular,
- the pads are on the top and on the bottom borders of the model area,
- the conductivity of the polymer is neglected,
- the particles have elliptic shape with arbitrary aspect ratio (length ratio of major to minor axes),
- the particles may initially have an arbitrary orientation and can overlap whereby the condition for the final orientation and position of the particles is found by a simulation of the particles' alignment as described below by the assumption that the size of the overlapping area is directly related to a repulsion force between the particles.

A 2D simulation of a percolation problem shows a systematic difference to the three-dimensional (3D) problem. In a 2D percolation the threshold is shifted to higher filler contents. In the case of an ICA this can be observed by an increase of the contact resistance when the thickness of an adhesive layer is reduced [31–33]. Despite of this fundamental difference between the 2D and 3D model the percolation theory shows that the principle behavior of a random particle network obeys the same rules [34]. This means the absolute values derived from the 2D do not directly correspond to the 3D case but show the analog dependencies on the particle parameters like shape, orientation or distribution. However, the computational effort can be reduced drastically for a 2D numerical simulation and the computing capacity saved in this way can be used to investigate a higher number of different dependencies. For this reason, the presented simulation is restricted to the 2D case and can consider the change of size of the contact areas at the mating surfaces of the particles in dependence of the contact force. Moreover, the simulation considers the dependency of the contact resistance on particle shape, particle orientation, etc.

21.2.2.1. Simulation of the Particles' Alignment The simulation is capable of establishing the geometrical alignment of the elliptical particles based on mechanical effects like inertia, conservation of momentum, Hooke's law, and friction. This method is known as discrete element modeling (DEM) which is an explicit numerical method where individual separate elements (particles) react with their neighbors by their contact areas through friction and adhesion (DEM was first applied by Peter Cundall in 1971 to geotechnical and granular flow problems [35]).

For each time step an overlapping between the particles at their contact points is assumed in order to consider a pressure-dependent deformation of the particles. Figure 21.3 illustrates the forces of one particle in a non-equilibrium calculation step. The considered particle is in contact with another one and the upper pad. Elastic forces F_{e1} and F_{e2} are defined as functions of the size of the respective overlapping area whereby the direction of these forces is perpendicular to the chord between the intersection points S_{1a}, S_{1b}, S_{2a}, S_{2b}.

In order to define an electrical contact resistance between neighboring particles a contact force between them has to be present. In an ICA these contact forces are assumed to result from shrinking of the polymer during the curing process and have to be considered in order to simulate the DC-resistance of an ICA joint [36,37]. For this reason, an attracting force F_a between intersecting particles resulting from the compaction of the particle network during the curing of the polymer is introduced. The direction of this attracting force is antiparallel to the respective elastic force. The norm is proportional to the length of the chord between the intersecting points.

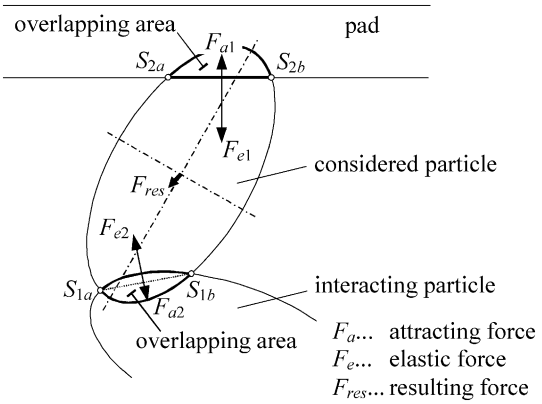


FIGURE 21.3. Schematic illustration of the overlapping area between adjacent particles, and mechanical forces leading to a rearrangement of the particles.

For each particle these attracting and repulsing elastic forces at different contact points can be added-up to one resulting force F_{Res} which causes a rearrangement. This rearrangement motion of the particles is controlled by inertia, a damping force and a friction force. Similarly, the resulting torsional moments are calculated for each particle. The distribution of the resulting rearrangement forces and moments are successively optimized in a looped calculation procedure leading to the final particle arrangement. Depending on the relation between the attracting force and the elastic force a defined overlapping between the particles remains at the equilibrium state.

21.2.2.2. Simulation of the Total Joint Resistance After the particle alignment is being calculated the electric resistance of the joint is simulated. This is done in several steps. The first step is to separate each single particle from the base model, whereby the voltage distribution inside the respective particle is calculated using a finite difference method [38]. As boundary condition, the potentials of the contact areas to the neighboring particles are defined. When the voltage distribution is calculated for $n-1$ different sets of boundary conditions it is possible to calculate an admittance matrix describing the particle as $(n+1)$-pole, where n denotes the number of neighbored particles. (The $(n+1)$ pole as an additional pole located in the particle center is introduced for simplifying indexing and set-up of the equation system.) This idea is depicted in Figures 21.4(a) and (b).

After an admittance matrix is being calculated for every particle, the resulting $(n+1)$-poles are connected to one network which contains the information on the particle distribution of the entire joint including the pad arrangement. The effort for calculating the contact resistance to the required extent is reduced by excluding particles which do not contribute to the current transportation from the network. In order to model more realistic contact properties a transition resistance between the particles is taken into account. For this purpose, additional resistors (R_{pp} and R_{ppad}) between the respective connections of the $(n+1)$-poles are introduced [Figure 21.4(c)]. The value of each individual transition resistance depends on the contact force which controls the size of the overlapping area. As a measure for the size of the overlapping area and therewith for the contact resistance between the two considered particles the length of the intersection line (e.g., $w_s = \overline{S_{1a}S_{1b}}$ in Figure 21.3) is used. The interface conductivity between two particles is then obtained as product of a normalized specific conductivity λ_{pp} and the length of the intersection line w_s

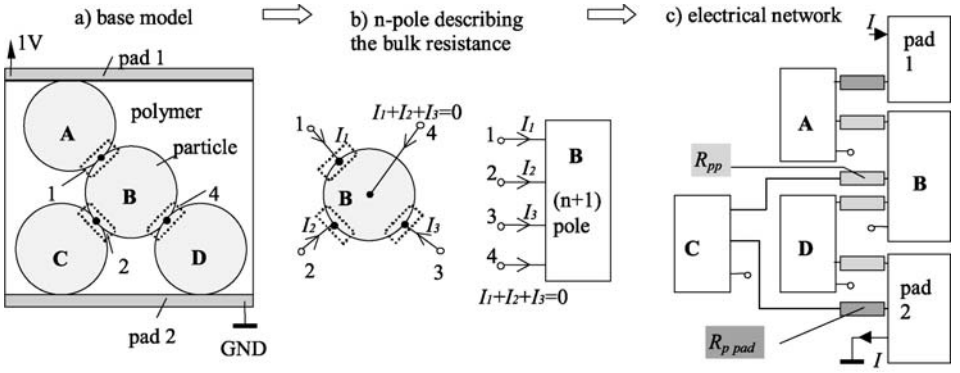


FIGURE 21.4. Basic principle for calculating the electric resistance.

of the involved particles. In turn, the normalized interface resistance R_{pp} between two particles or $R_{p\,pad}$ between a particle and a pad are calculated according to following relation, respectively:

$$R_{pp} = \frac{1}{\lambda_{pp} \cdot w_s} \quad \text{or} \quad R_{p\,pad} = \frac{1}{\lambda_{p\,pad} \cdot w_s}. \tag{21.1}$$

As final step the equations of this network are solved. Thereafter, the potential at each contact point is found and the total joint resistance is calculated.

21.2.3. Results

21.2.3.1. Percolation Behavior The established model allows us to analyze different fundamental properties of conductive adhesive joints. In the first step, only the percolation behavior is investigated. For this purpose, only the particle alignment is calculated and the probability of an interconnection p_p from one pad to the other is determined as a function of the filler content ϕ in % of area. For all simulations, a quadratic model area with a normalized edge length of 1 is chosen. Figure 21.5 shows the dependence of p_p for circular particles with different diameters d. It can be seen that the slope of the curve decreases for larger diameters in accordance with the theory. In the literature for the case of *hard-core* circles (neighboring particles only touch at one point), a critical filler content ϕ_c lies between 0.45 and 0.55 [34].

In Figure 21.5 the results of the simulation with the percolation theory is compared. At an interconnection probability of $p_p = p_c \equiv 0.5$ (which is defined as the percolation threshold for the finite system) the simulation provides filler content values of ϕ_c between 0.55 and 0.57. For the smaller particles the higher values are valid. The smaller the particles are, the more the infinite system is approached (which is described exactly by the percolation probability).

A further fact in this model is the deformability of the particles which corresponds to a combination of the *hard-core* and the *soft-core* model where the particles may touch along a line (not just a point) due to arbitrary overlapping. In the literature one can find the value $\phi_c = 0.68$ for the *soft-core* circles system [39].

In an ICA joint the electrical conductance is provided by the conductive particles, and the mechanical strength derives only from the polymer. This means: to preserve the

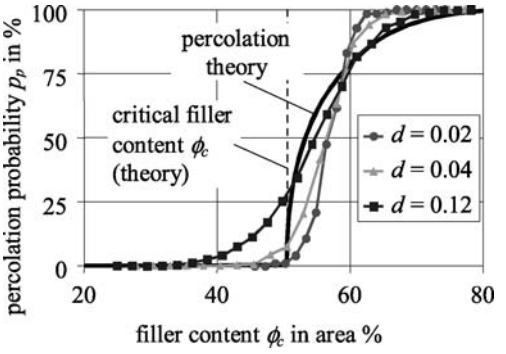


FIGURE 21.5. Comparison of percolation probability and probability of interconnection with circular particles as a function of the filler content.

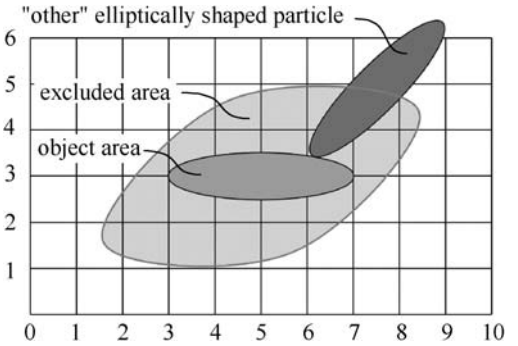


FIGURE 21.6. Excluded area for ellipses with the same shape but different orientation.

mechanical stability of the joint the particle content has to be as small as possible. This can be obtained with highly elongated particles like flakes, sticks, or tubes. However, for non-circular particles, a given filler content will lead to a different interconnection probability. In order to understand the influence of the particle shape on the filler content area which is necessary to obtain e.g., the 50% threshold to form an interconnection it is helpful to start at the definition of the so called *excluded volume* which is reduced to the *excluded area* in the two-dimensional case. The excluded area A_{ex} of an object (particle) is defined as the object area itself plus the area around this object into which the center of another similar object is not allowed to enter if overlapping of these two objects is to be avoided [34]. For example, for circles with the radius r the excluded area A_{ex} is $4r^2\pi$. For two ellipses of the same shape but different (constant) orientation the excluded area is shown in Figure 21.6. From a heuristic consideration it becomes obvious that the percolation threshold is lowered if the excluded volume is increased.

The excluded area is increased when circular particles are stretched to ellipses. The bigger the aspect ratio at a constant particle area the bigger is the excluded area. This tendency is shown in Figure 21.7 where three cases of excluded areas are formed with two ellipses of the same shape and size (relation between major and minor axis $X = 16$), respectively, but different mutual orientation ($\delta = 0°$, $\delta = 45°$, and $\delta = 90°$). One can see that the excluded area increases significantly with rising angle δ. The maximum possible value of the excluded area also depends on the shape of the particles. Figure 21.8 depicts

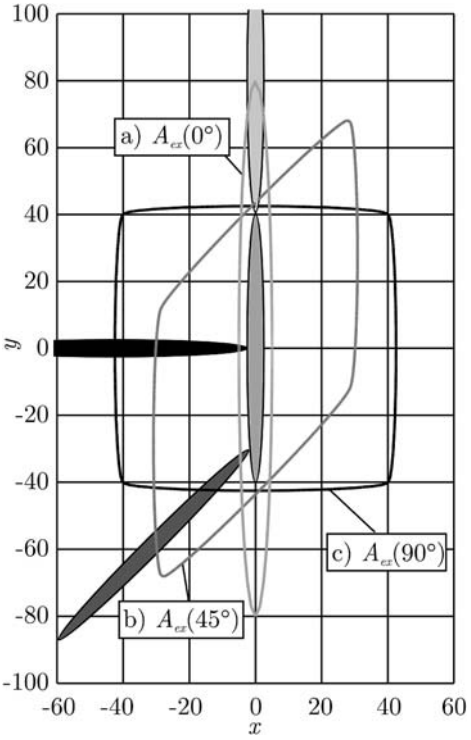


FIGURE 21.7. Comparison of three cases of excluded area for equal ellipses but different orientation: (a) $\delta = 0°$, $A_{ex} = 4 \cdot A$, (b) $\delta = 45°$, $A_{ex} = 16.5 \cdot A$, (c) $\delta = 90°$, $A_{ex} = 22.5 \cdot A$; relation between length of major and minor axis of ellipses $X = 16$.

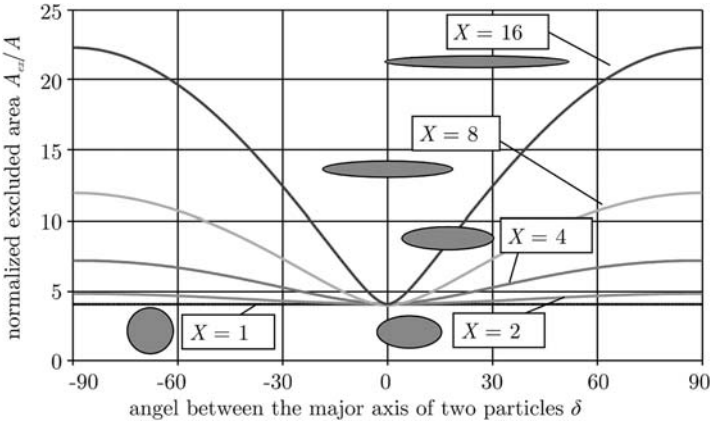


FIGURE 21.8. Normalized excluded area versus angle δ between elliptic particles with various aspect ratios ($1 \leq X \leq 16$).

the normalized excluded area versus angle δ between elliptic particles with various aspect ratios ($1 \leq X \leq 16$). It should be noted that the dependence of the excluded area on the orientation of the particles with respect to each other becomes stronger the more oblong the particles are.

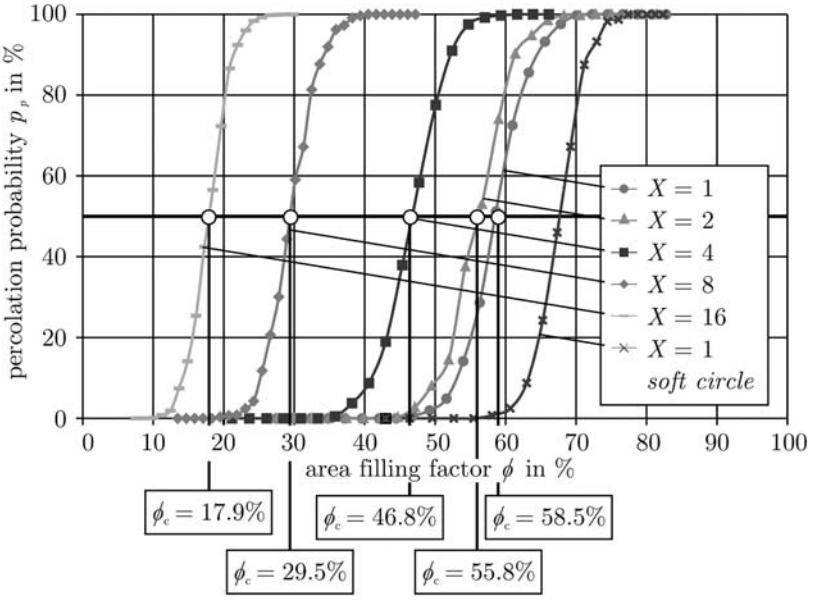


FIGURE 21.9. Probability of interconnection as a function of the filler content for elliptic particles with the same area but different aspect ratio.

The influence of the excluded area on the percolation behavior can be recognized in Figure 21.9 where the percolation probability p_p for elliptical particles with different aspect ratio is depicted as function of the filler content ϕ. It can clearly be seen the decreasing critical filler content (the respective filler content at a probability level of 50%) with the increasing aspect ratio of the particles. All particles in Figure 21.9 have the same area as a circular particle with the diameter $d = 0.04$.

A further important parameter is the mean number B of neighboring particles. With increasing filler content the number of particles being in contact with each other grows. On reaching the percolation threshold the following relation is valid:

$$B_c = \phi_c \frac{\langle A_{ex} \rangle}{\langle A \rangle}, \tag{21.2}$$

where B_c is the critical mean number of neighboring particles, $\langle A \rangle$ is the mean particle area, and $\langle A_{ex} \rangle$ is the mean excluded area. For circular particles B_c varies in the range between 1.8 and 2.2 [34]. One can expect that this range should also be valid for elliptically shaped particles. However, the values of ϕ_c obtained from the simulation described here are consistently some percent bigger than those from the percolation theory. These differences derive from two factors: One as already mentioned above is that the percolation threshold derived from the percolation theory is defined differently as ϕ_c. The relation between the particles' size and the ICA layer thickness is considered in the simulation more realistically as a finite value which means that p_p is always higher than the theoretical percolation probability. The second factor is that a contact force was introduced in this model which equalizes the contraction forces (e.g., from the cured polymer binder) by a variable overlapping area between the particles. This leads to an increased mean value of neighboring particles B_c and changes the system in direction to the soft core model. Moreover, B_c also

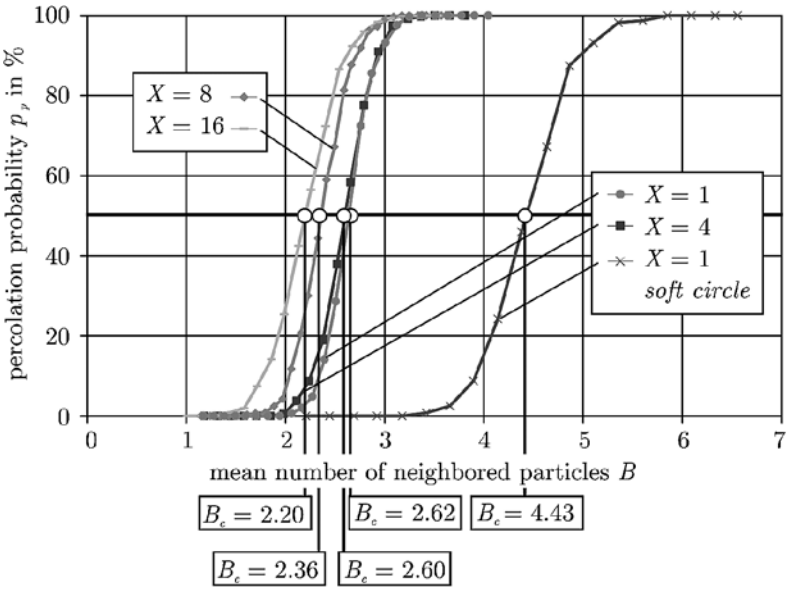


FIGURE 21.10. Percolation probability p_p versus the number of neighbor particles B for elliptical particles with different various aspect ratios ($1 \leq X \leq 16$).

varies with the particle geometry as can be seen in Figure 21.10 depicting the percolation probability versus the number of neighbor particles for elliptical particles with various aspect ratios ($1 \leq X \leq 16$). The diagram demonstrates that a higher number of neighboring particles with circular than with oblong form is needed to obtain the same percolation probability. For comparison, in Figure 21.10 the respective curve for soft circular particles is also displayed. In this case $B_c = 4.43$ which agrees with the respective value reported from Balberg et al. [34]. This comparison allows to understand the fundamental difference between *hard-core* and *soft-core* model.

Typical particle arrangements for illustration of the influence of the particle geometry on the respective filler content necessary to obtain an electrical connection between the pads are depicted in Figure 21.11. The particles have all the same area and in all cases the percolation probability is 50% ($\phi = \phi_c$). However, only arrangements are selected which provide at least one electrical connection between the mating pads. Particles contributing to an electrical path are displayed dark whereas the light particles are isolated or are linked to clusters which have electrical contact at most with only one pad.

In all subpictures it is conspicuous that only a comparatively small percentage of particles is involved in the current transportation. Frequently the current path is necked to a single chain of particles. This fact is directly related to the condition of a percolation probability far below 100%. The electrical consequences are discussed more thoroughly in the next subsection. In particular a strong dependence of the critical filler content on the particle geometry becomes obvious: ϕ_c varies from almost 55% for circular particles to less than 21% for elliptical particles with an aspect ratio of $X = 16$.

Until now we have considered the percolation probability under the condition that the particles may have an arbitrary orientation, i.e., δ was uniformly distributed in the interval from $-90°$ to $+90°$. However, oblong particles like silver flakes in an epoxy adhesive (in order to consider an example of the system most important for practical applications) may

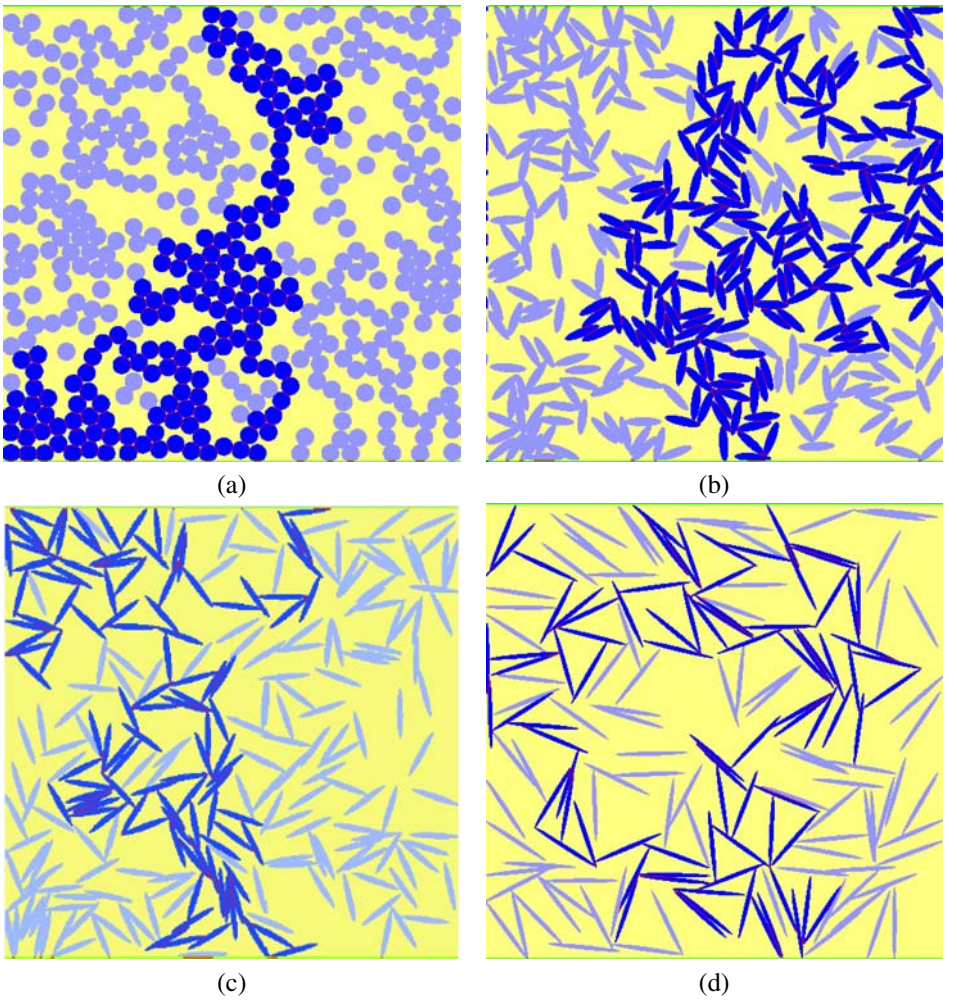


FIGURE 21.11. Particle arrangements with various filler content and various aspect ratio characterized by a percolation probability of 50%, respectively (all particles have the same area). (a) 445 particles, axis ratio 1, filler content 54.5%. (b) 380 particles, axis ratio 4, filler content 46.0%. (c) 255 particles, axis ratio 8, filler content 30.5%. (d) 170 particles, axis ratio 16, filler content 20.8%.

underlay constraints e.g., in the vicinity of the plane contact pads which, to some extent, force a parallel alignment of the particles as can be observed in microsections of ICA bonds. Therefore, a more realistic model should consider these constraints by limiting the freedom of orientation. In Figure 21.12 the effect of different degrees of freedom on the mean normalized excluded area $\langle A_{ex} \rangle$ is demonstrated. A restriction of the orientation to e.g., $-30° \leq \delta \leq +30°$ leads to a reduction of $\langle A_{ex} \rangle$ to about the half of the value for the unlimited freedom of orientation ($-90° \leq \delta \leq +90°$). If all particles are parallel ($\delta \approx 0$) the minimum of $\langle A_{ex} \rangle = 4$ is obtained. The practical meaning of this consideration is, that a higher filler content is needed in cases of a limited freedom of particle orientation.

Whereas in an infinite system the influence of the contact pads on the orientation of the particle does not exist, in the considered limited system this influence plays an impor-

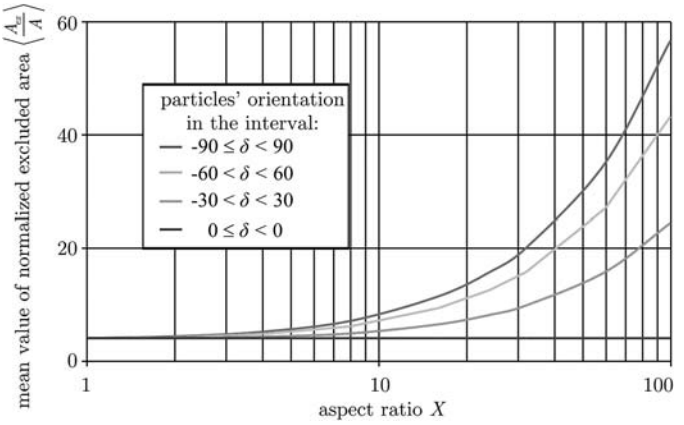


FIGURE 21.12. Mean normalized excluded area $\langle A_{ex} \rangle$ versus aspect ratio for different limitation degrees of possible particle orientation. Within the respective intervals δ is assumed to be uniformly distributed.

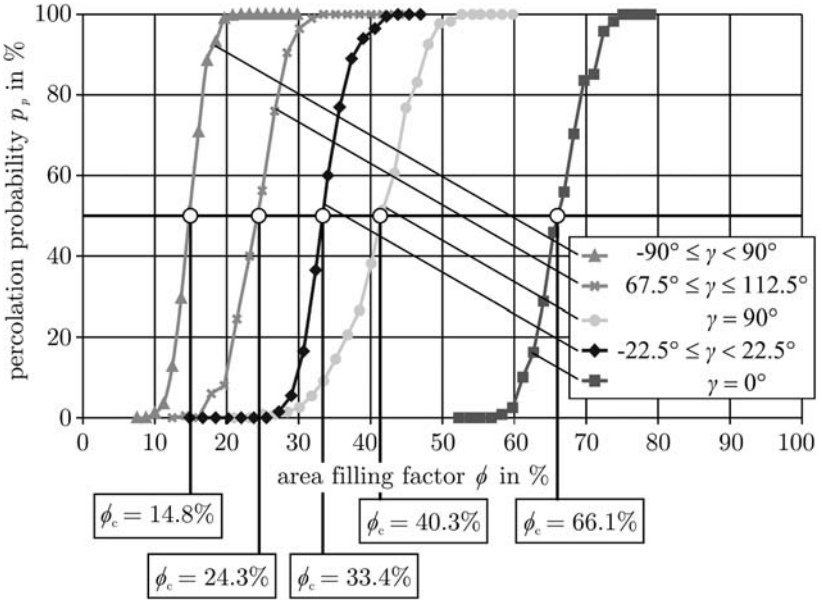


FIGURE 21.13. Percolation probability as function of the filler content for two cases of limited freedom of particle orientation ($\gamma = 0°$ and $\gamma = 90°$) and three different ranges of freedom (γ within $\pm 22.5°$ to the horizontal axis, $\pm 22.5°$ to the vertical axis, and no restriction: $-90° \le \gamma \le 90°$).

tant role. From a heuristic consideration one may expect a lower required filler content if the particles are oriented perpendicular to the pad surfaces. In order to analyze the influence of the particle orientation with respect to the contact pads the angle γ between the long particle axis and the pad plane (horizontal in this case) is introduced. The highest filler content is needed if all particles are aligned parallel to the pad plane ($\gamma = 0°$). On a first glance one could expect the lowest critical filler degree ϕ_c in the case when all particles are oriented perpendicular ($\gamma = 90°$). However, this is not the case. An analysis of the

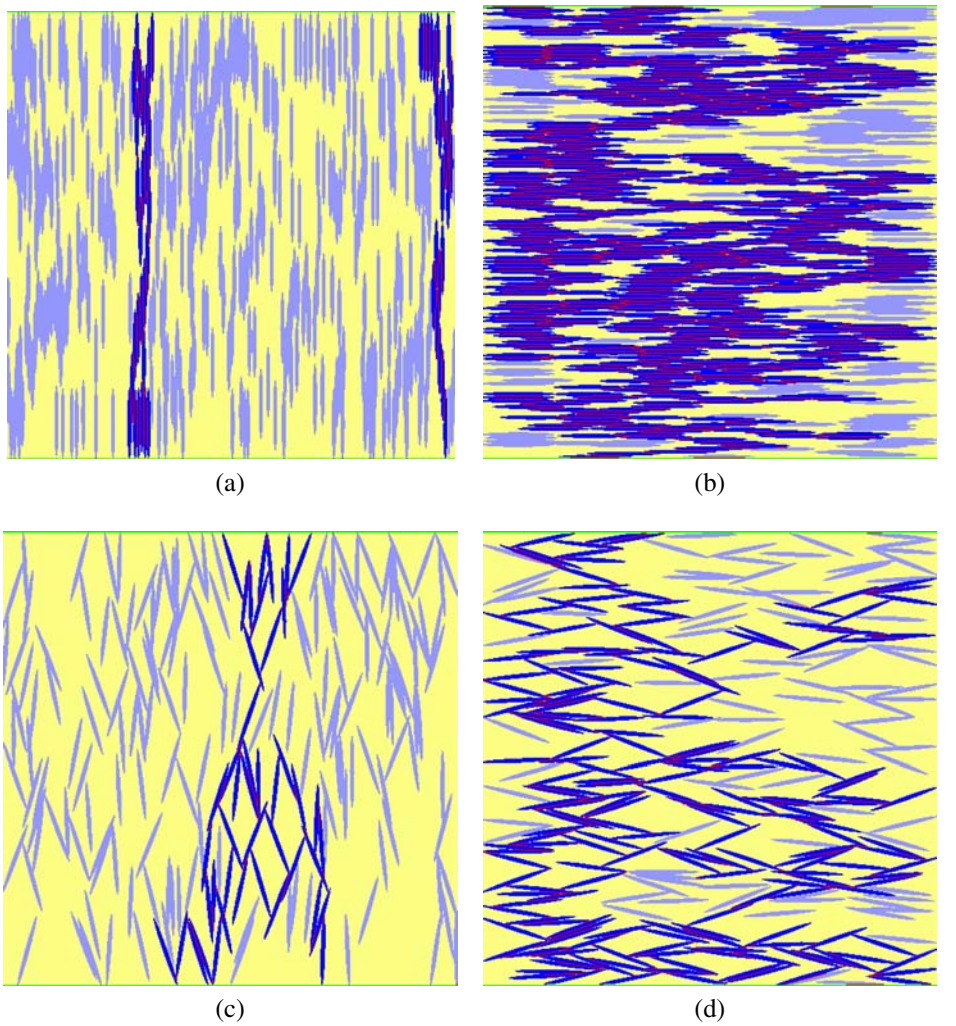


FIGURE 21.14. Accidental particle arrangements for different conditions (a) $\gamma = 90°$, (b) $\gamma = 0°$, (c) γ within $\pm 22.5°$ to the vertical axis, and (d) $\pm 22.5°$ to the horizontal axis. All particles have the same area ($0.02^2 \cdot \pi$), the same aspect ratio $X = 16$, the percolation probability is 50%. Particles contribution to the interconnection between the pads are displayed in dark color. (a) 365 particles, $\gamma = 90°$, filler content 41.6%. (b) 609 particles, $\gamma = 0°$, filler content 66.1%. (c) 199 particles, $67.5° \leq \gamma \leq 112.5°$, filler content 24.2%. (d) 279 particles, $-22.5° \leq \gamma \leq +22.5°$, filler content 33.3%.

percolation probability as function of the filler content clearly shows a lower critical filler content in cases when the particle orientation may vary to some extent (γ within $\pm 22.5°$ to the horizontal axis, $\pm 22.5°$ to the vertical axis, and no restriction: $-90° \leq \gamma \leq +90°$) compared to cases without any freedom ($\gamma = 0°$ and $\gamma = 90°$, Figure 21.13). The lowest possible critical filler content ϕ_c is again reached with an unlimited fan out of the particle orientation.

The role of the particle orientation for the percolation probability is visualized in Figure 21.14 along with four accidental particle arrangements as calculated by this simulation for the different conditions (a) $\gamma = 90°$, (b) $\gamma = 0°$, (c) γ within $\pm 22.5°$ to the vertical

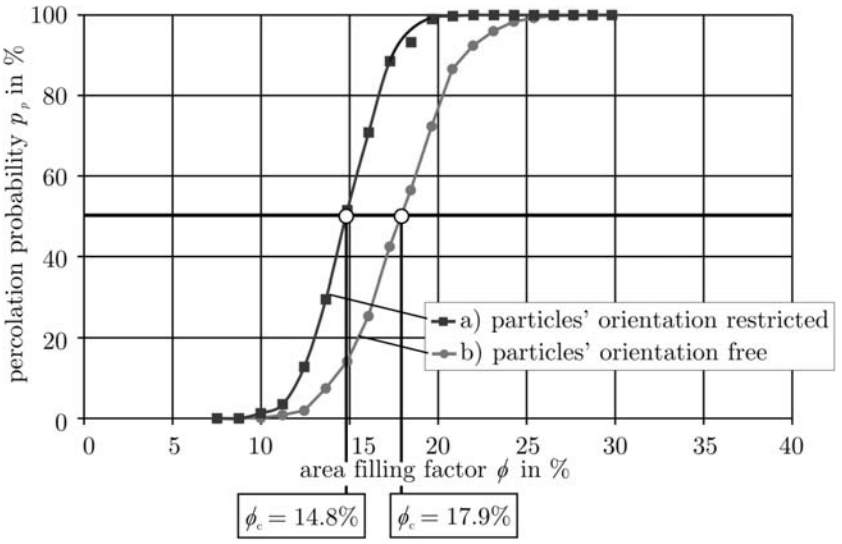


FIGURE 21.15. Percolation probability p_p versus particle filler content ϕ in case of (a) a model with purely translational particle rearrangement and (b) a model with translational and rotational rearrangement ($A = 0.02^2 \cdot \pi$, $X = 16$).

axis, and (d) within $\pm 22.5°$ to the horizontal axis. These simulations were carried out under following assumptions: All particles have the same area, the same aspect ratio, and the filler content is chosen such that the percolation probability is 50% ($A = 0.02^2 \cdot \pi$, $X = 16$, $p_p = p_c$). Again, only those arrangements are depicted where an interconnection between the pads is achieved.

From a comparison of Figure 21.9 and Figure 21.13 one can see that for the same cases of particles with elliptic shape $X = 16$ and no restrictions for the orientation the percolation threshold ϕ_c in one case is almost 18%, in the second case only 14.8%. The respective curves are shown in Figure 21.15. The difference of 3.2% results from the fact that the simulation of the particle arrangement has been carried out in different ways: In the first case the simulation started with the placement of particles with randomly and uniformly distributed orientation and only translational movements were allowed for the rearrangement of the particles. In this way the equidistribution of the orientation was constrained. By contrast, in case of the left curve of Figure 21.15 after placement of the particles besides the translational also a rotational rearrangement was permitted. These rotations allow the particles to increase the extent of parallel orientation of neighbored particles. This effect can counteract against the equidistribution of the orientation and in turn lead to an increase of the percolation threshold. The more the filler content is increased the more significant becomes this effect of paralleling of particles. This effect is clearly visible in Figure 21.16 showing a model with 400 particles being allowed to rotate after their placement. Due to the comparatively high filling degree of $\phi = 45\%$ a high degree of parallel clustering can be observed. Another aspect can also be recognized from this figure: not only the total number but also the percentage of electrically active particles (dark colored) is increased significantly when the percolation probability exceeds the threshold value. The number of parallel current paths per unit of contact area is of great practical importance for joints with high current loads. It must be noted that in any case the current is necked to the microscopic

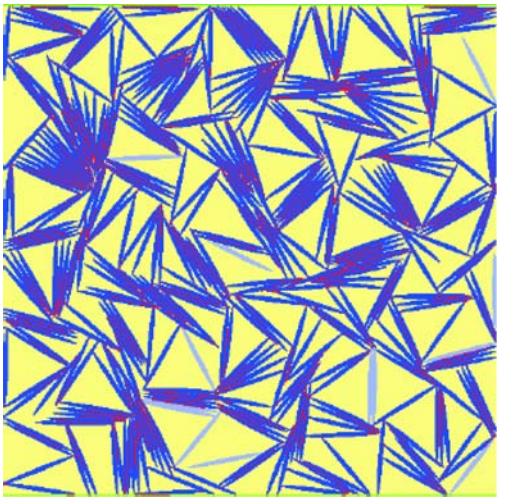


FIGURE 21.16. Model of arrangement with 400 particles. A filler content significantly above the percolation threshold ($\phi = 45\%$) leads to paralleling of particles ($A = 0.02^2 \cdot \pi$, $X = 16$).

interfaces between the particles where an extremely high current density can be reached. In order to increase the number of interparticle contacts the filler content must be far enough beyond the percolation threshold.

In Figure 21.17 a microsection of a real ICA joint is shown. Some regimes with a distinct uniformity of orientation becomes visible. It can be assumed that the final particle arrangement in a real joint does not only depend on production process parameters but also varies e.g., during the printing process and during placing the components due to plastic deformation. As can be recognized in this figure the microhomogeneity is not ensured. The distribution of the particle orientation may vary significantly in a volume with an elongation of several ten to hundred particles. Although the properties of an ICA cannot be quantified accurately with this two-dimensional model, it just provides a deeper understanding of some phenomena and can be used to predict tendencies. The next section deals with the investigation of electric joint properties on the base of simulated particle arrangements.

21.2.3.2. Joint Resistance For a reliable isotropically conductive adhesive joint the percolation probability has to be 100%. In the simulation this means in the case of elliptical particles with an axis ratio of $X = 4$ a filler content of $>60\%$ is needed. In order to establish the percolation probability p_p and the average value of the total joint resistance as functions of the filler content ϕ, 600 simulations with accidental starting condition and identical parameters (but different particle arrangements) were performed for every ϕ value. Results are shown in Figure 21.18. In order to get the resistance curve also for filler contents below a percolation probability of 100% it was necessary to exclude particle arrangements without electrical connection between the contact pads from this consideration, since they would have led to an infinite resistance. However, the lowered number of evaluable simulation results is related to an increase of uncertainty in the resistance calculation which causes widening of the average deviation of the resistance curve.

For simplicity, in a first approximation the contact resistance between the particles is assumed to be zero (ideal contact) and the bulk conductivity σ_{pad} of the two pads is assumed to be 1000 times that of the particles' one ($\sigma_{pad} = 1000 \cdot \sigma$) so that the potential

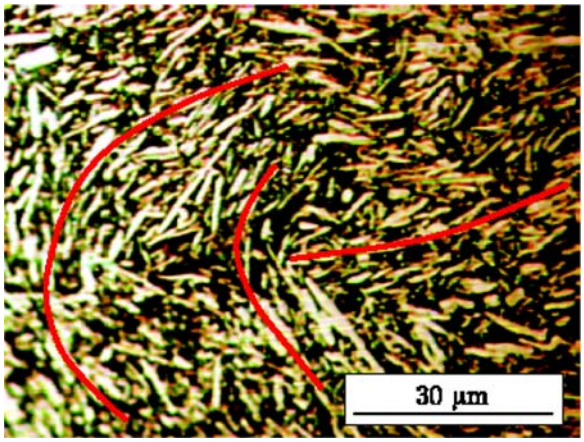


FIGURE 21.17. Microsection of an ICA joint demonstrating small areas with a distinct uniformity of orientation (along red lines). An orientation structure becomes visible.

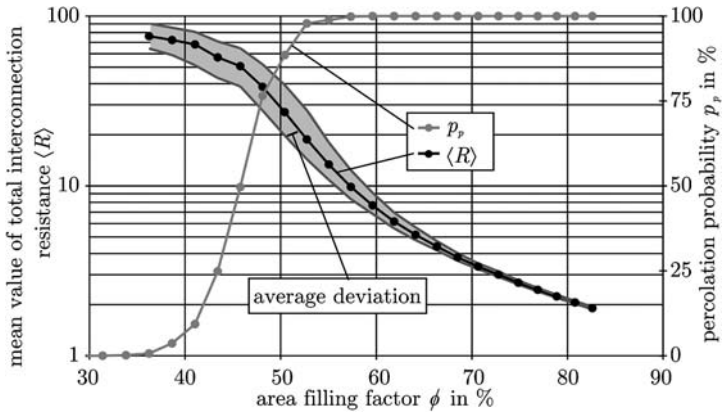


FIGURE 21.18. Probability of interconnection and the average joint resistance versus the filler content (elliptic particles with an axis ratio of $X = 4$).

in the pads can be considered as constant and the total joint resistance depends only on the particle resistance. In order to receive results which are independent from the system size a specific conductivity of the particles of $\sigma = 1$ is introduced. For the calculation of the joint resistance this means: If the quadratic model area would be completely filled with particle material ($\phi = 100\%$) the joint resistance would be 1. At a more realistic filler content of 77% the joint resistance reaches a value of 2.6. In the simulation the first contact occurred at $\phi = 38\%$ leading to a considerably higher joint resistance of 67. In between these two points, at $\phi = 60\%$, the interconnection probability just reaches the 100% value. At this level the normalized total joint resistance amounts 9.4. This shows that the major change of the contact resistance occurs in a range where the percolation probability is below 100%. By increasing the filler content above this point a strongly decreasing width of the average deviation of the joint resistance R can be observed indicating a regime of reliable interconnection.

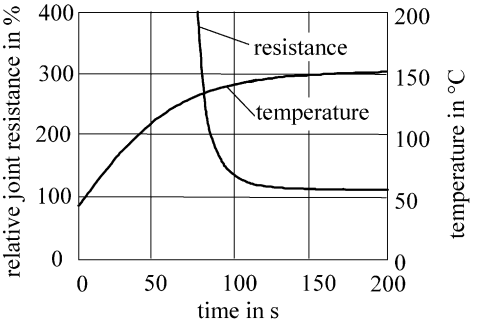


FIGURE 21.19. Change of the joint resistance during curing process.

Although, the filler content as a volume ratio is difficult to compare with that of a two-dimensional model, from a microsectional view of a real ICA joint one would obtain a value ϕ far beyond 60% which means definitely an interconnection probability of 100%. Nevertheless, an ICA just after being dispensed or printed on a substrate behaves like an insulator. The low joint resistance corresponding to an interconnection probability of 100% arises only successively during the curing process. In Figure 21.19 such a break-down of the joint resistance at the beginning of the curing process is depicted. This phenomenon is based on a strongly changing contact resistance R_{pp} between the particles which are brought into an intimate contact during the development of the contraction forces of the polymer binder as described in Section 21.2.2. It can be considered in this model e.g., by introducing a time and temperature dependent function describing the effect of the curing process. Such a simulation shows that the interface resistance between the particles and certainly also between particles and pads are the most significant parameters for modeling an ICA joint. In order to take into account different interface effects like corrosion at the pad-particle interfaces or oxidation between particles, both effects known as aging phenomena, the specific interface conductivity λ_{pp} between two particles and λ_{ppad} between particle and pad are introduced.

By contrast to Figure 21.18 where all transition resistances at interfaces where neglected let us assume the more realistic case of a finite conductivity at the interfaces. Taking the same particle arrangement but values of λ_{pp} and λ_{ppad} between 1 and 0.001 one can observe that in a logarithmic scale the total joint resistance as functions of the filler content have the same shape but are shifted to higher $\langle R \rangle$ values. This means that the effect of a change of transition conductivity is almost the same at any degree of filler. How strong this effect is, depends on the relation between the conductivity of the particles and their interfaces. The lesser the interface conductivity is, the more it controls the total joint resistance.

By keeping the interface conductivity between the particles constant at $\lambda_{pp} = 1$ and varying the one between particles and pad dependencies of the total joint resistance on ϕ are obtained (Figure 21.21). This corresponds to the practical case of a joint degradation e.g., due to galvanic corrosion at the interface between metalization and ICA.

It can be recognized that now the effect of λ_{ppad} on the total joint resistance is much weaker than in the case when $\lambda_{pp} = \lambda_{ppad}$ (Figure 21.20). A change of interface conductivity λ_{ppad} does not appear any more as resistance multiplier but rather acts as a supplement to the resistance of the ICA fill. This becomes clearer by considering the pad surfaces as equipontential planes and by defining a total transition resistance $\langle R_{tr} \rangle$ consisting of inter-

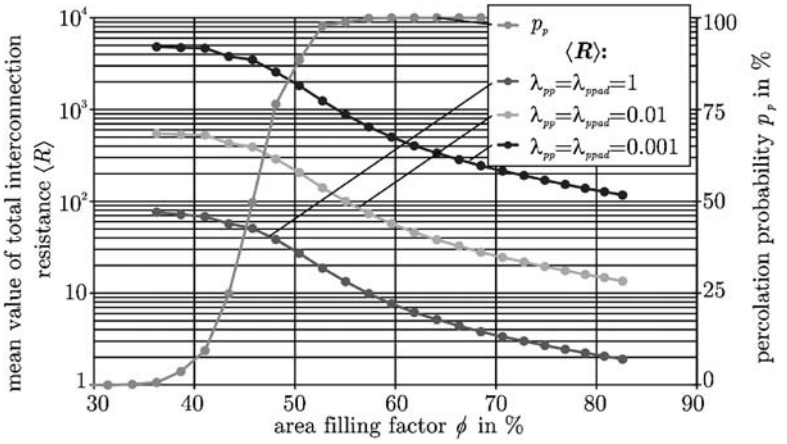


FIGURE 21.20. Normalized total joint resistance and percolation probability versus filler content for various values of normalized interface conductivity under the condition of $\lambda_{pp} = \lambda_{ppad}$.

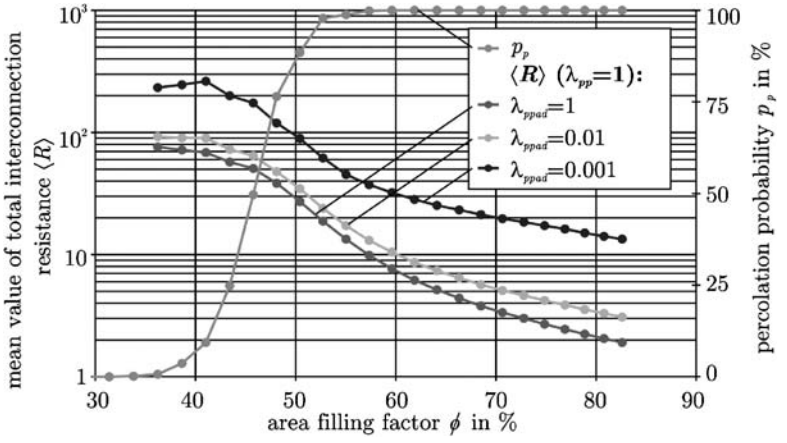


FIGURE 21.21. Normalized total joint resistance like in Figure 21.20 but for constant normalized particles' interface conductivity ($\lambda_{pp} = 1$).

face resistances of only those n particles touching the pad which are also involved in the current path according to the following relation:

$$\langle R_{tr} \rangle = \frac{1}{\sum_i^n \lambda_{ppad} \cdot w_{s,i}} = \frac{1}{\lambda_{ppad}} \cdot \frac{1}{\sum_i^n w_{s,i}}, \tag{21.3}$$

where $w_{s,i}$ is the length of the intersection line of the ith particle. The sum

$$w_s^* = \sum_i^n w_{s,i} \tag{21.4}$$

of all these lengths can be understood as *effective contact length* to which the current is necked at the respective pad interface. The total pad interface resistance is the sum of both

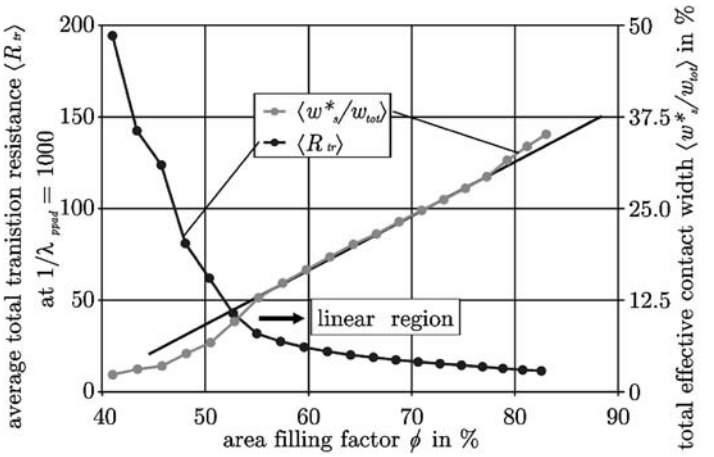


FIGURE 21.22. Total transition resistance at both pad-particle interfaces and normalized effective contact length as functions of the filler content.

interface effects (which are equal for both pads in this case but one could consider different interfaces as well):

$$\langle R_{tr} \rangle = \frac{2}{\lambda_{p\,pad} \cdot w_s^*}. \tag{21.5}$$

This resistance can again be calculated as function from the filler content. A comparison with the results depicted in Figure 21.20 shows that it corresponds very well to the difference of the total joint resistance functions of the two cases $\lambda_{pp} = \lambda_{p\,pad} = 1$ and $\lambda_{pp} = \lambda_{p\,pad} = 0.001$. This is illustrated by Figure 21.22. This diagram also shows that the average effective contact width normalized to the total width w_{tot} of the modeled space demonstrates a linear increase with the filler content in a wide regime above the 100% percolation threshold. In Figure 21.23 the particle arrangement of a model with 70% filler content is depicted where the effective contact length amounts around 25% of the total model width.

For practical applications it is important to note that there is a distinct decrease of the interfacial resistance with increasing filler content. One could expect an improvement of quality of an ICA joint with rising filler content. However, in this case the bonding forces resulting from the adhesion of the polymer to the pad surface are reduced to the complement part of the effective contact area. This consideration allows to understand that a compromise between high mechanical strength and low total joint resistance has to be found which depends on the respective application field.

A further parameter influencing the joint resistance is the particle shape, since this is a controlling parameter of the percolation behavior as discussed before. For this purpose the joint resistance for various aspect ratios of elliptic particles is investigated. Figure 21.24 shows the results for the cases $X = 8$, $X = 4$, and circular particles, when $\lambda_{pp} = \lambda_{p\,pad} = 1$. The results clearly show qualitatively similar dependencies versus the filler content whereby the critical filler content is shifted in almost equal steps from 30%

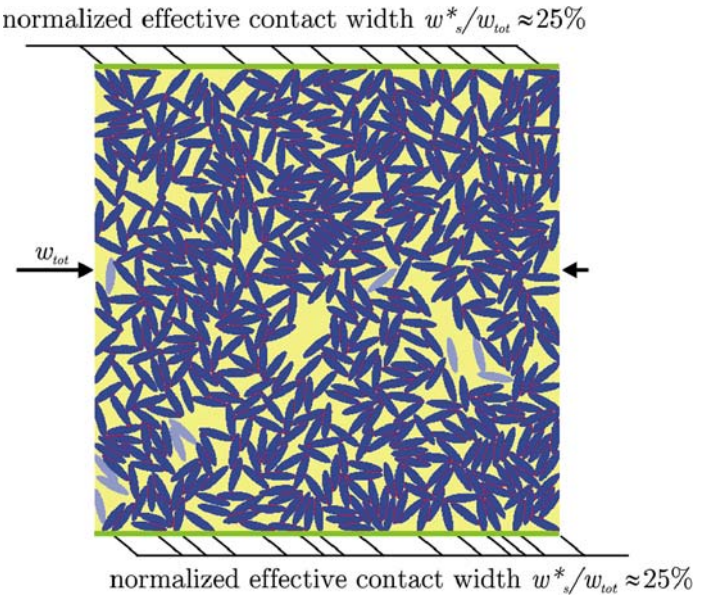


FIGURE 21.23. Particle arrangement of a model with 70% filler content providing a relation of around 25% effective contact width to total model width.

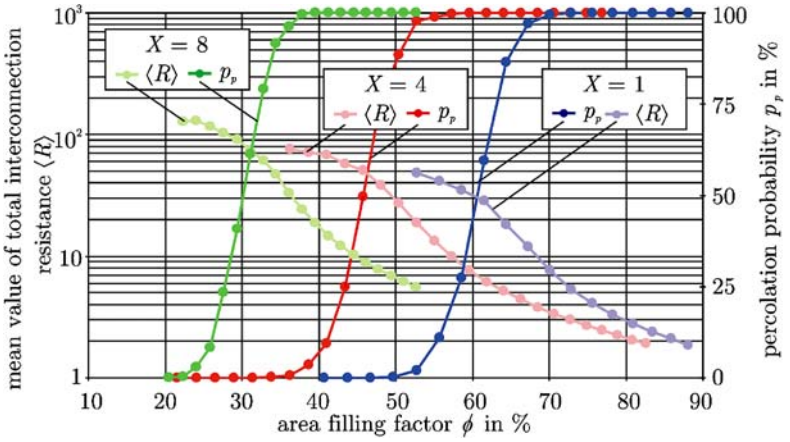


FIGURE 21.24. Total joint resistance and percolation probability versus filler content for particles with various aspect ratios.

for the oblong particles to around 60% for the circular ones. This comes out clearer by introducing the deviation $\Delta\phi$ of the filler factor ϕ from critical value ϕ_c as

$$\Delta\phi = \phi - \phi_c. \tag{21.6}$$

This transformation shifts the results of Figure 21.24 such that the point of percolation remains the same for all three cases and shows that also the minimum additional

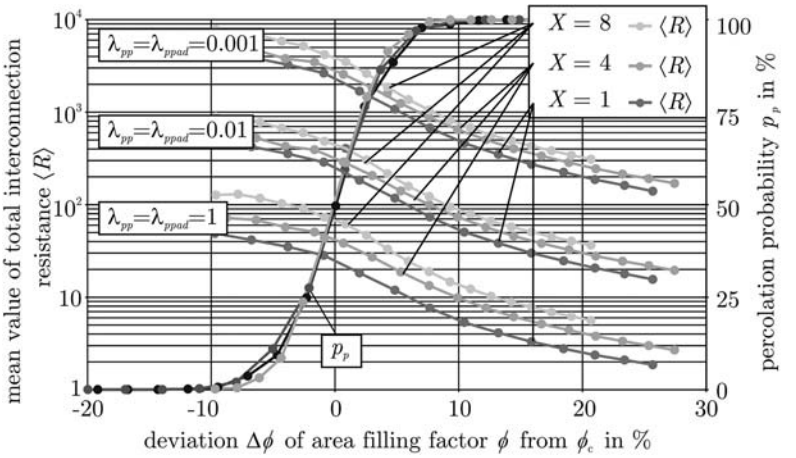


FIGURE 21.25. Results from Figure 21.24 but scaled as functions of the deviation $\Delta\phi$ of the filler content from the critical one for various cases of interfacial conductivity.

amount of filler content $\Delta\phi_{\min}$ which is needed in excess to ϕ_c to allow to form a reliably operating joint is the same in all three cases ($\Delta\phi_{\min} \approx 10\%$, Figure 21.25).

The practical meaning of this fact is that if the critical filler content of a system with a certain particle shape is known the necessary filler content of a system with a new particle shape can be found by increasing the filler content at least by the same percentage $\Delta\phi_{\min}$ as in the known system.

Having in mind the afore mentioned relation between the filler content and the mechanical qualities one would undoubtedly prefer adhesives with long particles, since the lower the filler content is the bigger are the expectable bonding forces. However, it turns out that the joint resistance at the same $\Delta\phi$ is also shifted to higher values in case of the longer particles: In the case of $X = 8$ the joint resistance is between two and three times as high as in case of $X = 1$. Since the electrical parameters of all simulations ($X = 1$, $X = 4$, and $X = 8$) remained unchanged the shift of joint resistance can only be explained by geometrical effects such as the size of overlapping areas characterizing the intensity of the contacts, the number B of neighboring particles which is responsible for the density of the resistance network, and the geometry of the particles themselves. However, it can be shown that B doesn't change significantly in all discussed cases. Thus, B is not responsible for the big resistance changes. If it is further determined that the size of the overlapping contact areas and therewith the respective widths $w_{s,i}$ of the particles are only slightly depending on the filler content one can realize that the main part of the joint resistance change has to be assigned directly to the particle shape.

Besides the results of the case $\lambda_{pp} = \lambda_{ppad} = 1$ the joint resistance functions of the cases $\lambda_{pp} = \lambda_{ppad} = 0.01$ and $\lambda_{pp} = \lambda_{ppad} = 0.001$ are also demonstrated in Figure 21.25. A constant vertical shift of corresponding curves indicates that in all cases an increase of the axis relation X causes an increment of $\langle R\rangle$ of almost the same amount. This applies also for the dependence of the joint resistance functions on the interparticle conductivity.

For practical applications it can be concluded from this model that the particle shape (in the model represented by the aspect ratio for elliptical particles) has a certain influence on the joint resistance. However, the influence of the same relative change of filler content is much stronger. Moreover it can be observed that the joint resistance has a weaker depen-

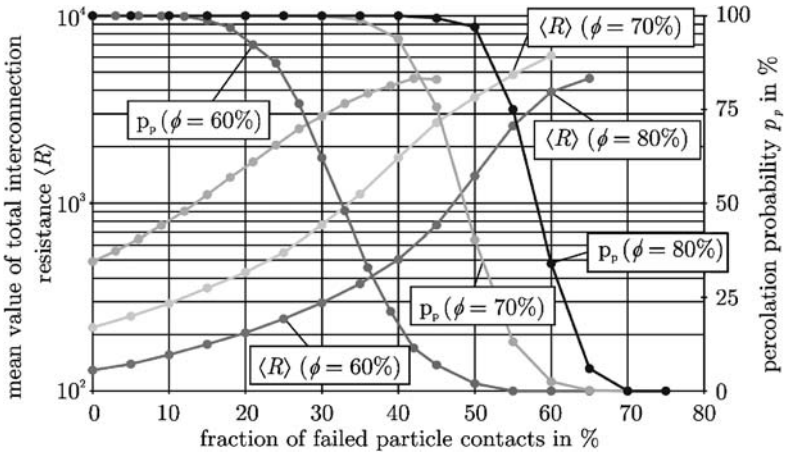


FIGURE 21.26. Joint resistance and percolation probability versus percentage κ_o of failed interparticle contacts for three values of filler content (particle parameters: $X = 4$, $A = 0.02^2 \cdot \pi$).

dency on the interparticle conductivity at high values of λ_{pp} and λ_{ppad} (a change of two orders of magnitude causes a resistance increase of only one order of magnitude). However, the more the interparticle conductivity decreases the more the joint resistance approaches to the inversely proportional dependence on the interparticle conductivity. Such a change could take effect if the molecular bonding forces of the polymer are weakened and oxides are formed at the particles interfaces.

As known from numerous investigations of the aging behavior under real operation conditions as well as under accelerated aging conditions the joint resistance can increase significantly [40–43]. By anticipating that changes do not occur in the volume of the conducting particles the resistance increase has to be assigned to degradation of the interfaces between particles and pad. The following part of the discussion is devoted to changes of the joint behavior due to aging.

In Figure 21.25 the shift of joint resistance with falling interparticle conductivity can be observed in cases when all interparticle contacts become homogeneously deteriorated to the same extent in the whole contact. However, this will remain the exception, since with growing thickness the contribution of tunneling becomes the dominant effect for the conductivity through the interface barriers. The initial specific tunneling conductivity e.g., on a noble metal surface is in the order of 10^{12} S/m^2 and with continuous growth of an oxide layer the conductivity decreases rapidly [44]. With respect to microscopic inhomogeneities in the ICA in a more realistic consideration, therefore, it should be assumed that due to e.g., oxidation more and more transitions between particles fail rather than that the interparticle conductivity decreases continuously and equally distributed. The results of simulations of such an aging behavior for joints with three different values of filler content ($\phi = 60\%$, $\phi = 70\%$, and $\phi = 80\%$) are demonstrated in Figure 21.26. For this purpose, the interparticle conductivity of a certain percentage of particle transitions are changed from $\lambda = 0.001$ to $\lambda = 0$ which means a total drop-out of the respective transitions. This percentage can be interpreted as the probability κ_o for such a drop-out. Thus, an increase of κ_o corresponds to an aging process in the ICA interconnection. At the beginning of the increase of κ_o only the joint resistance increases whereas the percolation probability remains constant at 100%. Only with continuing aging also the percolation probability starts to fall, indicating failing

of the joint to the respective percentage. It is obvious, that joints with a higher filler content have a lower joint resistance and will withstand a longer aging process without failing which could be interpreted some how as a better long-term reliability. However, it is remarkable that the joint resistance in the moment when the percolation probability starts falling below 100% is almost independent from the filler content.

The change of contact resistance for various types of ICAs, different curing processes, and aging conditions were also experimentally investigated [45] and are summarized in the following section.

21.3. AGING BEHAVIOR AND QUALITY ASSESSMENT

21.3.1. Introduction

Besides the benefits mentioned in the introduction, ICAs provide an environmentally friendly alternative to solders for interconnections in electronic applications with the advantage of a superior fine pitch capability. However, unstable electrical conductivity under elevated temperature and humidity conditions and a low impact resistance of the interconnections were major obstacles preventing ICAs from becoming a general replacement for solders in SMT until now.

It is a well accepted opinion that corrosion is involved in the shift of contact resistance [20,24,46,47]. It is reported that galvanic corrosion rather than simple oxidation is the dominant interfacial degradation mechanism. However, galvanic corrosion doesn't only need an electrochemical potential difference but would also require a ions-containing aqueous phase which would have to be formed at e.g., 85% r.h.—that means at a humidity level where wet surfaces usually dry-up. On the other hand, galvanic corrosion would be an explanation why a much faster increase in resistance is attained under elevated humidity than under dry conditions at the same elevated temperature [25].

Recently developed ICAs promise an improved electrical and mechanical reliability on non noble metalizations as conventionally used in SMT and can be completely cured in the same short duration as known from a typical reflow soldering cycle. In order to investigate connections between processing parameters and the aging behavior eight different ICAs for solder replacement from four mayor manufactures were evaluated with respect to the behavior of the contact resistance and the shear force during a forced aging process. For this purpose test assemblies with ICA sample contacts were fabricated using PCBs with the four most common surface finishes. The test assemblies were then exposed to an elevated temperature and humidity environment (85°C/85% r.h.) for 1000 h, whereby the changes of the electrical contact resistance and the shear force were monitored.

The test assemblies consisted of SMD-chip components (1206 and 0805) which were mounted on FR4-printed circuit boards with the sample contacts between different surface finishes and the respective component metalization. Two different curing methods were investigated: curing in a conventional convection oven and curing in a vapor phase device.

The goal of this study was to experience advantages and drawbacks of ICAs presently available on the market and to present some recent progresses in ICA technology development.

21.3.2. Material Selection and Experimental Parameters

All selected adhesives are one component, epoxy based, thermosetting, silver filled and isotropically conductive. All are especially designed to replace solder pastes in SMT.

According to the manufactures data sheets, some of them promise a stable contact resistance even when applied on tin, tin/lead and OSP-coated copper. Table 21.1 gives an overview about the selected products and curing conditions as recommended by their manufacturers.

The test boards were pluggable and allowed to combine monitoring of contact resistance shift with the four-point probe method of 40 contacts simultaneously during forced aging (Figure 21.27), and after that mechanical quality testing.

In this way the voltage drop of the series connection of two joint resistances and the resistance of the component itself is measured. In order to keep uncertainties during evaluation of the joint resistance as small as possible chip resistors (1206 and 0805 package) with a low value and a small tolerance (51 mΩ $\pm$5%) were used provided with lead free galvanic tin coated contacts. The joint resistance is defined as the resistance between contact pad and metalization of the component. The voltage drop within the copper pads and the metalization of the component can be neglected in this arrangement.

TABLE 21.1.
Chosen ICAs with the curing conditions as recommended by the manufactures.

Manufacturer	Adhesive	Recommended curing conditions
A	1	120 min @ 120°C
		30 min @ 140°C
	2	5 min @ 125°C-Reflow
		3 min @ 150°C-Reflow
		8 min @ 125°C-Convection
		5 min @ 150°C-Convection
B	1	30 min @ 125°C
		15 min @ 150°C
	2	120 min @ 125°C
		60 min @ 150°C
C	1	10 min @ 125°C
		6 min @ 150°C
		3 min @ 175°C
	2	6 min @ 130°C
		3 min @ 150°C
D	1	30 min @ 140°C
	2	N.A.

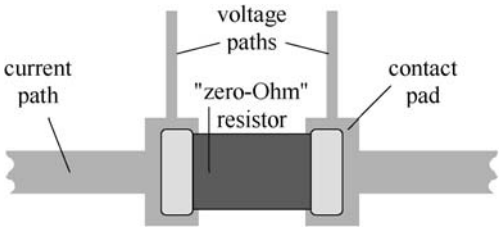


FIGURE 21.27. Conductor structure of one component.

With respect to the most commonly used surface finishes of PCBs test boards with the following types of surface treatments were manufactured:

– organic surface protection (OSP): ENTEK Plus Cu-106A
– electroless Ag-coating
– electroless Ni/Au-coating
– electroless Sn-coating.

The curing cycles were optimized by monitoring the electrical conductivity of the bondings during the curing process. This method is explained in the following subsection.

21.3.3. Curing Parameters and Definition of Curing Time

The results of quality testing of ICA joints depend strongly on how the adhesive has been cured. A comparison of quality, therefore, needs a thorough definition of curing conditions. Two different curing methods were applied in comparison. The first one as reference was curing in a conventional convection oven at a preselected temperature (two values were considered: 150°C or 200°C). The second method was curing in a saturated vapor phase of perfluorinated fluids with boiling points of 155°C and 200°C. The idea of using vapor phase condensation is to take advantage of the much higher heat transfer rates compared to gas convection allowing to shorten down the relatively long curing time of most ICAs to the length of a typical reflow soldering cycle by reducing the heating-up time.

In general ICAs have a high electrical resistance before curing. The conductivity develops during the curing process [48]. It is assumed that shrinkage of the resin matrix during curing causes the silver flakes to contact more intimately which leads to the dramatic contact resistance decreases [49]. If the relation between conductivity and curing degree of the adhesive is known for the respective combination of adhesive and curing method it can be used to determine the needed curing time. Figure 21.28 shows how the conductivity develops during the curing process. For this purpose, calibration measurements were performed with samples where the adhesive was printed over a spacing between two contact pads forming a quadratic resistor (length and width: 1 mm, thickness: 100 μm). The resistance between these contact pads was observed during the curing process. Reliable representa-

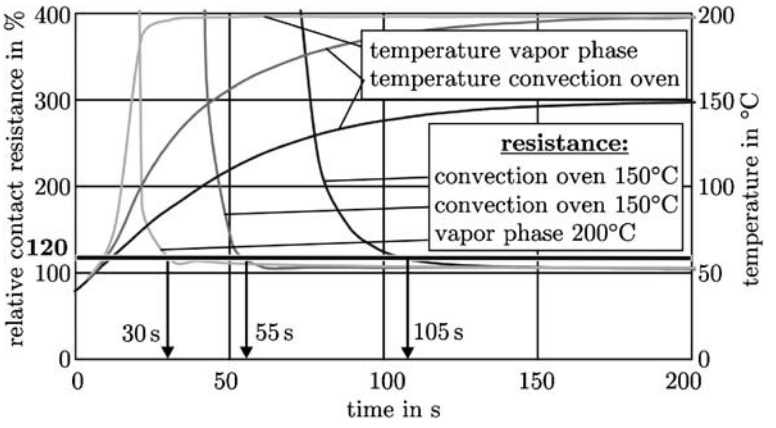


FIGURE 21.28. Change of contact resistance during curing process (adhesive type A1).

TABLE 21.2.
Applied curing times in minutes required to achieve 120% of the minimum possible resistivity.

Adhesive	Convection		Vapor phase	
	150°C	200°C	155°C	200°C
A1	8	3	2	2
A2	8	4	5	3
B1	20	4	15	3
B2	40	20	30	10
C1	8	2	3	2
C2	8	2	3	2
D1	8	5	6	3
D2	8	5	6	3

tive calibration curves were obtained by averaging results from six measurements at each combination. This kind of contact resistance measurement is an easy and reliable method of online monitoring of the curing progress allowing to closed-loop control the curing time and to optimize the curing condition for PCBs with different thermal properties.

Figure 21.28 shows the result of these measurements for curing the A1-type adhesive at 150°C and 200°C in the convection oven and for curing at 200°C in the vapor phase device. Curing at 200°C in the convection oven causes a faster decrease of the resistance as at 150°C due to the faster polymerization process. However, the resistance drop at curing the adhesive with 200°C in the vapor phase device is still much earlier and faster due to the significantly reduced heating-up time of the sample. As an indicator for the completion of the curing process the moment is used when the average resistance reaches the 120% mark of its final value (which was obtained from reference samples after a curing time of 1 hour). The time span until this moment multiplied by a safety factor was rounded up to at least 2 min. This time was defined as curing time for each evaluated adhesive/curing method combination and used for the further sample preparation. Table 21.2 gives a summary of the curing times determined in this way. In general the curing time as defined above is shorter than the one recommended in Table 21.1.

21.3.4. Testing Conditions, Typical Results, and Conclusions

For each particular combination of material and curing method the testing results from 80 adhesive joints were established. To evaluate the reliability of the contacts these samples were exposed to an elevated temperature and humidity (85°C/85% r.h.) environment for 1000 hours. During this forced aging process the electrical contact resistance was measured every 100 hours at room temperature and the results were averaged. The shear force was measured in the initial state, after 400, and after 1000 hours. At every measurement 6 to 7 components were sheared off. If either the contact resistance was higher than 5 Ω or the component dropped off from the test board the contact was counted as "failed" and not considered in further evaluations.

21.3.4.1. Influence of PCB finish on aging behavior The A1 type is a commercially applied ICA but not especially designed for the use on less noble metal surfaces. Therefore, it is not surprising that joints fabricated with this adhesive show a drastic increase of contact resistance during the testing period on PCB finishes like tin or copper [50,51]. Since the

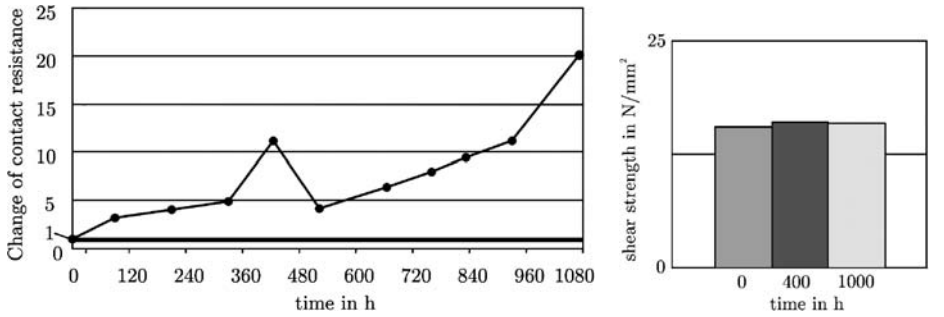


FIGURE 21.29. Adhesive A1 on Ag-coated PCB during forced aging test (85°C/85% r.h.); curing at 150°C in the convection oven. (a) Change of joint resistance (initial value: 10 mΩ). (b) Change of shear force.

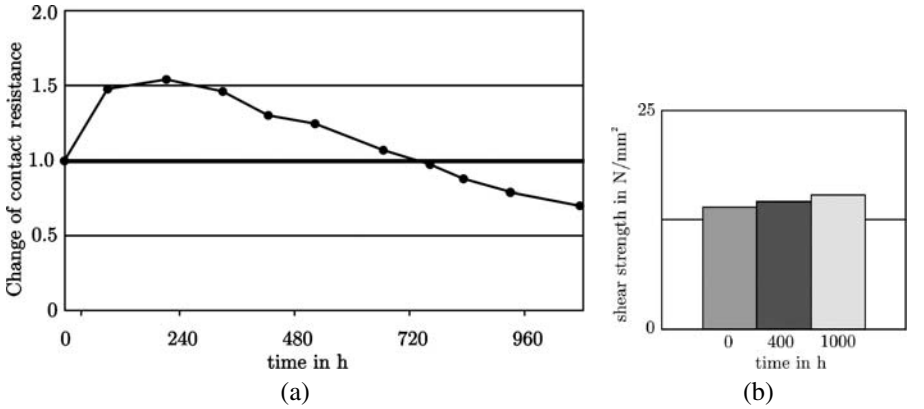


FIGURE 21.30. Adhesive A1 on Sn-coated PCB during forced aging test (85°C/85% r.h.); curing at 150°C in the convection oven. (a) Change of joint resistance (initial value: 37 mΩ). (b) Change of shear force.

increase of contact resistance is observable on each of the tested PCB-finishes it is assumed that the interface between the ignoble tin surface of the component and the adhesive is responsible for this effect. Despite of the high shift of contact resistance the shear strength did not decrease. Figure 21.29 shows the results measured at the Ag-coated PCB cured in the convection oven at 150°C.

The more recently developed product A2 is designed for solder replacement and suitable also for mounting components with tin surfaces. On all samples the average contact resistance after the forced aging test was even lower than the initial value. For the samples cured in the convection oven at 150°C only a slight increase of contact resistance in the first quarter of the testing time and a subsequent decrease of the resistance were observed. Figure 21.30 shows the results obtained from Sn-coated PCBs cured in the convection oven at 150°C. Also the shear force shows stable values over the entire testing period. Only at the NiAu coated PCB a decrease of the shear force was noticeable.

Another example is a product C2 offered as a comparatively quickly curing adhesive (3 minutes at 200°C) showing a strong increase of resistance only on the Sn-coated PCB. Rather high initial values decreased significantly after about 300 hours of aging (Figure 21.31). The used curing time of only 2 minutes was certainly insufficient.

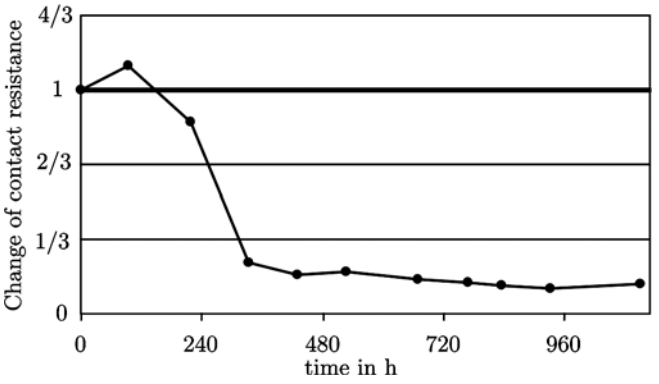


FIGURE 21.31. Change of contact resistance; adhesive C2 on OSP-coated PCB during forced aging test (85°C/85% r.h.); cured at 200°C in the convection oven.

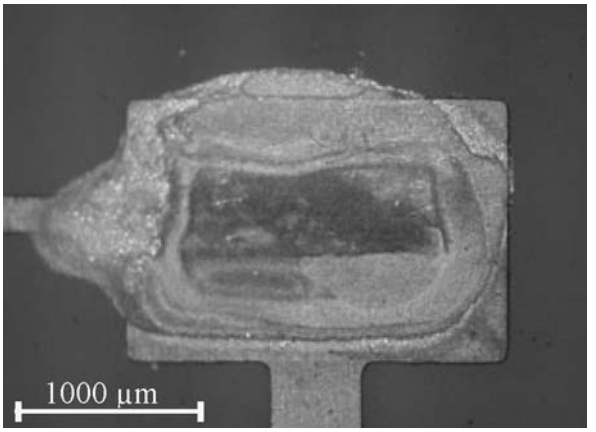


FIGURE 21.32. Area of fracture after shear test after 1000 hours forced aging (adhesive D2, curing condition: vapor phase 155°C, PCB OSP-coated).

Figure 21.32 depicts an area of fracture after a shear test of a sample joint manufactured with adhesive D2 after 1000 hours of aging. The fracture occurred at the PCB/adhesive–interface. A discoloration of the pad surface is visible resulting from a strong degradation of this interface. It should be noted that a low shear force can appear independent from a good conductivity.

By contrast, Figure 21.33 shows the area of fracture of a contact formed with adhesive D1 using the same parameters as in Figure 21.32. In this case the fracture occurred mainly at the component/adhesive interface. At the PCB pad where the fracture occurred a blank, shiny copper surface appeared indicating a low degradation. It is not surprising that this sample showed good mechanical and electrical properties.

21.3.4.2. Summary Eight different ICAs from four mayor manufactures were evaluated with respect to there suitability to serve as replacement of solder pastes. As a general result it should be noted that significantly different behavior of the tested adhesives could be observed in spite of consisting of quite similar compositions: silver-filled one-component epoxy resin.

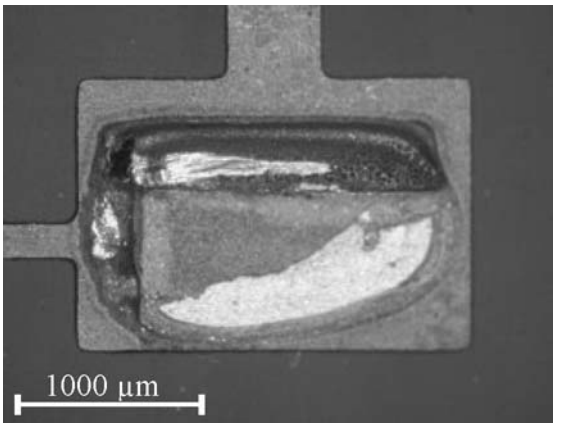


FIGURE 21.33. Area of fracture after shear testing a sample with a type D1 adhesive after 1000 hours forced aging (curing condition: in vapor phase at 155°C, PCB: OSP-coated).

Adhesives of the A1-type and D1-type are not recommended for oxygen rich systems or oxidizing surfaces from their manufacturers. Although, initial values of joint resistivity and shear strength were acceptable, the results during and after a forced aging process exhibit large shifts and an unstable values.

Using the vapor phase device generally the curing time of the adhesives can be decreased drastically due to the high heat transfer rate obtained by condensation. Although, with some adhesives perfect results were obtained with others the shear strength showed significantly lower values compared to equal samples cured in the convection oven. Further investigations would be required to fully understand the reasons for these different behaviors.

Using the NiAu-finished PCBs an increased failure rate was observable for the most adhesives. At the first glance this might be surprising since due to the low inclination to oxidation of noble metal surfaces one could expect a rather moderate tendency to changes of the contact resistance. However, the increase of resistance during aging might to a certain extent be assigned to the weaker adhesion between gold and the epoxy resin. The latter argument is in good agreement with the results of the shear tests frequently showing the area of fracture coinciding with the gold-adhesive interface.

The newer ICA types didn't show a significant increase of contact resistance during a forced aging under elevated temperature and humidity even on ignoble metal surfaces as well as on the OSP coated copper pads. Although, the protective layer could be assumed to impede an intimate contact between the filler particles and the metalization, rather the opposite turns out from the investigations as described afore. No retardation of the onset of contact formation during curing can be observed and a stable joint resistance during aging can be understood as an indicator for a reliable long-term behavior.

As a summarizing result of this investigation it can be concluded that the progress of the adhesive technology has reached a stage were a replacement of solder by an isotropic conductive adhesive can be performed in many cases without changing the involved surfaces treatments whereby a reliability level can be achieved comparable to solder joints.

21.4. ABOUT TYPICAL APPLICATIONS

Due to the large variety of filler content and chemical composition of the polymer the industrial use of ICAs is not restricted to the attachment of SMDs. The following examples demonstrate the numerous application fields.

21.4.1. ICA for Attachment of Power Devices

For optimizing device performance and reliability of power devices an accurate thermal design is a critical issue. The basic need is to remove the power loss produced in the device (junction) under operation condition at the lowest possible temperature drop with respect to the ambient. This quality is expressed as the junction-to-ambient thermal resistance $R_{th,j-a}$. It consists of the sum of the junction-to-case thermal resistance $R_{th,j-c}$ and the thermal resistance $R_{th,c-a}$ from the case to the ambient. Whereas $R_{th,j-c}$ is provided in the datasheet from the component's manufacturer $R_{th,c-a}$ highly depends on the mounting technique and can vary orders of magnitude. In power assemblies the device case is attached to a heat sink using a thermal grease, by adhesive bonding, or by soldering. Normally, heat dissipation from the case by radiation and convection can be neglected. $R_{th,c-a}$ is mainly controlled by the heat flow through the bond line between the device case and the heat sink. Therefore, $R_{th,j-a}$ can be expressed as

$$R_{th,j-a} = R_{th,j-c} + R_{th,c-a} = R_{th,j-c} + R_{th,bond} + R_{th,heatsink}, \tag{21.7}$$

where $R_{th,heatsink}$ is defined as the thermal resistance between the mounting surface of the heat sink and the ambient. The bond line thermal resistance $R_{th,bond}$ can in principle be calculated by dividing the expected or measured bond line thickness by the adhesive's intrinsic thermal conductivity λ_{bond} measured on a free-standing cured sample. However, at a typical bond line thickness of 15–75 μm, the interface thermal resistance $R_{th,if}$ between the adhesive and its adherents can be significant compared to the intrinsic thermal resistance of the adhesive and thus the bond line thermal resistance itself must be considered as a sum of the two components [52]

$$R_{th,bond} = \frac{t_{bond}}{A \cdot \lambda_{bond}} + R_{th,if}, \tag{21.8}$$

where A and t_{bond} are the bond line area and thickness. The practical importance can be seen on the example of a TO-247 package which is attached to a heat sink with a copper heat spreader as mounting surface (Figure 21.34). Two types of ICAs are compared: a conventional one with a low thermal conductivity and an ICA (type Diemat DM6030Hk) with a silver particle content of more than 95 weight percent.

For mounting the device on a heat sink, the ICA must flow as freely as grease to eliminate air voids and reduce the thermal resistance of the interface. However, a complete avoidance of pores and gas enclosures in the bond line is not always possible. In order to consider the thermal meaning of enclosures in the bond line Equation (21.8) has to be modified such that the thermally conducting cross section of the ICA is reduced by the area percentage p of the voids:

$$R_{th,bond} = \frac{1}{A} \cdot \frac{t_{bond}}{\lambda_{bond} \cdot \left(1 - \dfrac{p}{100}\right) + \lambda_{air} \cdot \dfrac{p}{100}} + R_{th,if}. \tag{21.9}$$

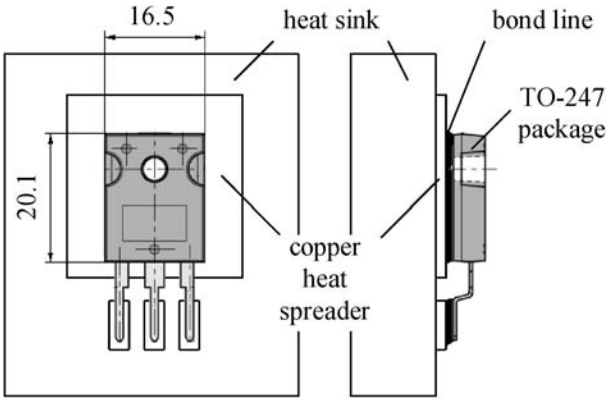


FIGURE 21.34. Power device assembly with TO-247 package (dimensions in millimeter).

TABLE 21.3.
Geometry and thermal data used for evaluation as demonstrated in Figure 21.35.

Bond line thickness	in µm	50	
Bonding area	in cm^2	2.31	
Thermal conductivity of voids (air)	in W/(m°C)	0,03	
Thermal conductivity of conventional ICA	in W/(m°C)	2	[53]
Thermal conductivity of DIEMAT ICA	in W/(m°C)	60	[53]
Thermal conductivity of Sn96Ag3.5Cu	in W/(m°C)	57	[54]
$R_{th,if}$ at bonding area of 2.31 cm^2	in °C/W	8.7×10^{-2}	[52]

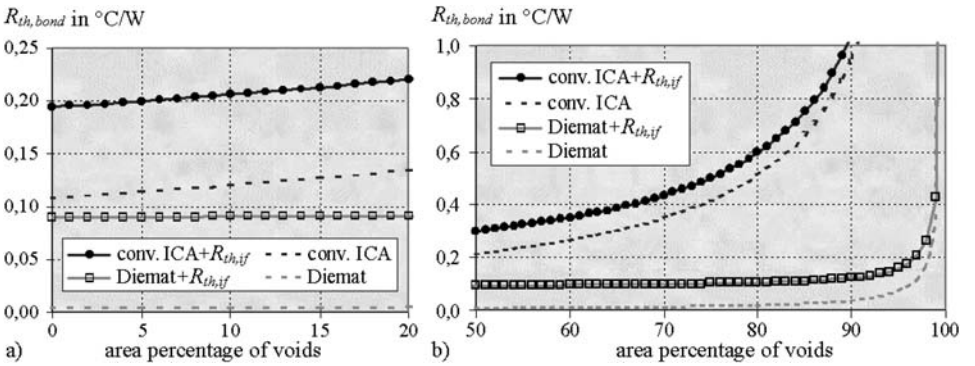


FIGURE 21.35. Bond line thermal resistance $R_{th,bond}$ versus percentage of gas enclosures; (a) at acceptable quantities of enclosure, (b) in an unacceptable range.

Using the values as listed in Table 21.3 Equation (21.9) is evaluated for different ICA types.

At low percentages of enclosures an almost linear increase of $R_{th,bond}$ can be observed in bond lines with conventional ICAs. No measurable thermal effect of enclosures exists in case of a high intrinsic thermal conductivity (DIEMAT or lead-free solder). In order to recognize the meaning of the interface thermal resistance for both ICA types $R_{th,bond}$

is also calculated as a reference with $R_{th,if}$ set to zero (dashed curves), respectively. In both cases the interface thermal resistance clearly predominates the thermal resistance increase due to enclosures [Figure 21.35(a)]. It is surprising that only if voids are unacceptably large from mechanical point of view they become thermally important [Figure 21.35(b)].

For $R_{th,j-c}$ of silicon devices in TO-247 packages frequently values below 1°C/W are provided from manufacturers. It should be noted that in cases of GaAs power devices due to the lower thermal conductivity of the substrate compared to Si the discussed phenomena are particularly important [53].

21.4.2. ICA for Interconnecting Parts with Dissimilar Thermal Expansion Coefficient

For many types of radiation sensors sealed housings are needed consisting of a metal cap and a base plate which carries the sensor substrate. The base plate including the bonding process for mounting the sensor substrate is dispensable if all functions of the base plate can be fulfilled by the sensor substrate itself. The cap can either be soldered to the substrate using a high-temperature solder in order to allow a second soldering process (e.g., for its attachment to a PCB) or an adhesive can be applied. The latter is favored whenever a low processing temperature is essential or when thermomechanical requirements cannot be fulfilled with a soldered joint. In the following example an alumina substrate with a wiring structure produced by thick film technology was required to be bonded to a cap of aluminum whereby soft soldering had to be excluded from the possible joining techniques, since due to very dissimilar coefficients of thermal expansion of base plate and cap thermal cycling caused microcracks in the solder layer. The basic structure of the sensor is schematically illustrated in Figure 21.36.

Adhesives with a high glass transition temperature T_g are reported as to be generally more susceptible to the formation of cracks than those with a low T_g which is understood as a different ability to stress relaxation [55]. However, due to the thermoelastic behavior of adhesives there are e.g., the possibility of directly attaching of 15 mm × 15 mm large silicon chips on FR-4 PCB without damage during temperature cycling as demonstrated in the DACTEL project [56].

A further essential parameter besides the thermoelastic behavior is the thickness of the ICA layer between the mating surfaces. The thicker the layer is, at a given thermal expansion mismatch, the lower is the shear plane angle and the inclination to formation of cracks. This is well known from reliability investigations of solder joints e.g., between chip components and PCBs where it was found that the number of thermal cycles until failure increases significantly with rising stand-off height [57].

Besides aspects of long-term reliability the housing of the sensor has to fulfill thermal requirements. An accurate calibration of the sensor frequently needs a constant temperature distribution in the housing. Therefore, the heat loss has to be dissipated from the substrate through a clamping part at the front face of the cap at the lowest possible temperature drop. From this consideration a thin bonding layer is needed which is in contradiction with the mechanical requirements. An optimum can be found by selecting an adhesive with a high thermal conductivity. Figure 21.37 depicts the result of a comparison of two axial temperature profiles in the cap obtained by thermal simulation. This study shows that the temperature drop perpendicular to an ICA layer can be reduced significantly by increasing the filler content. Whereas the thermal conductivity of adhesives for conventional surface mounting applications even if offered as thermally conductive adhesives is rather low (ranging in the order of some W/mK, [58], represented by type 1 ICA in Figure 21.37) there are some

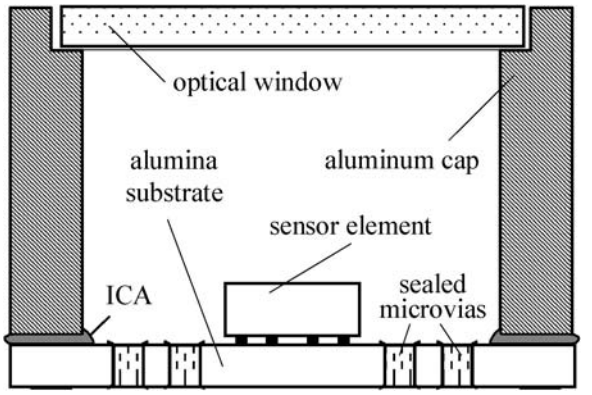


FIGURE 21.36. Schematic structure of the sensor housing.

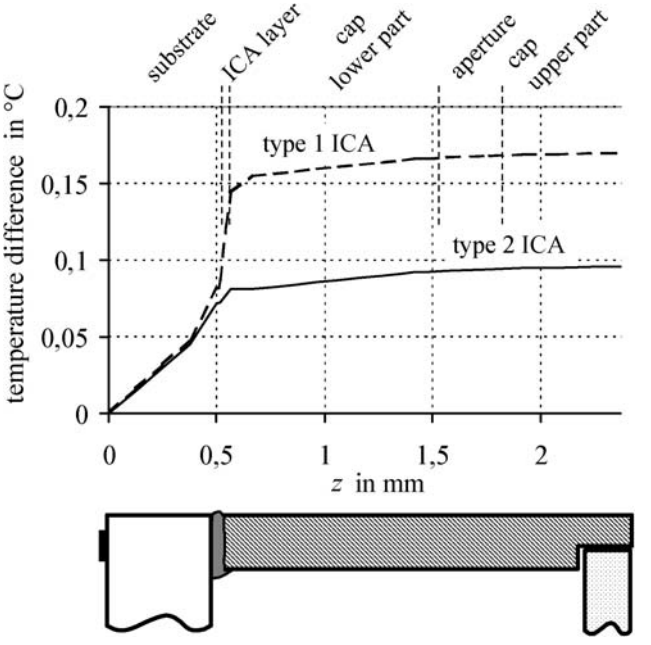


FIGURE 21.37. Influence of the thermal conductivity of the ICA layer on the temperature profile in axial direction (type 1 ICA: 3 W/mK, type 2 ICA: 30 W/mK).

types of ICA on the market with an extremely high silver content with a thermal conductivity of more than 50 W/mK (type 2 ICA, filler content: 93 weight % silver, 60 W/mK), allowing to achieve a negligible temperature drop across the ICA layer.

Due to the excellent thermoelastic behavior of the adhesive a lateral displacement of up to 20 μm due to different thermal expansions of substrate and cap could be handled with an adhesive layer thickness of only about 30 μm [59].

TABLE 21.4.
Comparison of number of process steps for SMD assembling either using soldering (left column) or adhesive bonding (right column).

...	...
Printing the solder (stencil) for SMDs and the transistor chip	Printing the adhesive (stencil) for all components
Placing SMDs and transistor	Placing all components
Reflow soldering	Curing the adhesive
Cleaning (flux removing)	–
Dispensing the adhesive for the remaining chips	–
Placing the chips	–
Curing	–
Wire bonding	Wire bonding
Applying glob top	Applying glob top
...	...

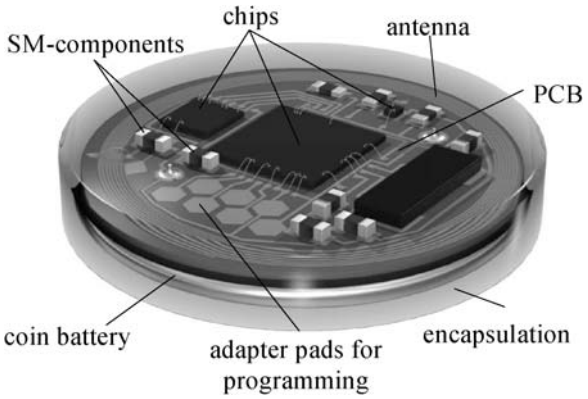


FIGURE 21.38. Schematic illustration of the temperature logger module.

21.4.3. ICA for Cost-Effective Assembling of Multichip Modules

Frequently multichip modules are the solution when a high degree of miniaturization is needed in an electronic assembly with complex functions. The application described here is a temperature logger system smaller than a one-Euro cent coin in diameter embedded in a removable brace to observe the wearing habits of the patient [60]. For this purpose the temperature is measured with a pn junction and recorded using a microcontroller which acts as a temperature logger and is capable of storing the thermal history from several months. This device is hermetically sealed using polymeric encapsulants. A wireless data exchange is performed using a radio frequency identification device (RFID). The electronic assembly contains a printed circuit board which is populated with SMDs in 0201 package (0.6 mm $\times$ 0.3 mm) and bare microprocessor and RFID chips (Figure 21.38).

The SMDs can be assembled by soldering or ICA bonding. The die components were not available for solder attachment as is frequently the case for small-batch production. Thus, wire bonding was necessary in the vicinity of the SMDs. In order to prevent the wire bonding pads from flux residues, the attachment of the SMDs by soldering would be related to additional process steps for protection and for cleaning of the bonding pads. The production process with the lowest possible number of process steps was found using ICA

for the attachment of chips and SMDs [59]. The manufacturing processes are compared in Table 21.4.

21.5. SUMMARY

As a historical background the development of ICA technology is briefly discussed. Already decades ago some principle advantages were recognized such as the low processing temperature and the possibility to use non-wettable and base metal surfaces for interconnections. For this reason, the attachment of semiconductors was the first application of ICAs in electronic packaging. In the meanwhile, much progress has been achieved in the quality and stability of ICA materials. Nevertheless, the ICA technology is not widely seen as an equivalent replacement of soldering, although it is a lead-free and a flux-free technology and both of these facts are important environmental aspects. One reason might be the experience with the former ICA technology concerning a moderate long-term reliability due to the inclination of oxide formation and galvanic corrosion in humid environment at elevated temperature. One goal of this contribution is to provide a deeper understanding of silver filled epoxy-based adhesives. For this purpose a model of an ICA joint which considers the particle alignment and distribution as well as the voltage and current distribution within the adhesive is discussed. Using this model the influence of parameters like filler content, particle arrangement, and particle size on the joint resistance is estimated and tendencies of changes during aging are clarified. The behavior of real ICA joints under accelerated aging condition and the results of quality assessment is also discussed. Finally, some practical examples demonstrate advantages and drawbacks in manifold fields of applications.

NOTATIONS AND DEFINITIONS

A	particle area
A_{ex}	excluded area
$\langle A_{ex} \rangle$	mean value of excluded area
A_{ex}/A	normalized excluded area
B	mean number of neighbored particles
B_c	mean critical number of neighbored particles
δ	angle between major axes of two particles
ϕ	portion of area filled with particles (filler content in area %)
ϕ_c	critical portion of area filled with particles (critical area filling factor)
$\Delta\phi$	deviation of area filling factor ϕ from critical value ϕ_c
γ	angle between major axes of a particles and the (horizontal) pad plane
κ_o	probability for a nonconductive (failing) interconnection between two particles
λ_{pp}	specific interface conductivity between two particles, see Equation (21.1).
$\lambda_{p\,pad}$	specific interface conductivity between particle and pads
p_c	percolation probability of 50%
p_p	percolation probability
R	normalized total resistance of interconnection (joint resistance)
$\langle R \rangle$	mean value of the normalized total interconnection resistance
R_{pp}	transition resistance between two particles

$R_{p\,pad}$	transition resistance between a particle and a pad
$\langle R_{tr} \rangle$	total transition resistance at both pad-particle interfaces, average value
σ	specific conductivity of particle
σ_{pad}	specific conductivity of pad
w_{tot}	total width of the modeled space (= length of the contact pad)
$w_{s,i}$	contact width at the ith particle (length of intersection line between two considered particles or between particle and the pad as e.g., shown in Figure 21.3)
$w_s^* = \sum_i^n w_{s,i}$	total effective width of contact pad (= active width of pad involved in current flow), n is the number of particles in direct contact with the respective pad and contributing to current flow
$\langle w_s^*/w_{tot} \rangle$	percentage of total effective contact width with respect to the pad length, average value
X	length relation between major and minor axis of elliptical particles (axis relation)

REFERENCES

1. R.L. Henry, Improvements in commutator brushes, British Patent #246.972, 1926.
2. H. Schuhmann, Conductive varnish, U.S. Patent #1,913,214, June 6, 1933.
3. H. Bienfait and W.L. Carolus van Zwet, Electrical conductor and method of making the same, U.S. Patent #2,018,343, Oct. 22, 1935.
4. H.J. Loftis, Molded electrically conductive body, U.S. Patent #2,361,220, Oct. 24, 1944.
5. N.H. Collings and R.J. Heaphy Beverton, Electrically conductive adhesive, U.S. Patent #2,444,034, June 29, 1948.
6. H. Wolfson and G. Elliott, Electrically conducting cements containing epoxy resins and silver, U.S. Patent #2,774,747, Dec. 18, 1956.
7. C. Mitchel and H. Berg, Use of conductive epoxies for die-attach, Int. Symp. Microelectron, 1976.
8. D. Gerber and W. Scheel, Kleben elektronischer Baugruppen, in W. Scheel, Ed., Baugruppentechnologie der Elektronik, Verlag Technik Berlin, Eugen G. Leuze Verlag, Saulgau, 1997, pp. 393–394.
9. H. Shirakawa, Nobel lecture: The discovery of polyacetylene film—the dawning of an era of conducting polymers, Reviews of Modern Physics, 73(July), pp. 713–718 (2001).
10. P.M. Raj, J. Liu, and P. Öhlckers, Fundamentals of packaging materials and processes, in R.R. Tummala, Ed., Microsystems Packaging, McGraw-Hill, 2001, p. 739.
11. I. Watanabe and K. Takemura, Anisotropic conductive adhesive films for flip-chip interconnection, in J. Liu, Ed., Conductive Adhesives for Electronics Packaging, Electrochemical Publications Ltd., ISBN 0 901150 37 1, 1999, pp. 256–271.
12. A. Aghzout, Über die Leitfähigkeit von elektrisch anisotropen Klebern und deren Anwendungen in der Mikroelektronik, Dissertation thesis, Vienna University of Technology, June, 2002.
13. K.-i. Shinotani, J. Malmodin, and R. Trankell, Fundamentals of microsystems design for environment, in R.R. Tummala, Ed., Microsystems Packaging, McGraw-Hill, 2001, p. 859.
14. G. Hanreich, K.-J. Wolter, and J. Nicolics, Rework of flip-chip polpulated PCBS by laser desoldering, in H. Hauser, Ed., Sensors & Packaging, ÖVE-Schriftenreihe, Nr. 35, 2003, pp. 283–289.
15. L.M. Yu and W.C. Qing, Solder joints design attribute to no solder bridge for fine pitch device, Fifth International Conference on Electronic Packaging Technology, ICEPT2003, Shanghai, China, 28–30 Oct., 2003, pp. 70–75.
16. S. Kang, R.S. Rai, and S. Purushothaman, Development of high conductivity lead (Pb)-free conductive adhesives, IEEE Transactions on Components Packaging, and Manufacturing Technologies, 21(4), pp. 18–22 (1998).
17. J.S. Hwang, Fine pitch soldering and solder paste, in J.H. Lau, Ed., Handbook of Fine Pitch Surface Mount Technology, Van Nostrand Reinhold, New York, 1993, pp. 81–133.
18. Directive 2002/95/EC of the European Parliament and of the Council of 27 January 2003 on the Restriction of the Use of Certain Hazardous Substances in Electrical and Electronic Equipment, Official Journal of the European Union L37, 13.02.2003, pp. 19–23.

19. R. Dudek, H. Berek, T. Fritsch, and B. Michel, Reliability investigations on conductive adhesive joints with emphasis on the mechanics of the conduction mechanism, IEEE Transactions on Components and Packaging Technologies, 3(3), pp. 462–469 (2000).
20. C.P. Wong and D. Lu, Recent advances in electrically conductive adhesives for electronics applications, 4th International Conf. on Adhesive Joining and Coating Technology in Electronics Manufacturing, Proceedings, 2000, pp. 121–128.
21. S. Ganesan and M. Pecht, Lead-free electronics—2004 Edition, CALCE EPSC Press, ISBN 0-9707174-7-4, 2004.
22. J.H. Lau, C.P. Wong, N.-C. Lee, and R.S.W. Lee, Electronics Manufacturing—with Lead-Free, Halogen-Free, and Conductive-Adhesive Materials, McGraw-Hill, ISBN 0071386246, July 2002.
23. J.E. Morris, F. Anderssohn, S. Kudtarkar, and E. Loos, Reliability studies of an isotropic electrically conductive adhesive, 1st Int. IEEE Conference on Polymers and Adhesives in Microelectronics and Photonics, October 21–24, Potsdam, Germany, 2001, pp. 61–69.
24. D. Lu and C.P. Wong, Development of conductive adhesives for solder replacement, IEEE Transactions on Components and Packaging Technologies, 23(4), pp. 620–626 (2000).
25. D. Lu, C.P. Wong, and Q.K. Tong, Mechanisms underlying the unstable joint resistance of conductive adhesives, IEEE Transactions on Components, Packaging, and Manufacturing Technology, Part C, 22(3), pp. 228–232 (1999).
26. M. Zwolinski, J. Hickman, H. Rubin, Y. Zaks, S. McCarthy, T. Hanlon, P. Arrowsmith, A. Chaudhuri, R. Hermansen, S. Lau, and D. Napp, Electrically conductive adhesives for surface mount solder replacement, IEEE Transactions on Components, Packaging, and Manufacturing Technology, Part C, 19(4), pp. 251–256 (1996).
27. E. Stanley, J.S. Andrade, S. Havlin, H.A. Makse, and B. Suki, Percolation Phenomena: a broad-brush introduction with some recent applications to porous media, liquid water, and city growth, Physica A: Statistical and Theoretical Physics, 266(1-4), pp. 5–16 (1999).
28. A. Bunde, Percolation in composites, Journal of Electroceramics, 5(2), pp. 81–92 (2000).
29. J.E. Morris, Conduction mechanisms and microstructure development in isotropic, electrically conductive adhesives, in J. Liu, Ed., Conductive Adhesives for Electronics Packaging, Electrochemical Publications Ltd., ISBN 0 901150 37 1, 1999, pp. 37–77.
30. M. Mündlein, J. Nicolics, and G. Hanreich, Accelerated curing of isotropically conductive adhesivemns by vapor phase heating, Third International IEEE Conference on Polymers and Adhesives in Microelectronics and Photonics (Polytronic 2003), Montreaux, Switzerland, October 21–23, 2003, pp. 101–105.
31. L. Li and J.E. Morris, Electrical conduction models for isotropically conductive adhesive joints, IEEE Transactions on Components, Packaging, and Manufacturing Technology, Part A, 20(1), pp. 3–8 (1997).
32. X.-G. Liang and X. Ji, Thermal conductance of randomly oriented composites of thin layers, International Journal of Heat and Mass Transfer, 43, pp. 3633–3640 (2000).
33. E. Sancaktar and B. Lan, Modeling filler volume fraction and film thickness effects on conductive adhesive resistivity, 4th IEEE International Conference on Polymers and Adhesives in Microelectronics and Photonics (Polytronic 2004), Portland, USA, 12–15 Sept., 2004, pp. 38–49.
34. I. Balberg, C.H. Anderson, S.Alexander, and N. Wagner, Excluded volume and its relation to the onset of percolation, Physical Review B, 30(7), pp. 3933–3943 (1984).
35. P.A. Cundall, A computer model for simulating progressive, large-scale movements in blocky rock systems, Proc. Symp. Int. Rock Mech. 2(8), Nancy, Vol. 2, 1971, pp. 129–136.
36. B. Su and J. Qu, A micromechanics model for electrical conduction in isotropically conductive adhesives during curing, Electronic Components and Technology 2004, Volume 2, 1–4 June 2004, pp. 1766–1771.
37. G.G.W. Mustoe, M. Nakagawa, X. Lin, and N. Iwamoto, Simulation of particle compaction for conductive adhesives using discrete element modeling, Electronic Components and Technology Conference 1999, 1–4 June, 1999, pp. 353–359.
38. A. Drory, I. Balberg, and B. Berkowitz, Application of the central-particle-potential approximation for percolation in interacting systems, Physical Review E, 52, pp. 4482–4494 (1995).
39. B. Berkowitz, Percolation theory and network modeling applications in soil, Physics: Surveys in Geophysics, 19(1), pp. 23–72 (1998).
40. S. Xu, D.A. Dillard, and J.G. Dillard, Environmental aging effects on the durability of electrically conductive adhesive joints, International Journal of Adhesion and Adhesives, 23(3), pp. 235–250 (2003).
41. E. Suganuma and M. Yamashita, High temperature degradation mechanism of conductive adhesive/Sn alloy interface, International Symposium on Advanced Packaging Materials: Processes, Properties and Interfaces, Proceedings, 2001, March 11–14, 2001, pp. 19–22.

42. H. Li, K.-S. Moon, Y. Li, L. Fan, J. Xu, and C.P. Wong, Reliability enhancement of electrically conductive adhesives in thermal shock environment, Electronic Components and Technology, ECTC 2004, Vol. 1, June 1–4, Las Vegas, USA, 2004, pp. 165–169.
43. F. Kriebel, Leitkleben—eine Alternative zum Löten in der Oberflächenmontagetechnik, VTE, (4)(Aug.), pp. 182–191 (1998).
44. H. Hauser, Kontaktwerkstoffe, Anhang D, in G. Fasching, Ed., Werkstoffe für die Elektrotechnik, Springer, 4th edition, 2005, pp. 517–520.
45. M. Mündlein, J. Nicolics, and J.E. Morris, Reliability investigations of isotropic conductive adhesives, 25th International Spring Seminar on Electronics Technology, ISSE 2002, Prague, Czech Republic, May 11–14, ISBN 0-7803-9824-6, 2002, pp. 329–333.
46. K. Gilleo, Evaluating polymer solders for lead-free assembly part I, Circuits Assembly, (January), pp. 50–56 (1994).
47. J.C. Jagt, P.J.M. Beris, and G.F.C.M. Lijten, Electrically conductive adhesives: a prospective alternative for SMD soldering? IEEE Transactions on Components, Packaging, and Manufacturing Technology, Part C, 18(2), pp. 292–298 (1995).
48. D. Klosterman, L. Li, and J.E. Morris, Materials characterization, conduction development, and curing effects on reliability of isotropically conductive adhesives, IEEE Transactions on Components, Packaging, and Manufacturing Technology, Part A, 21(1), pp. 23–31 (1998).
49. D. Lu, Q.K. Tong, and C.P. Wong, Conductivity mechanisms of isotropic conductive adhesives (ICAs), IEEE Transactions on Electronics Packaging Manufacturing, 22(3), pp. 223–227 (1999).
50. J.C. Jagt, Reliability of electrically conductive adhesive joints for surface mount applications: a summary of the state of the art, IEEE Transactions on Components, Packaging, and Manufacturing Technology, Part A, 21(2), pp. 215–225 (1998).
51. M.G. Perichaud, J.Y. Deletage, H. Fremont, Y. Danto, C. Faure, and M. Salagorty, Evaluation of conductive adhesives for industrial SMT assembles, 23rd IEEE/CPMT Electronics Manufacturing Technology Symposium, 1998, pp. 377–385.
52. R.C. Campbell, S.E. Smith, and R.L. Dietz, Measurements of adhesive bondline effective thermal conductivity and thermal resistance using the laser flash method, 15th Annual IEEE Semiconductor Thermal Measurement and Management Symposium, 9–11 March, 1999, pp. 83–97.
53. M. Mayer, J. Nicolics, G. Hanreich, and M. Mündlein, Thermal aspects of GaAs power FET attachment using isotropically conductive adhesive, 2nd Int. IEEE Conf. on Polymers and Adhesives in Microelectronics and Photonics, Polytronic 2002, Zalaegerszeg, Hungary, June 23–26, 2002, pp. 38–43.
54. http://www.npl.co.uk/ei/iag/leadfree/propertiespbf.html, Lead-Free Alloy Properties, National Physical Laboratory.
55. D.W. Swanson and L.R. Enlow, Stress effects of epoxy adhesives on ceramic substrates and magnetics, Microelectronics Reliability, 41(4), pp. 499–510 (2001).
56. J. Vanfleteren, A. Vervaet, and A.V. Calster, Highlights of the DACTEL project: development of adhesive chip technologies for dedicated electronic applications, Proc. of 23rd Int. Spring Seminar on Electronics Technology, ISSE 2000, Balatonfüred, HU, May 6–10, 2000, pp. 315–318.
57. M. Sumikawa, T. Sato, C. Yoshioka, and T. Nukii, Reliability of soldered joints in CSPs of various designs and mounting conditions, IEEE Transactions on Components and Packaging Technology, 24(2), pp. 293–299 (2001).
58. R. Prasad, Adhesives and its applications, in G.R. Blackwell, Ed., The Electronic Packaging Handbook, CRC Press, IEEE Press, 2000, pp. 10-1–10-28.
59. M. Mündlein, G. Hanreich, and J. Nicolics, Application of an ICA for the production of a radiation sensor, Proc. of 3rd Int. IEEE Conf. on Polymers and Adhesives in Microelectronics and Photonics, Polytronic 2003, Montreux, Schweiz, October 21–23, 2003, pp. 123–127.
60. M. Mündlein, J. Nicolics, and M. Brandl, Packaging concept for a miniaturized wirelessly interrogable temperature logger, Proceedings of the 27th Int. Spring Seminar on Electronics Technology, Sofia, Bulgaria, May 14–16, 2004, pp. 68–73.

22

Recent Advances of Conductive Adhesives: A Lead-Free Alternative in Electronic Packaging

Grace Y. Li and C.P. Wong

School of Materials Science and Engineering, Georgia Institute of Technology, 771 Ferst Drive, Atlanta, GA 30332-0245, USA

22.1. INTRODUCTION

Although the electronics industry has made considerable advances over the past few decades, the essential requirements of interconnects among all types of components in all electronic systems remain unchanged. The components need to be electrically connected for power, ground and signal transmissions, where tin/lead (Sn/Pb) solder alloy has been the *de facto* interconnect material in most areas of electronic packaging. Such interconnection technologies include pin through hole (PTH), surface mount technology (SMT), ball grid array (BGA) package, chip scale package (CSP), and flip chip technology. Tin-lead alloy solder has been a primary means of all electronic systems [1,2]. Figure 22.1 shows the use of Sn-Pb solder in typical PTH, SMT and BGA configurations.

There are increasing concerns with the use of tin-lead alloy solders. First, tin-lead solders contain lead, a material hazardous to human and environment. Each year, thousands tons of lead are manufactured into various products especially consumer electronic prod-

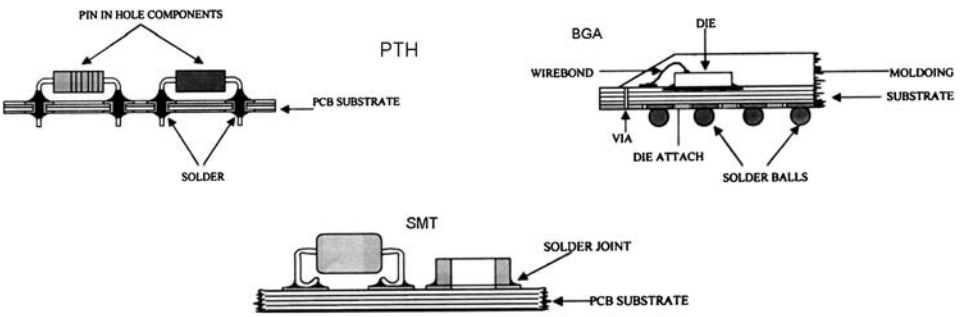


FIGURE 22.1. Schematic structures of PTH, SMT and BGA packages using solder interconnects.

ucts (e.g., cell phones, pagers, electronic toys, PDA, etc.) which tend to have a short (2–3 years) life cycle and millions of such lead-containing products simply end up in landfills. According to the latest report [U.S. Geological Survey, 2001], the total lead consumption by the U.S. industries was 52,400 metric tons in 2000. More than 10% of that (5,430 metric tons) was used to produce alloy solders. Worst of all, most of the electronic products have a very short service life. Recycling of lead-containing consumer electronic products has proven to be very difficult. Japan has banned the use of lead in all their electronic products in January of 2005, and European Union (E.U.) plans to ban all imports of lead-containing electronics in July, 2006. In response to the new legislation, most major electronic manufacturers all over the world, have stepped up their search for alternatives to lead-containing solders.

To date, these efforts have focused on two alternatives: lead-free solders and polymer based electrically conductive adhesives (ECAs) [3,4]. The most promising lead-free alloys contain tin as the primary element, because it melts at a relatively low temperature (232°C) and easily wets other metals. Depending on the applications, a number of lead-free solder alloys have found their way into commercial products [5,6]. However, most currently commercial lead-free solders, such as tin/silver (Sn/Ag), tin/silver/copper (Sn/Ag/Cu), have higher melting temperatures than that of tin-lead eutectic solder (183°C), typically at least 30°C higher. Therefore, the reflow temperature during electronic assembly must therefore be raised by 30°C to 40°C. This temperature increase reduces the integrity, reliability and functionality of printed wiring boards, components and other attachment, therefore severely limits the applicability of these metal alloys to organic/polymer packaged components and low-cost organic printed circuit boards. Although some low melting point alloys are available such as Sn/In (120°C), Sn/Bi (138°C), Sn/Zn/Ag/Al/Ga (189°C) [7], their material properties and processability in assembly are still of concern.

One the other hand, electrically conductive adhesives (ECA) can be processed at a much lower temperature. ECAs mainly consist of an organic/polymeric binder matrix and metal filler. These composite materials provide both physical adhesion and electrical conductivity. Compared to the solder technology, ECA offers numerous advantages, such as environmentally friendly, lower processing temperature, fewer processing steps (reducing processing cost), and especially, the fine pitch capability due to the availability of small-sized conductive fillers, especially for ACA [8–14].

ECA can be categorized with respect to conductive filler loading level into anisotropically conductive adhesives (ACA, with 3–5 μm sized conductive fillers, or sometimes in film form, ACF) and isotropically conductive adhesives (ICA, with 5–10 μm sized fillers), which are shown in Figure 22.2(a) and (b) [8]. The difference between ACA and ICA is based on the percolation theory (Figure 22.3). The percolation threshold depends on the shape and size of the fillers, but typically in the order of 15–25% volume fraction. For ICA, the loading level of conductive fillers is much more than the percolation threshold, providing electrical conductivity in x, y and z directions. For ACA, on the other hand, the loading level of conductive fillers is far below the percolation threshold, and the low volume loading is insufficient for inter-particle contact and prevents conductivity in the X–Y plane of the adhesive. Therefore, they provide a unidirectional electrical conductivity in the vertical or Z-axis. Both adhesive types are being adapted as interconnect materials for surface mount technology processes, such as chip on glass (COG), chip on flex (COF) and flip-chip bonding technologies in electronic packaging industries.

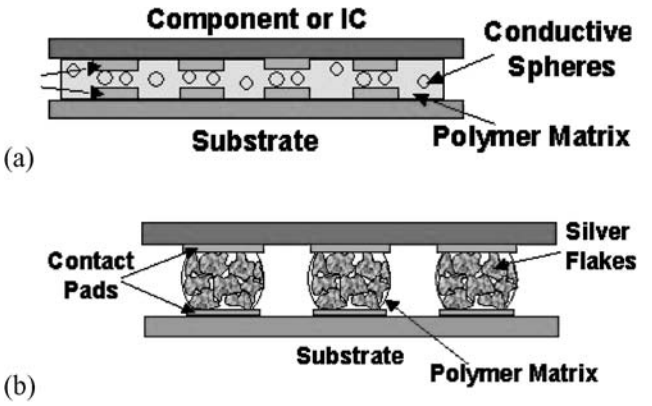


FIGURE 22.2. Schematic illustrations of (a) ACA and (b) ICA in flip-chip bonding.

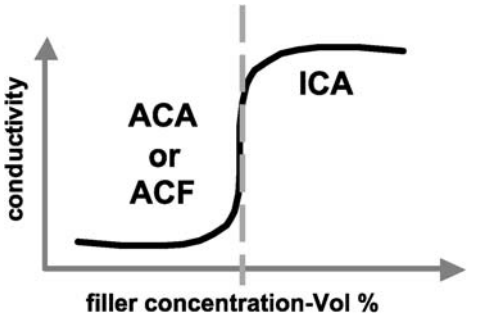


FIGURE 22.3. A typical percolation curve showing the abrupt drop in resistance at the percolation threshold.

22.2. ISOTROPIC CONDUCTIVE ADHESIVES (ICAs)

Isotropic conductive adhesives, also called as "polymer solder," are composites of polymer resin and conductive fillers. The adhesive matrix is used to form a mechanical bond at an interconnection. Both thermosetting and thermoplastic materials are used but thermoset epoxies are by far the most common binders due to the superior balanced properties, such as excellent adhesion, good chemical resistance and low cost. The conductive fillers provide the composite with electrical conductivity through tunneling and physical contact between the conductive particles. Silver flakes are the most commonly used conductive fillers for current commercial ICAs because of the unique properties of high conductivity of silver oxide and the maximum contact with flakes. ICAs have been used in the electronic packaging industry, primarily as die attach adhesives [15–17]. Recently, ICAs have been proposed as an alternative to tin/lead solders in surface mount technology (SMT) applications and a number of efforts have been conducted to improve the properties of ICAs in the past few years. Recent advances in material design and formulation have targeted to improved electrical conductivity, contact resistance stability, impact performance enhancement and metal migration control of ICAs. These will be reviewed in the following sections.

22.2.1. Improvement of Electrical Conductivity of ICAs

Polymer–conductive filler based electrically conductive adhesives (ECA) typically have lower electrical conductivity than Sn/Pb solders. To enhance the electrical conductivity of ECAs, various methods have been conducted and significant improvement of conductivity of ECA has been achieved.

22.2.1.1. Increase Polymer Matrix Shrinkage In general, ICA pastes exhibit insulative property before cure, but the conductivity increases dramatically after they are cured. ICAs achieve electrical conductivity during the polymer cure process caused by the shrinkage of polymer binder. Accordingly, ICAs with high cure shrinkage generally exhibit the best conductivity. Table 22.1 shows the relationship of shrinkage and conductivity for three different types of ECA, ECA1, ECA2 and ECA3 [18]. With increasing crosslinking density of ECA, the shrinkage of the polymer matrix increased, and subsequently, the obviously decreased resistivity of ECA was observed. Therefore, increasing the cure shrinkage of a polymer binder could improve electrical conductivity. For epoxy-based ICAs, a small amount of a multi-functional epoxy resin can be added into an ICA formulation to increase cross-linking density, shrinkage, and thus increase conductivity.

22.2.1.2. In situ Replacement of Lubricant on Ag Flakes An ICA is generally composed of a polymer binder and Ag flakes. There is a thin layer of organic lubricant on the Ag flake surface. This lubricant layer plays an important role for the performance of ICAs, including the dispersion of the Ag flakes in the adhesives and the rheology of the adhesive formulations [19–22]. This organic layer is a Ag salt formed between the Ag surface and the lubricant which typically is a fatty acid such as stearic acid [22,23]. This lubricant layer affects conductivity of an ICA because it is electrically insulating. To improve conductivity, the organic lubricant layer should be partially or fully removed or replaced. A suitable lubricant remover is a short chain dicarboxylic acid because of the strong affinity of carboxylic functional group ($-$COOH) to silver and stronger acidity of such short chain dicarboxylic acids. With the addition of only small amount of short chain single bond dicarboxylic acid, the conductivity of ICA can be improved significantly due to the easier electronic tunneling/transportation between Ag flakes and subsequently the intimate flake–flake contact [24].

22.2.1.3. Incorporation of Reducing Agent in Conductive Adhesives Silver flakes are by far the mostly used fillers for conductive adhesives due to the unique properties of high conductivity of silver oxide compared to other metal oxides, most of which are insulative. However, the conductivity of silver oxide is still inferior to metal itself. Therefore, incorporation of reducing agent would further improve the electrical conductivity of ICAs. Aldehydes were introduced into a typical ICA formulation and obviously improved con-

TABLE 22.1.
Relationship of shrinkage and electrical conductivity of ECAs.

Formulation	Cross-linking density (10^{-3} mol/cm^3)	Shrinkage (%)	Bulk resistivity (10^{-3} Ohm cm)
ECA1	4.50	2.98	3.0
ECA2	5.33	3.75	1.2
ECA3	5.85	4.33	0.58

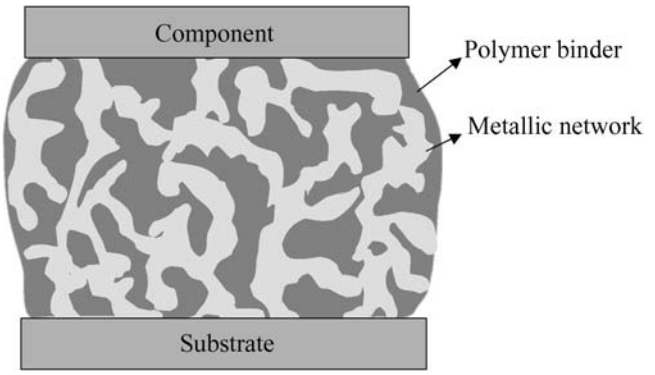


FIGURE 22.4. Diagram of transient liquid phase sintering conductive adhesives.

ductivity was achieved due to reaction between aldehydes and silver oxides that exist on the surface of metal fillers in ECAs during the curing process.

$$\mathrm{R{-}CHO} + \mathrm{Ag_2O} \rightarrow \mathrm{R{-}COOH} + 2\mathrm{Ag}. \tag{22.1}$$

The oxidation product of aldehydes, carboxylic acids, which are stronger acids and have shorter molecular length than stearic acid, can also partially replace or remove the stearic acid on Ag flakes and contribute to the improved electrical conductivity [25].

22.2.1.4. Low Temperature Transient Liquid Phase Fillers Another approach for improving electrical conductivity is to incorporate transient liquid-phase sintering metallic fillers in ICA formulations. The filler used is a mixture of a high-melting-point metal powder (such as Cu) and a low-melting-point alloy powder (such as Sn-Pb). The low-melting-alloy filler melts when its melting point is achieved during the cure of the polymer matrix. The liquid phase dissolves the high melting point particles. The liquid exists only for a short period of time and then forms an alloy and solidifies. The electrical conduction is established through a plurality of metallurgical connections formed in situ from these two powders in a polymer binder (Figure 22.4). The polymer binder with acid functional ingredient fluxes both the metal powders and the metals to be joined and facilitates the transient liquid bonding of the powders to form a stable metallurgical network for electrical conduction, and also forms an interpenetrating polymer network providing adhesion. High electrical conductivity can be achieved using this method [26,27]. One critical limitation of this technology is that the numbers of combinations of low melt and high melt fillers are limited. Only certain combinations of two metallic fillers which are mutually soluble exist to form this type of metallurgical interconnections.

22.2.2. Stabilization of Contact Resistance on Non-Noble Metal Finishes

22.2.2.1. Mechanism of Unstable Contact Resistance Contact resistance between an ICA and non-noble metal finished components increases dramatically during an elevated temperature and humidity aging. The National Center of Manufacturing and Science (NCMS) set a criterion for solder replacement conductive adhesives. The criterion is if the contact resistance shift after 500-hour at 85°C/85%RH aging is less than 20%, then the contact resistance is defined as "stable" [28]. Previous study in our laboratory has shown that galvanic corrosion rather than simple oxidation at the interface between an ICA and non-noble

metal is the dominant mechanism for the unstable contact resistance. Under the aging conditions (for example, 85°C/85%RH), the non-noble acts as an anode, and is oxidized by losing electrons, and then turns into metal ion ($M - ne^- = M^{n+}$). The noble metal acts as a cathode, and its reaction generally is $2H_2O + O_2 + 4e^- = 4OH^-$. Then M^{n+} combines with OH^- to form a metal hydroxide or metal oxide. After corrosion, a layer of metal hydroxide or metal oxide is formed at the interface. Because this layer is electrically insulating, the contact resistance increases dramatically [29]. A galvanic corrosion process has several characteristics: (1) happens only under wet conditions, (2) an electrolyte must be present, and (3) oxygen generally accelerates the process. Based on the mechanism of unstable contact resistance of ECA, several methods could be applied to stabilize the contact resistance.

22.2.2.2. Oxygen Scavengers Since oxygen accelerates galvanic corrosion, oxygen scavengers could be added into ECAs to slow down the corrosion rate. When ambient oxygen molecules diffuse through the polymer binder, they react with the oxygen scavenger and are consumed. However, when the oxygen scavenger within the ECA is depleted, then oxygen can again diffuse onto the interface and accelerate the corrosion process. Therefore, oxygen scavengers can only delay the galvanic corrosion process, but does not solve the corrosion problem completely. Some commonly used oxygen scavengers include sulfates (such as Na_2SO_4), hydrazine ($H_2N{=}NH_2$), carbohydrazide ($H_2N{-}NH{-}CO{-}NH{-}NH_2$), diethylhydroxylamine ($(C_2H_5)_2N{-}OH$), and hydroquinone ($HO{-}C_6H_4{-}OH$).

22.2.2.3. Corrosion Inhibitors Another method of preventing galvanic corrosion is to introduce organic corrosion inhibitors into ICA formulations [29–32]. In general, organic corrosion inhibitors act as a barrier layer between the metal and environment when forming a film over the metal surfaces [33]. Some chelating compounds are especially effective in preventing metal corrosion [34]. Chelating agents are organic molecules with at least two polar functional groups capable of ring closure with a metal cation. The functional groups may either be basic groups, such as $-NH_2$, which can form bonds by electron donation, or acidic groups, such as $-COOH$, which can coordinate after loss of the proton. Most organic corrosion inhibitors react with the epoxy resin at a specific temperature. Therefore, if an ICA is epoxy-based, the corrosion inhibitors must not react with the epoxy resin during curing which would cause them to be consumed and lose their effectiveness. Appropriate selection of corrosion inhibitors could be very effective in protecting the metal finishes from corrosion. However, the effectiveness of the corrosion inhibitors is highly dependent on the types of contact materials. As examples, Figure 22.5 and Figure 22.6 show the effect of different corrosion inhibitors on contact resistance of ICA on Sn/Pb and Sn surface, respectively. It can be seen that by using suitable corrosion inhibitors, stabilized contact resistance was achieved but the effective corrosion inhibitors were different for different metal finishes [31,32].

22.2.2.4. Sacrificial Anode To improve the contact resistance stability, applying a sacrificial anode is another method. For galvanic corrosion of ECAs during aging, the larger the different in potential, the faster the corrosion develops. Also, the self-corrosion rates of both metals will change: the comparably active metal (the anode) corrodes faster while the other (the cathode) corrodes slower than they would do without contact. When applying a third and more active metal (or alloy) on a dissimilar metal couple in electrical contact, the comparably active metal can be protected from galvanic corrosion by oxidation of the third metal. This corrosion control is very important in reliability issues of the conductive adhesive joints. The addition of low corrosion potential individual metals, metal mixtures

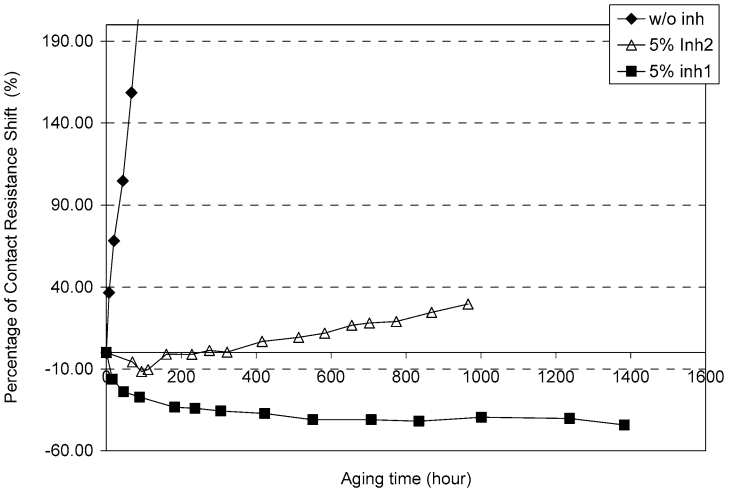


FIGURE 22.5. Shifts of contact resistance of conductive adhesives on Sn/Pb surface with and without corrosion inhibitors.

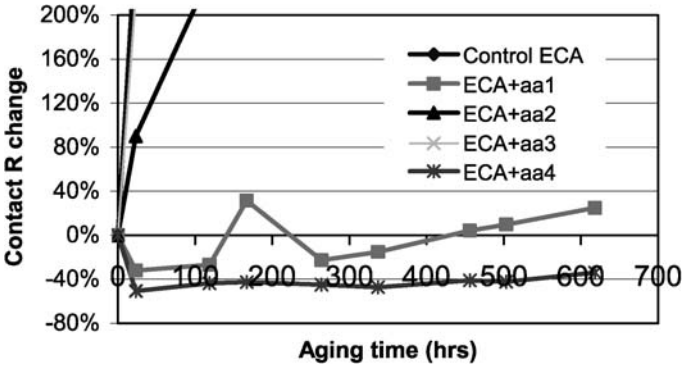


FIGURE 22.6. Shifts of contact resistance of conductive adhesives on Sn surface with and without corrosion inhibitors.

or metal alloys greatly reduces the electrode potential of ECAs, or in other words, narrows down the potential gap between the ECA and the metal finishes. Thus, these sacrificial anode materials act as an anode in this configuration and they are corroded first instead of the metal finishes, resulting in protecting the surfaces at the cathode. Comparing different types of sacrificial anodes, metal alloys are more effective than individual metals or metal mixtures due to the less easily oxidization characteristics for alloys [35].

22.2.2.5. Oxide-Penetrating Particles Another approach of improving contact resistance stability during aging is to incorporate some electrically conductive particles, which have sharp edges, into the ICA formulations. The particle is called oxide-penetrating filler. Force must be provided to drive the oxide-penetrating particles through oxide layer and hold them against the adherend materials. This can be accomplished by employing polymer binders that show high shrinkage when cured (Figure 22.7) [36]. This concept is used in polymer-

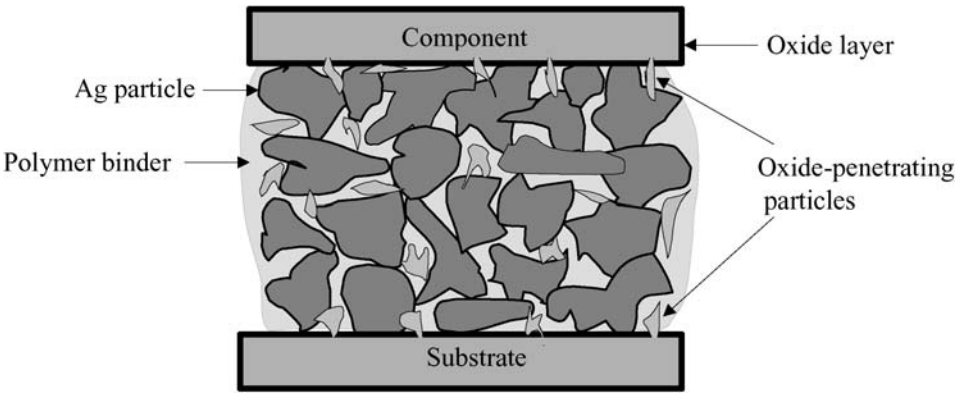


FIGURE 22.7. A Joint connected with an ICA containing oxide-penetrating particles and silver powders.

solder which has good contact resistance stability with standard surface-mounted devices (SMDs) on both solder-coated and bare circuit boards.

22.2.3. Silver Migration Control of ICA

Silver electromigration has long been a reliability concern in the electronic industry. The electromigration is an electrochemical process whereby, silver, in contact with an insulating material, in a humid environment and under an applied electric field, leaves its initial location in ionic form and deposits at another location [37]. Such migration may lead to a reduction in electrical spacing or cause a short circuit between interconnections. The migration process begins when a thin continuous film of water forms on an insulating material between oppositely charged electrodes. When a potential is applied across the electrodes, a chemical reaction takes place at the positively biased electrode where positive metal ions are formed. These ions, through ionic conduction, migrate toward the negatively charged cathode and over time, they accumulate to form metallic dendrites. As the dendrite growth increases, a reduction of electrical spacing occurs. Eventually, the silver growth reaches the anode and creates a metal bridge between the electrodes, resulting in an electrical short circuit [38]. The rates of silver migration is increased by: (1) an increase in the applied potential; (2) an increase in the time of the applied potentials; (3) an increase in the level of relative humidity; (4) an increase in the presence of ionic and hydroscopic contaminants on the surface of the substrate; and (5) a decrease in the distance between electrodes of the opposite polarity [37].

Several methods have been used to control the silver migration. The methods include: (1) alloying the silver with an anodically stable metal such as palladium [38] or platinum [39] or even tin [40]; (2) use of hydrophobic coating over the PWB to shield its surface from humidity and ionic contamination [41]; (3) plating of silver with metals such as tin, nickel or gold; and (4) coating the substrate with polymer [42]. Some other research is still going on for the migration prevention and high reliability ICAs.

22.2.4. Improvement of Reliability in Thermal Shock Environment

Besides the contact resistance increase due to galvanic corrosion, the poor thermal cycling (TC) performance of the ECA joints has been another serious issue for board level

interconnects. It has been found rather difficult to control the TC failures of the ECA joints. Generally, the failure of the electrical interconnection during the TC test could be caused by many factors such as coefficient of the thermal expansion (CTE) mismatch between the IC component chip/the interconnection materials/the substrates, elastic modulus difference of these components, adhesion strength of the interconnect materials on the IC chip and the substrate, the mechanical properties of the IC chips, the glass transition or the softening point of the ECA materials, moisture uptake in the interface and the bulk ECAs, the surface or interface property change and so forth. Especially, the thermal stress in the ECA joints generated by a huge temperature difference during the TC and the interfacial delamination due to the adhesion degradation could be the critical reasons. In this aspect, a feasible solution to the TC failure problem is to introduce flexible molecules into the epoxy resin matrix. By releasing the thermal stress with the flexible molecules, the thermomechanical stresses could be dramatically reduced and the ECA/component joint interfaces could keep intact through the thermal cycling test.

22.2.5. Improvement of Impact Performance of ICA

Impact performance is a critical property of solder replacement ICAs. Effort has been continued in developing ICAs that have better impact strength and will pass the drop test, a standard test used to evaluate the impact strength of ICAs.

Nano-sized metal particles were used in ICAs to improve the electrical conduction and mechanical strength. Using nano-sized particles, agglomerates are formed due to surface tension effect [43]. Another approach is simply to decrease the filler loading to improve the impact strength [44]. However, such a process reduces the electrical properties of the conductive adhesives. A recent development was reported where conductive adhesives were developed using resins of low modulus so that this class of conductive adhesives could absorb the impact energy developed during the drop [45]. However, the electrical properties of these materials were not mentioned in the paper. Conformal coating of the surface-mounted devices was used to improve mechanical strength. It was demonstrated that conformal coating could improve the impact strength of conductive adhesives joints [46]. Furthermore, elastomer-modified epoxy resins were also used to enhance the damping properties and the impact performance of ICAs [47].

22.3. ANISOTROPIC CONDUCTIVE ADHESIVES (ACAs)/ANISOTROPIC CONDUCTIVE FILM (ACF)

Anisotropic conductive adhesives (ACAs) or anisotropic conductive films (ACFs) provide unidirectional electrical conductivity in the vertical or Z-axis. This directional conductivity is achieved by using a relatively low volume loading of conductive filler (5 to 20 volume percent) [48–50]. The low volume loading is insufficient for inter-particle contact and prevents conductivity in the X–Y plane of the adhesive. The Z-axis adhesive, in film or paste form, is interposed between the surfaces to be connected. Heat and pressure are simultaneously applied to this stack-up until the particles bridge the two conductor surfaces on the two adherends. Figure 22.8 shows the configuration of a component and a substrate bonded with ACF. Once the electrical continuity is produced, the polymer binder is hardened by thermally initiated chemical reaction (for thermosets) or by cooling (for thermoplastics). The hardened dielectric polymer matrix holds the two components together and helps maintain the pressure contact between component surfaces and conductive

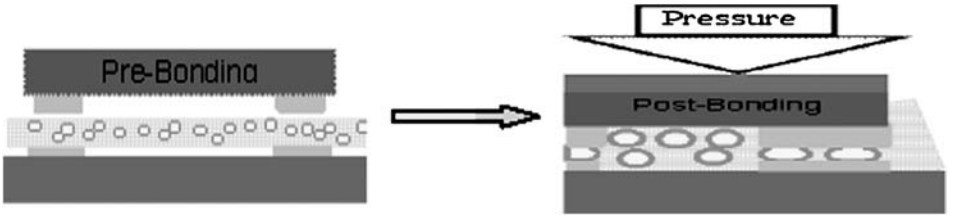


FIGURE 22.8. Schematic diagram of accomplishment of z-direction interconnects using ACF.

particles. Because of the anisotropy, ACA/ACF may be deposited over the entire contact region, greatly facilitating materials application. Also, an ultra fine pitch interconnection (<0.04 mm) could be achieved easily. The fine pitch capability of ACA/ACF would be limited by the particle size of the conductive filler, which can be a few microns or a few nanometers in diameter.

22.3.1. Materials

When designing materials to achieve fine pitch interconnections, several important variables must be considered and are application dependent. These variables include adhesive characteristics as well as particle type.

Two basic types of adhesives are available: thermosetting and thermoplastic materials. Thermoplastic adhesives are rigid materials at temperatures below the glass transition temperature (T_g) of a polymer. Above the T_g, polymers exhibit flow characteristics. When using this type of material, assembly temperatures must exceed the T_g to achieve good adhesion. The principal advantage of the thermoplastic adhesives is the relative ease with which the interconnection can be disassembled for repair operations. Thermosetting adhesives, such as epoxies and silicones, form a three-dimensional cross-linked structure when cured under specific conditions. Curing techniques include: heat, UV light, and added catalyst. As a result of this irreversible cure reaction, the initial uncross-linked material is transformed into a rigid solid. The curing reaction is not reversible. This fact may hinder disassembly and interconnection repair. The ability to maintain strength at high temperature and the deformation of robust adhesive bonds are the principal advantages of these materials. For the selection of the adhesive, the robust bonds should be formed to all surfaces involved in the interconnection. Numerous materials surfaces can be found in the interconnection region including: SiO_2, Si_3N_4, polyester, polyimide, FR-4, glass, gold, copper, and aluminum. Adhesion to these surfaces must be preserved after standard tests such as temperature-humidity-bias aging and temperature cycling. Some surfaces may require chemical treatments to achieve good adhesion. In addition, the adhesive must not contain ionic impurities that would degrade electrical performance of the interconnections.

The materials used as conductive particles must also be carefully selected. Silver offers moderate cost, high electrical conductivity, high current carrying ability, and low chemical reactivity, but problems with electromigration may occur. Nickel is a lower cost alternative, but corrosion of nickel surfaces has been found during accelerated aging tests. The material that offers the bests properties is gold; however, costs may be prohibitive for large-volume applications. Plated polymeric particles may offer the best combination of properties at moderate cost. Some ACA materials use solder particles to ensure electrical contacts with high reliability by creating a metallurgical bond.

22.3.2. Application of ACA/ACF in Flip Chip

Conventional flip-chip assembly involves two main steps: solder reflow and application of underfill, which is an organic adhesive placed between the IC chip and the substrate to improve mechanical reliability. In recent years, ACA flip-chip technology has been employed in many applications where flip chips are bonded to rigid chip carriers [51]. This includes bare chip assembly of ASICs in transistor radios, personal digital assistants (PDAs), sensor chip in digital cameras, and memory chip in lap-top computers. In all the applications, the common feature is that ACA flip-chip technology is used to assembly bare chips where the pitch is extremely fine, normally less than 120 μm. For those fine applications, it is apparently use of ACA flip chip instead of soldering is more cost effective.

ACA flip-chip bonding exhibits better reliability on flexible chip carriers because the ability of flex provides compliance to relieve stresses. For example, the internal stress generated during resin curing can be absorbed by the deformation of the chip carrier. ACA joint stress analysis conducted by Wu et al. indicated that the residual stress is larger on rigid substrates than on flexible substrates after bonding [52].

22.3.3. Improvement of Electrical Properties of ACAs

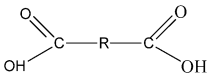
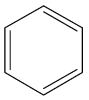
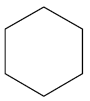
ACA flip chip technology has been employed in many applications because of the primary advantages such as fine pitch capability, lead-free, low processing temperature, absence of flux residue, and generally lower cost. Also, ACA flip chip technology does not require an additional underfilling process because the ACA resin acts as an underfill. However, ACAs have lower electrical current capability because of the restricted contact area and poor interfacial bonding of the ACAs and metal bond pad.

22.3.3.1. Self-Assembled Monolayer (SAM) In order to enhance the electrical performance of ACA materials, a class of chemicals that form a self-assembled monolayer (SAM) of conjugated molecular nano wires is introduced. These SAM molecules adhere to the metal surface and form physico-chemical bonds, which allow electrons to flow, as such, it reduces electrical resistance and enables a high current flow. The unique electrical properties of SAM originate from their linear chain structure and electron delocalization along the conjugated chain. Preliminary data have also shown that the current density of a SAM molecular wire such as 1,4-dithiol benzene can be up to 1×10^8 Å/cm^2, so the potential for conductivity improvement and high current carrying capability are significant.

An important consideration when examining the advantages of SAM compounds pertains to the affinity of SAM compounds to specific metal surfaces. Table 22.2 gives the examples of SAM molecules preferred for maximum interactions with specific metal finishes; although only molecules with symmetrical functionalities for both head and tail groups are shown, molecules and derivatives with different head and tail functional groups are possible for interfaces concerning different metal surfaces.

22.3.3.2. Electrical Properties Improvement of ACA with SAM Different SAM compounds, dicarboxylic acid and dithiol have been introduced into ACA joints for silver-filled and gold-filled ACAs, respectively [53,54]. Due to the strong affinity between those SAM compounds and metal fillers and metal bond pads, the physic-chemical bonding was introduced on the interface between the metal particles and between the ACAs/metal bond pads. The physico-chemical bonding could allow electrons to flow freely, as such, the molecules can reduce electrical resistance and achieve a high current density in the interface. For

TABLE 22.2.
Potential SAM interfacial modifier for different metal finishes.

Formula	Compound	Metal finish
H−S−R−S−H	Dithiol	Au, Ag, Sn, Zn
N≡C−R−C≡N	Dicyanide	Cu, Ni, Au
O=C=N−R−N=C=O	Diisocyanate	Pt, Pd, Rh, Ru
	Dicarboxylate	Fe, Co, Ni, Al, Ag
	Imidazole and derivative	Cu
R =  ,  , −C=C−C=C−, −C−C−C−, etc.		

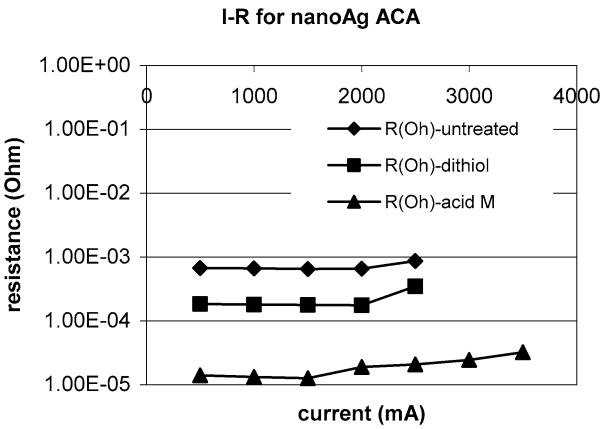


FIGURE 22.9. Electrical properties of nano Ag filled ACA with SAM.

dithiol incorporated ACA with micron-sized gold fillers, significantly lower joint resistance and higher maximum allowable current (highest current applied without inducing joint failure) was achieved for low temperature curable ACA (<100°C). For high curing temperature ACA (150°C), the improvement is not as significant as low curing temperature ACAs, due to the partial degradation of SAM compounds at the relatively high temperature. However, when dicarboxylic acid or dithiol was introduced into the interface of nano-silver filled ACA, significantly improved electrical properties could be achieved at high temperature curable ACA, suggesting de-bonding/degradation of SAM did not occur on silver nanoparticles at the curing temperature (150°C) (Figure 22.9). The enhanced bonding could be attribute to the larger surface area and higher surface energy of nano particles, which enabled the SAMs to be more readily coated and relatively thermally stable on the metal surfaces [54].

22.3.4. Thermal Conductivity of ACA

For the ACA interconnect joints to deliver high current, not only a low electrical resistance, but also a high thermal conductivity of the interconnect materials is required. The higher thermal conductivity can help dissipate heat more effectively from adhesive joints generated at high current. Therefore, a higher thermal conductivity could contribute to an improved current carrying capability. Reported shown that with the addition of high thermal conductivity fillers (e.g., SiC) into the ACA formulation, the higher thermal conductivity could be achieved, which also rendered a high current carrying capability [55] (Figure 22.10). By improving the interface properties of ACA joints with additives, a higher thermal conductivity could also be achieved.

22.4. FUTURE ADVANCES OF ECAs

There have been a significant number of studies to improve ECA technology. However, several critical issues associated with this technology must also be addressed before it can be used as a complete replacement for soldering technology.

22.4.1. Electrical Characteristics

The electrical conductivity must be further improved since their electrical conductivity is typically one or two orders of magnitudes lower than solders. Silver-filled ECAs achieve their electrical conductivity through a physical contact among the Ag flaks. Due to the high particle-particle contact resistance, ECAs exhibit a relatively higher bulk resistivity. Although some work has been conducted to improve electrical conductivity by in situ replacing or removing the lubricants on Ag flakes and helping electrons tunneling in ECA, additional work is still necessary to achieve the desired electrical conductivity.

Also, the electrical performance of ECAs on different metal surfaces under high temperature/humidity/bias conditions is not completely satisfactory. Although the mechanism for the reliability failure under aging environment has been well under-stood and some effective methods have been developed to stabilize the contact resistance on some metal finishes, further study is still needed to completely under-stand the reason for the reliability improvement. At the same time, it is necessary to look for new materials and methods to stabilize the contact resistance on different metal surfaces.

22.4.2. High Frequency Compatibility

The number of high-frequency applications and utilizations are increasing rap-idly, thus it is important to characterize the cross-linking between particles, coupling with semiconductor devices and other fundamental behavior of ECAs under high-frequency conditions. It is also necessary to maximize the current carrying capability of ECAs (especially ACAs) for high performance microprocessors and at high frequency range for wireless communication products, especially after exposure to various environment tests.

22.4.3. Reliability

It is necessary to understand the effect of the chip carrier material on ECA join reliability. This is a key issue before ECA technology is widely utilized in manufacturing (i.e.,

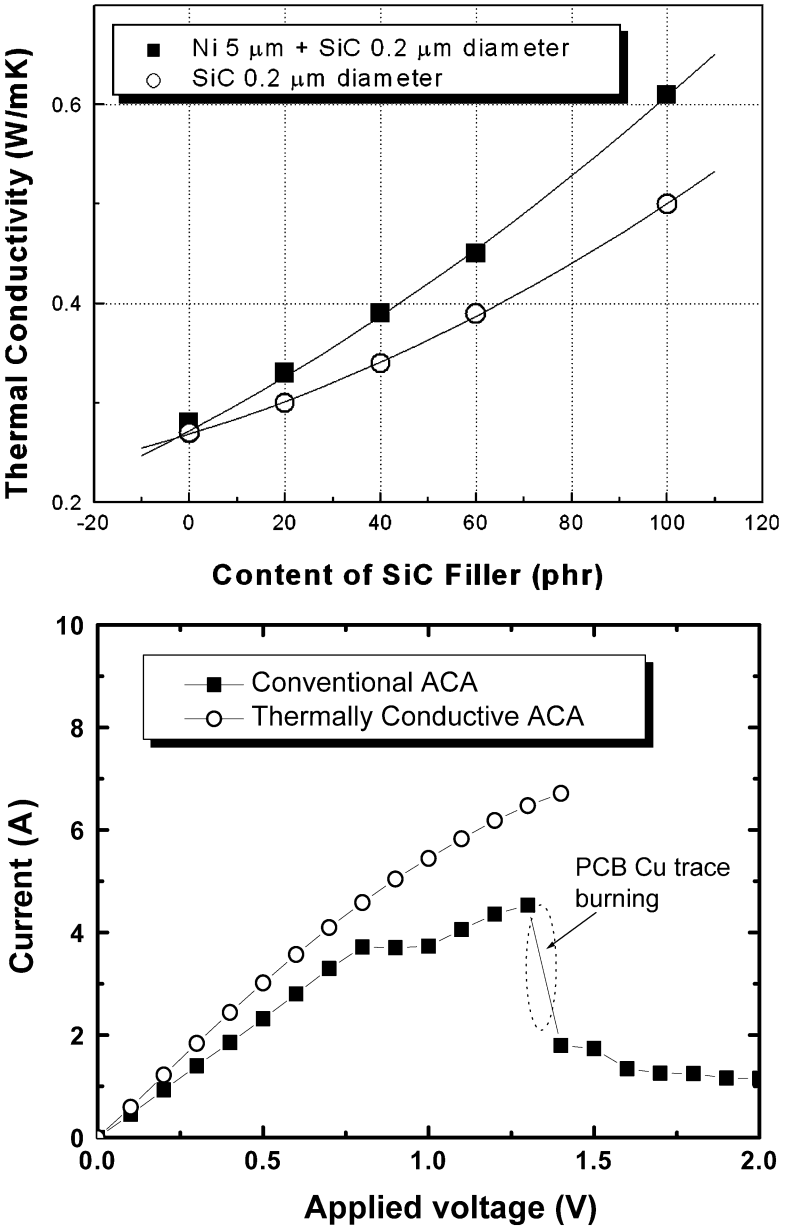


FIGURE 22.10. Thermal conductivity and I–V measurement of ACA with high thermal conductivity fillers.

in high-volume and low-cost applications). It is also necessary to establish failure rate prediction models for ECA joints for a wide variety of field conditions. It is essential to gain full understanding on effects of high current and high power on ECA joints, degradation and stress relaxation of polymeric matrices; and the effects of temperature, humidity, and other environments on matrix materials and the effects of fillers.

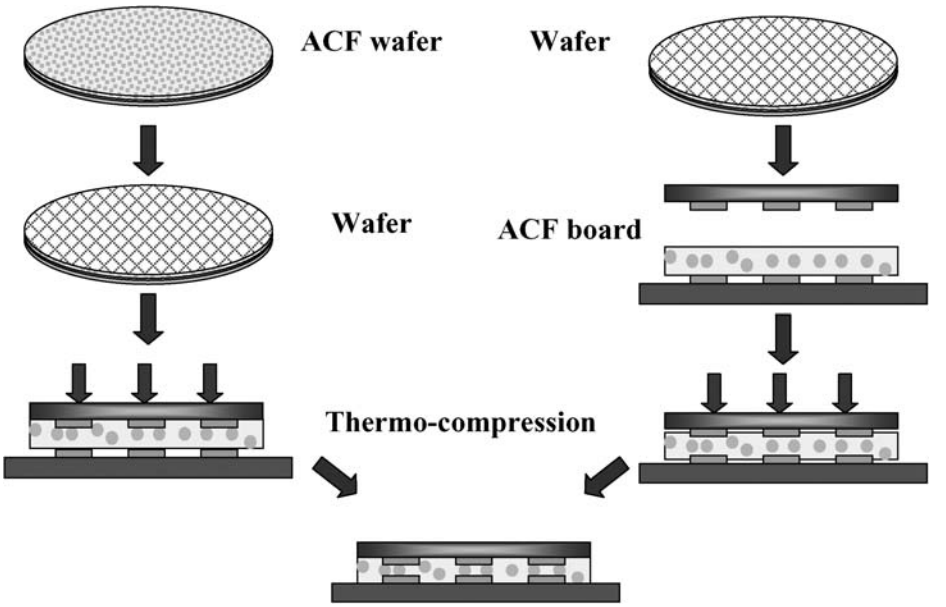


FIGURE 22.11. Schematic illustration of wafer level ACF process.

Silver-filled ICAs have a possibility for silver migration which causes electrical shorts especially in fine-pitch applications. Palladium treated silver filler exhibited much improved anti-migration characteristics compared to standard silver-filled ICAs. However, the preparation of the Pd-coated silver particles is expensive. Some low cost approaches must be developed. More comprehensive understanding of the fatigue resistance of ICA joints is required. Activities in this area have been limited and nonconclusive.

22.4.4. ECAs with Nano-filler for Wafer Level Application

For the next generation low cost, high efficiency ACA/ACF assembly, wafer level application might be a solution. Figure 22.11 shows the process of nano wafer level ACF (WLACF). Instead of using ACF in the assembly process, the ACF is applied on the wafer level before dicing. This eliminates the dispensing step in the component level and thus makes the ACF interconnect compatible with standard thermo-compression assembly process. The WLACF provides a lead-free and fine-pitch-capable interconnect, as well as a wafer level package to protect the wafer during test and burn-in. The cost of packaging can be dramatically reduced because it avoids the solder bumping process, combines interconnect and encapsulation, and enables wafer level test and burn-in. After the chip assembly, the ACF layer also acts as an underfill to redistribute the thermo-mechanical stress generated from the CTE mismatch between the chip and the substrate.

REFERENCES

1. M. Abet and G. Selvaduray, Lead-free solders in microelectronics, Materials Science & Engineering, 27, pp. 95–141 (2000).

2. J. Lau, C.P. Wong, N.C. Lee, and S.W.R. Lee, Electronics Manufacturing: With Lead-Free, Halogen-Free, and Conductive-Adhesive Materials, McGraw-Hill, New York, NY, 2002.
3. K.J. Puttlitz and K.A. Stalter, Eds., Handbook of Lead-free Solder Technology for Microelectronic Assemblies, Marcel Dekker, Inc., New York, NY, 2004, Chap. 1.
4. J. Lau, C.P. Wong, J. Prince, and W. Nakayama, Electronic Packaging; Design, Materials, Process and Reliability, McGraw-Hill, New York, 1998, Chap. 5, p. 393.
5. A.Z. Miric and A. Grusd, Lead-free alloys, Soldering and Surface Mount Technology, 10(1), p. 19 (1998).
6. J.S. Hwang, Solder materials, Surface Mount Technology, 18, p. 46 (2000).
7. K. Chen and K. Linz, Effects of gallium on wettability, microstructures and mechanical properties of the Sn-Zn-Ag-Ga and Sn-Zn-Ag-AI-Ga solder alloys, Int'l Symposium on Electronic Materials and Packaging, p. 49 (2002).
8. J. Liu, Ed., Conductive adhesives for Electronics Packaging, Electrochemical Publications Ltd. Isle of Man, British Isles, 1999.
9. E.P. Wood and K.L. Nimmo, In search of new lead-free electronic solders, J. Electron. Mater., 23(8), pp. 709–713 (1994).
10. E.R. Monsalve, Lead ingestion hazard in hand soldering environments, Proceedings of the 8th Annual SolderingTechnology and Product Assurance Seminar, Naval Weapons Center, China Lake, CA, February 1984.
11. Y. Li, K. Moon, and C.P. Wong, Electronics without lead, Science, 308(June 3), pp. 1419–1420 (2005).
12. S. Jin, D.R. Frear, and J.W. Morris, Jr., Foreword, J. Electron. Mater., 23(8), pp. 709–713 (1994).
13. Environmental Protection Agency, National Air Quality and Emission Trend Report, 1989, EPA-450/4-91-003, Research Triangle Park, NC, 1991.
14. E. Perrot, Electronic packaging for the 21st century, Advanced Packaging, (July-August) (1995).
15. J. Greaves, Evaluation of solder alternatives for surface mount technology, Proceedings of Nepcon West Technical Program, 1993, pp. 1479–1488.
16. M.A. Gaynes, R.H. Lewis, R.F. Saraf, and J.M. Roldan, Evaluation of contact resistance for isotropic conductive adhesives, IEEE Transactions on Components, Packaging, and Manufacturing Technology, Part B, 18(2), pp. 299–304 (1995).
17. G. Nguyen, J. Williams, F. Gibson, and T. Winster, Electrical reliability of conductive adhesives for surface mount applications, Proceedings of International Electronic Packaging Conference, 1993, pp. 479–486.
18. D. Lu, Q.K. Tong, and C.P. Wong, Conductivity mechanisms of isotropic conductive adhesives (ICAs), IEEE Transactions on Components, Packaging, and Manufacturing Technology, Part C, 22(3), p. 223 (1999).
19. L. Smith-Vargo, Adhesives that posses a science all their own, Electronic Packaging & Production, (August), pp. 48–49 (1986).
20. E.M. Jost and K. McNeilly, Silver flake production and optimization for use in conductive polymers, Proceedings of ISHM, 1987, pp. 548–553.
21. S.M. Pandiri, The behavior of silver flakes in conductive epoxy adhesives, Adhesives Age, Oct. 1987, pp. 31–35.
22. D. Lu, Q.K. Tong, and C.P. Wong, A study of lubricants on silver flakes for microelectronics conductive adhesives, IEEE Transactions on Components, Packaging and Manufacturing Technology, Part A, 22(3), pp. 365–371 (1999).
23. D. Lu, Q. Tong, and C.P. Wong, A fundamental study on silver flakes for conductive adhesives, 1998 International Symposium on Advanced Packaging Materials, Braselton, GA, 1998, pp. 256–260.
24. Y. Li, K. Moon, H. Li, and C.P. Wong, Conductivity improvement of isotropic conductive adhesives with short-chain dicarboxylic acids, Proceedings of 54th IEEE Electronic Components and Technology Conference, Las Vegas, Nevada, June 1–4, 2004, p. 1959.
25. Y. Li, A. Whitman, K. Moon, and C.P. Wong, High performance electrically conductive adhesives (ECAs) modified with novel aldehydes, Proceedings of 55th IEEE Electronic Components and Technology Conference, Lake Buena Vista, Florida, May 31–June 3, 2005, pp. 1648–1652.
26. C. Gallagher, G. Matijasevic, and J.F. Maguire, Transient liquid phase sintering conductive adhesives as solder replacement, Proceedings of 47th Electronic Components and Technology Conference, San Jose, CA, May 1997, p. 554.
27. J.W. Roman and T.W. Eagar, Low stress die attach by low temperature transient liquid phase bonding, Proceedings of ISHM, San Francisco, CA, Oct. 1992, p. 52.
28. M. Zwolinski, J. Hickman, H. Rubon, and Y. Zaks, Electrically conductive adhesives for surface mount solder replacement, Proceedings of the 2nd International Conference on Adhesive Joining & Coating Technology in Electronics Manufacturing, Stockholm, Sweden, June 1996, p. 333.
29. D. Lu, Q.K. Tong, and C.P. Wong, Mechanisms underlying the unstable contact resistance of conductive adhesives, IEEE Transactions on Components, Packaging, and Manufacturing Technology, Part C, 22(3), pp. 228–232 (1999).

30. Q.K. Tong, G. Fredrickson, R. Kuder, and D. Lu, Conductive adhesives with superior impact resistance and stable contact resistance, Proceedings of the 49th Electronic Components and Technology Conference, San Diego, CA, May 1999, p. 347.
31. D. Lu and C.P. Wong, Novel conductive adhesives for surface mount applications, Journal of Applied Polymer Science, 74, p. 399 (1999).
32. Y. Li, K. Moon, and C.P. Wong, Development of conductive adhesives with novel corrosion inhibitors for stabilizing contact resistance on non-noble lead-free finishes, Proceedings of 55th IEEE Electronic Components and Technology Conference, Lake Buena Vista, Florida, May 31–June 3, 2005, pp. 1462–1467.
33. H. Leidheiser, Jr., Mechanism of corrosion inhibition with special attention to inhibitors in organic coatings, Journal of Coatings Technology, 53(678), p. 29 (1981).
34. G. Trabanelli, Corrosion inhibitors, in F. Mansfeld, Ed., Corrosion Mechanisms, Marcel Dekker, Inc., New York, NY, 1987, p. 119.
35. H. Li, K. Moon, and C.P. Wong, A novel approach to stabilize contact resistance of electrically conductive adhesives on lead-free alloy surfaces, Journal of Electronic Materials, 33(2), p. 106 (2004).
36. D. Durand, D. Vieau, A.L. Chu, and T.S. Weiu, Electrically conductive cement containing agglomerates, flake and powder metal fillers, U.S. Patent 5,180,523, Nov. 1989.
37. G. Davies and J. Sandstrom, How to live with silver tarnishing/silver migration, Circuits Mfg., (Oct.), pp. 56–62 (1976).
38. G. Harsanyi and G. Ripka, Electrochemical migration in thick-film IC-S, Electrocomp. Science And Tech., 11, pp. 281–290 (1985).
39. R. Wassink, Notes on the effects of metalization of surface mounted components on soldering, Hybrid Circ., (13), pp. 9–13 (1987).
40. Y. Shirai, M. Komagata, and K. Suzuki, Non-migration conductive adhesives, 1st International IEEE Conference on Polymers and Adhesives in Microelectronics and Photonics, 21–24 Oct., 2001, pp. 79–83.
41. A. Der Marderosian, The Electrochemical Migration of Metal, Ratheon Co. Equipment Division, Equipment Development Laboratories, pp. 134–141.
42. H. Schonhorn and L.H. Sharpe, Prevention of surface mass migration by a polymeric surface coating, U.S. Patent #4,377,619, 1983.
43. S. Kotthaus, R. Haug, H. Schafer, and O.D. Hennemann, Proceedings of 1st IEEE International Symposium on Polymeric Electronics Packaging, Norrkoping, Sweden, Oct. 26–30, 1997, pp. 64–69.
44. S. Macathy, Proceedings of Surface Mount International, San Jose, August 27–31, 1995, pp. 562–567.
45. S.A. Vona and Q.K. Tong, Surface mount conductive adhesives with superior impact resistance, Proceedings of 4th International Symposium and Exhibition on Advanced Packaging Materials, Processes, Properties and Interfaces, Braselton, GA, March 14–16, 1998, pp. 261–267.
46. J. Liu and B. Weman, Modification of processes and design rules to achieve high reliable conductive adhesive joints for surface mount technology, Proceedings of the 2nd International Symposium on Electronics Packaging Technology, Dec. 9–12, 1996, pp. 313–319.
47. D. Lu and C.P. Wong, Conductive adhesives for solder replacement in electronics packaging, Proceedings—International Symposium on Advanced Packaging Materials: Processes, Properties and Interfaces, Braselton, GA, United States, Mar. 6–8, 2000, pp. 24–31.
48. A.O. Ogunjimi, O. Boyle, D.C. Whalley, and D.J. Williams, A review of the impact of conductive adhesive technology on interconnection, Journal of Electronics Manufacturing, 2, p. 109 (1992).
49. P.G. Harris, Conductive adhesives: a critical review of progress to date, Soldering & Surface Mount Technology, (20), p. 19 (1995).
50. K. Gilleo, Assembly with conductive adhesives, Soldering & Surface Mount Technology, (19), p. 12 (1995).
51. J. Liu, ACA bonding technology for low cost electronics packaging applications-current status and remaining challenges, Proceedings of 4th International Conference on Adhesive Joining and Coating Technology in Electronics manufacturing, Helsinki, Finland, June 2000, pp. 1–15.
52. C.M.L. Wu, J. Liu, and N.H. Yeung, Reliability of ACF in flip chip with various bump height, Proceedings of 4th International Conference on Adhesive Joining and Coating Technology in Electronics manufacturing, Helsinki, Finland, June 2000, pp. 101–106.
53. Y. Li, K. Moon, and C.P. Wong, Adherence of self-assembled monolayers on gold and their effects for high performance anisotropic conductive adhesives, Journal of Electronic Materials, 34(3), pp. 266–271 (2005).
54. Y. Li and C.P. Wong, Nano-Ag filled anisotropic conductive adhesives (ACA) with self-assembled monolayer for high performance fine pitch interconnect, Proceedings of 55th IEEE Electronic Components and Technology Conference, Lake Buena Vista, Florida, May 31–June 3, 2005, pp. 1147–1154.
55. M. Yim, J. Hwang, J. Kim, H. Kim, W. Kwon, K.W. Jang, and K.W. Paik, Anisotropic conductive adhesives with enhanced thermal conductivity for flip chip applications, Proceedings of 54th IEEE Electronic Components and Technology Conference, Las Vegas, Nevada, June 1–4, 2004, p. 159.

23

Die Attach Quality Testing by Structure Function Evaluation

Márta Rencz[a], Vladimir Székely[b], and Bernard Courtois[c]

[a]*MicReD Ltd., Hungary*
[b]*Budapest University of Technology and Economics, Hungary*
[c]*TIMA Laboratory Grenoble, France*

Abstract In this chapter simulation and measurement experiments prove that the structure function evaluation of the thermal transient testing is capable to locate die attach failure(s), even in case of stacked die packages. Both the strength and the location of the die attach failure may be determined with the methodology of a fast thermal transient measurement and the subsequent computer evaluation. The paper presents first the theoretical background of the method. After this, application on single die packages is presented. In the rest of the paper first simulation experiments show the feasibility of locating die attach problem in stacked die structures with the presented algorithm and a large number of measured experiments prove that the methodology is applicable also in practice. At the end of the chapter, in the evaluation of the methodology the special advantage of the method, that normally it does not require any additional circuit elements on any of the possibly-stacked-dies is also presented.

NOMENCLATURE

A (m^2)	area, surface
c (Ws/Km^3)	volumetric heat capacity
C, C_{th} (Ws/K)	thermal capacitance
K ($W^2\,s/K^2$)	value of the differential structure function
R, R_{th} (K/W)	thermal resistance
Z (K/W)	thermal impedance

Greek symbols

λ (W/mK)	thermal conductivity
τ (s)	time constant
ω (1/s)	angular frequency

Subscripts

Σ cumulative values

23.1. INTRODUCTION

The reliability of packaged electronics strongly depends on the quality of the die attachment. Any void in it or a small delamination may cause instant temperature increase in the die, leading sooner or later to failure in the operation. Since die attach problems are built in time bombs, they have to be eliminated by any means. They usually can not be detected by the electronic test, only in special, very bad cases. They can be detected by steady state thermal measurements, but these measurements are on one hand time consuming on the other hand they can not locate the actual problem. In case of stacked die structures, that will dominate the packages in the near future the die attach quality control problem is multiplied: the structure contains as many die attach layers as the number of chips, which are very difficult to test, resulting in pronounced reliability concerns [1,2].

It has been presented in earlier papers that the structure functions [3], obtainable from the thermal transient measurements are applicable to indicate the die attach failures of single die packages [4]. In the paper of [4] die attach quality measurements on experimental packages of ST-Microelectronics were presented. It was demonstrated that the presence of die attach voids in different forms and locations influences the thermal resistance of the glue layer such, that the voids in the die attach material can be conveniently detected with the help of fast thermal transient measurements and the subsequent evaluation. A special advantage of the applied structure function evaluation methodology is that it delivers not only the value but also the location of the increased thermal resistance in the heat flow path.

Because of the extremely small thermal resistance value of the stacked dies it was not straightforward that the method can be used to uncover the locations of die attach failures of stacked dies as well. And even if the method is applicable up to a certain number of dies, it has to be experimented what are the limits in the resolution of the method. It had to be also investigated whether the methodology is applicable for stacked die structures of different die sizes as well; this was done in [5].

In this chapter we first give a short summary of the structure function evaluation methodology, presenting how it can be used to detect the thermal properties of various layers in the heat flow path. In the rest of the chapter we demonstrate the applicability of the method with simulation and measurement experiments.

23.2. THEORETICAL BACKGROUND

The simplest representation of a thermal system is presented in Figure 23.1: it is characterized by one thermal resistance and one thermal capacitance.

If we apply a P power step to this model the temperature will rise at its port according to

$$T(t) = P R_{th}(1 - \exp(-t/\tau)), \tag{23.1}$$

where $\tau = R_{th}C_{th}$ is the time constant of the system.

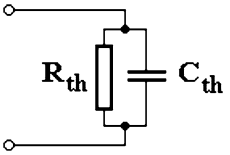


FIGURE 23.1. The simplest representation of a thermal system.

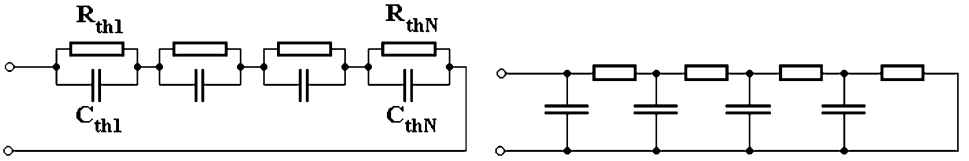


FIGURE 23.2. Foster (a) and Cauer (b) type representation of physical structures with finite number of time constants.

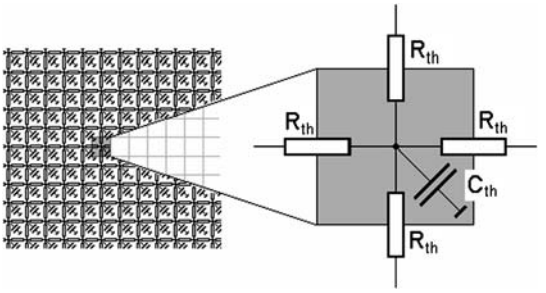


FIGURE 23.3. Real physical structures have infinite number of time constants.

Real physical structures have usually more time constants, in this case the temperature response function is the sum of the appropriate exponential functions, as

$$T(t) = P \sum_{i=1}^{N} R_{thi}(1 - \exp(-t/\tau_i)). \tag{23.2}$$

To such response functions two different model networks may be ordered. These are not independent, one may be calculated from the other: a Foster and a Cauer type network, see Figure 23.2.

Thermal systems are to be represented by Cauer networks, since the thermal capacitances are always connected to the ground, but since every Cauer network has its Foster equivalent, the Foster approximation of thermal networks exists as well.

Real physical structures, however, are represented with an infinite number of time constants, since any infinitesimal cube of a matter shows a certain thermal resistance and capacitance value, see Figure 23.3 resulting in as many number of time constants in the response function as the number of its capacities.

These time constants may be represented only with their density function, see Figure 23.4.

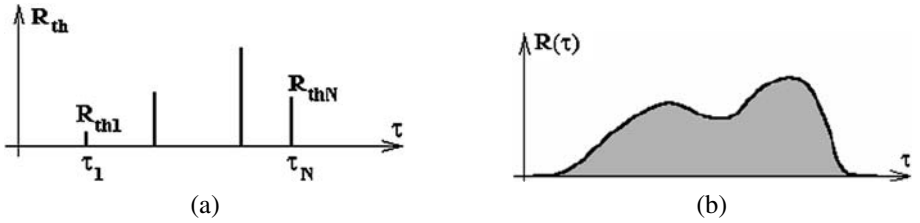


FIGURE 23.4. Time constants in a lumped element system (a) and in an infinite distributed parameter system (b). In the latter case the time constants form a continuous curve, the time constant density function.

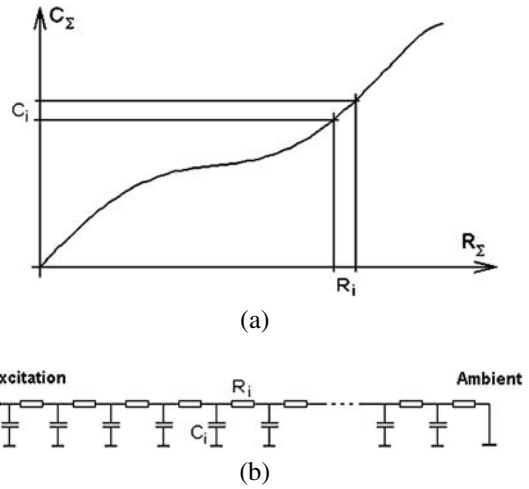


FIGURE 23.5. The cumulative structure function and the related Cauer equivalent circuit. (© 2002 IEEE SEMI-THERM Proceedings, "Determining partial thermal resistances with transient measurements and using the method to detect die attach discontinuities," M. Rencz, V. Szekely, A. Morelli, C. Villa, Fig. 1. © 2004 IEEE SEMI-THERM Proceedings, "Structure function evaluation of stacked dies," M. Rencz, V. Székely, Fig. 1.)

Knowing the time constant density of a system, an arbitrarily well approximating Foster equivalent circuit may be constructed by approximating the $R(\tau)$ function by infinitesimally narrow boxes with the height of $R(\tau)$, determining for each a parallel RC pair in the Foster chain. Finding the Cauer equivalent of this network by textbook transformations, we obtain a true physical equivalent of the heat transport in thermal systems. From this circuit we can draw up the so-called *Cumulative structure function* or PROTONOTARIOS-WING function [6], that gives the sum of the thermal capacitances C_Σ versus the sum of the thermal resistances R_Σ of the thermal system, measured from the point of excitation toward the ambient, see Figure 23.5.

On this monotonously increasing function, the plateaus represent the new materials with relatively low thermal capacity value and their widths give the related thermal resistances. The constant cross section intervals represent fields with heat propagation over linearly increasing areas.

The structure functions are calculated by direct mathematical transformations from the heating or cooling curves [3]. These latter curves may be obtained either from measurements or from the simulations of the detailed structural model of the heat flow path. In both cases a step function powering has to be applied on the structure, and the resulting increase

(or decrease, in case of switching off) in the temperature at the location of the powering has to be measured, following the action of switching on or off has to be recorded.

The differential structure function is defined as the derivative of the cumulative thermal capacitance with respect to the cumulative thermal resistance, by

$$K(R_\Sigma) = \frac{dC_\Sigma}{dR_\Sigma}. \tag{23.3}$$

Considering a dx wide slice of a single matter of cross section A, we can calculate this value. Since for this case $dC_\Sigma = cAdx$, and the resistance is $dR_\Sigma = dx/\lambda A$, where c is the volumetric heat capacitance, λ is the thermal conductivity and A is the cross sectional area of the heat flow, the K value of the differential structure function is

$$K(R_\Sigma) = \frac{cAdx}{dx/\lambda A} = c\lambda A^2. \tag{23.4}$$

This value is proportional to the c and λ material parameters, and to the square of the cross sectional area of the heat flow, consequently it is related to the structure of the system. In other words: this function provides a map of the square of the cross section area of heat current-flow as a function of the cumulative resistance. In these functions the local peaks and valleys indicate reaching new materials in the heat flow path, and their distance on the horizontal axis gives the partial thermal resistances between them. More precisely, the peaks point usually to the middle of any new region where both the areas, perpendicular to the heat flow and the material are uniform. After identifying the peaks that refer to the chip and the chip carrier their distance on the horizontal axis can be read from the diagram. This value gives the die attach thermal resistance.

The method can be used for detecting die attach failures as follows:

1. The structure function of a good sample has to be created. This can be obtained by mathematical transformation [3] of the measured or simulated thermal transient curves.
2. The structure function of the sample package that we wish to evaluate has to be created similarly.
3. The structure function of the known good device has to be compared to the structure function of the device under test. The shift in the appropriate points in the structure functions gives the increased thermal resistance of the layer in question.

In the next section it is presented how to apply the methodology to single die packages. In the following sections it is first demonstrated with simulation experiments that the method is applicable also for quality testing of stacked die structures, after this we show how to use the method in practice. The presented measured results on real stacked die packages prove the applicability of the method for the detection of voids and delamination in the different die attach and soldering layers of stacked die packages as well.

23.3. DETECTING VOIDS IN THE DIE ATTACH OF SINGLE DIE PACKAGES

Fast detection of die attach problems is a crucial question in semiconductor packaging. In order to verify the applicability of the structure function evaluation for the detection and the analysis of die attach discontinuities in single die packages ST Microelectronics has fabricated experimental packages [4] with well defined die attach problems, see Figure 23.6. In these experimental samples voids were produced in the corners, at the sides and in the centers of the packages and a large number of experiments were carried out on a large number of samples with each type of failures to assure reliable results. Out of all these experimental data only a few are presented to demonstrate our findings: we selected to present here the curves that represent the averages of the centre void or corner void samples.

In Figure 23.7 average differential structure functions of samples with corner voids and central voids are presented together with the differential structure function of a reference device. These curves were obtained by carrying out thermal transient measurements on the structures, and calculating the structure functions, as described in [6]. We can see in this figure that the first peaks, representing the chips, are at the same R_{th} value, but on the right hand side of these, the R_{th} "distance" of the next peak is increased in case of corner voids, and highly increased in case of central voids.

Similarly large differences can be noticed in the dominant time constants of the samples in the time constant density functions. Figure 23.8 presents the time constant density functions of the samples, where the main time constants of the selected average samples are those which have the highest intensity values. It is observable in the figure that a dominant time constant appears with the value of 0.01 sec in case of the samples with central voids. Similar, though less dominant increase can be noticed in case of the samples with corner voids as well.

Figure 23.7 and Figure 23.8 demonstrate in one hand, that the highest increase in the R_{th} value is resulted in the cases when the voids are centrally located in the package. This result shows that in this package structure the heat flow path is such that the thermal properties of the package very strongly depend on the quality of the heat removal path at the middle of the structure. On the other hand it is also suggested from these figures that the heating curves of the samples have to be significantly different already after 0.01 sec. This can be in fact noticed on the measured Z_{th} curves of Figure 23.9 as well: at $t = 0.01$ sec

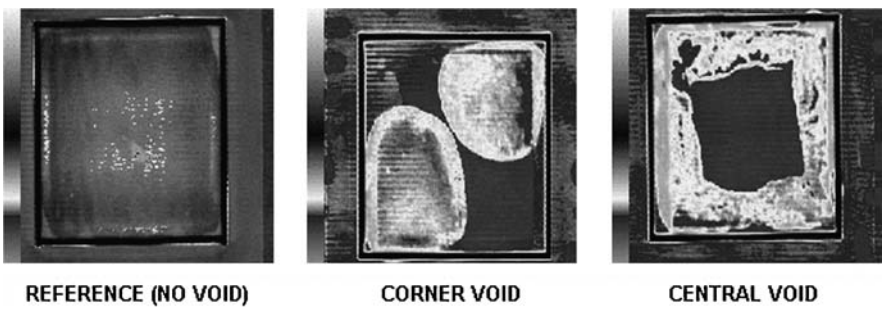


FIGURE 23.6. Experimental package samples with die attach voids prepared to verify the accuracy of the detection method (acoustic microscopic images). (© 2002 IEEE SEMI-THERM Proceedings, "Determining partial thermal resistances with transient measurements and using the method to detect die attach discontinuities," M. Rencz, V. Szekely, A. Morelli, C. Villa, Fig. 6.)

of the measurement the heating curves of the samples with die attach failures demonstrate significantly higher Z_{th} values than the perfect ones.

This last finding is extremely important from the point of view of in-production-line testing, because it means, that the method is applicable also for the in-line detection of die attach failures. In-production-line testing has to be done in milliseconds, and can not wait for the generation of the structure functions. For the generation of the structure function the whole thermal transient curve is needed, reaching the steady state, which may take minutes, or even longer, depending on the structure. To analyze however the origin of the Z_{th} increase, that is, for failure analysis the structure function method has to be used.

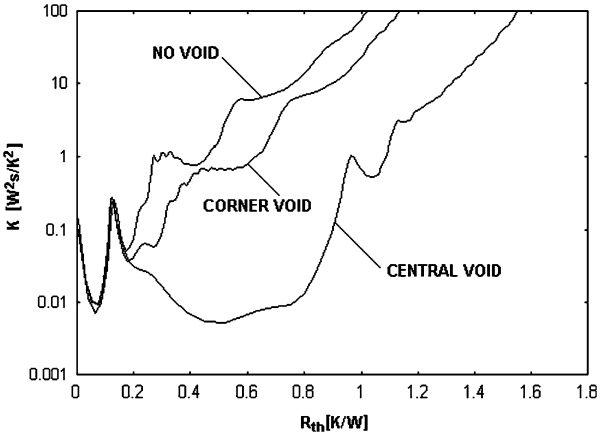


FIGURE 23.7. Differential structure functions of the experimental samples. The samples with central voids show highly increased R_{th} between the chip and the package. (© 2002 IEEE SEMI-THERM Proceedings, "Determining partial thermal resistances with transient measurements and using the method to detect die attach discontinuities," M. Rencz, V. Szekely, A. Morelli, C. Villa, Fig. 7.)

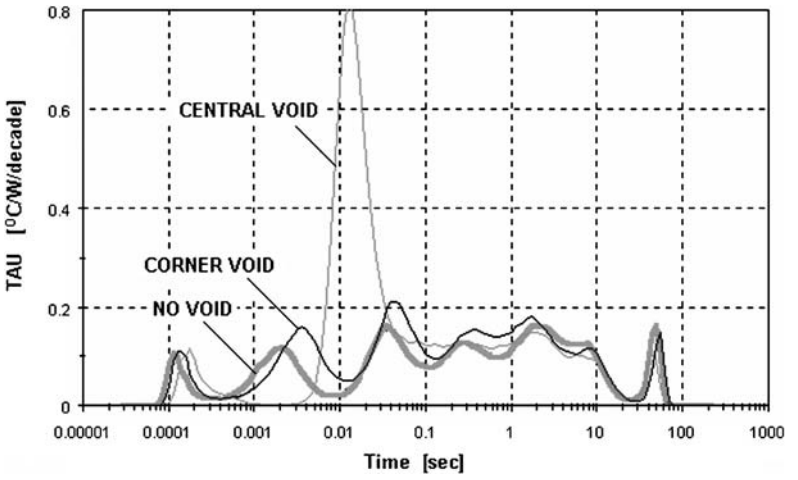


FIGURE 23.8. Main time constants of the experimental samples. (© 2002 IEEE SEMI-THERM Proceedings, "Determining partial thermal resistances with transient measurements and using the method to detect die attach discontinuities," M. Rencz, V. Szekely, A. Morelli, C. Villa, Fig. 8.)

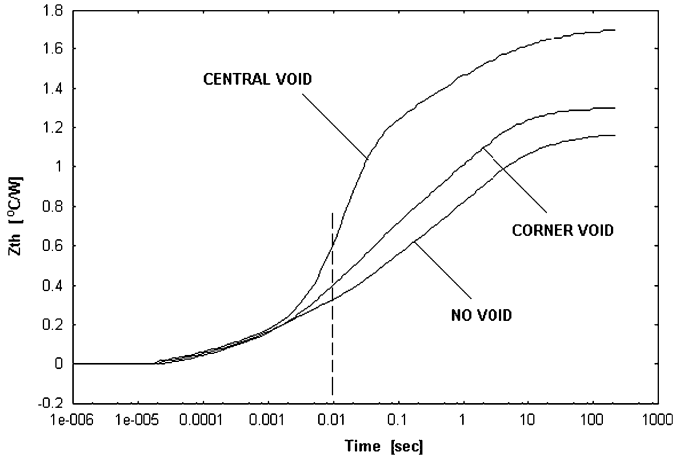


FIGURE 23.9. Measured Z_{th} curves of the average samples with no void, central void and corner voids in the die attach. At 0.01 sec the curves are with and without die attach voids are already clearly distinguishable. (© 2002 IEEE SEMI-THERM Proceedings, "Determining partial thermal resistances with transient measurements and using the method to detect die attach discontinuities," M. Rencz, V. Szekely, A. Morelli, C. Villa, Fig. 9.)

23.4. SIMULATION EXPERIMENTS FOR LOCATING THE DIE ATTACH FAILURE ON STACKED DIE PACKAGES

Today the major challenge to the packaging industry is the reliable increase of the packaging density with the help of stacking several dies in single packages. In case of stacked die packages the die attach quality control problem is multiplied: the structure contains as many die attach layers as the number of dies, and these are very difficult to test. Since any failure in the die attach may contribute to an early ruination of the whole circuit, testing the die attach quality of stacked dies is today a major task in assuring reliability.

The structure function evaluation methodology has demonstrated excellently the capability of detecting die attach problems of single die packages. It can be expected that the method will be applicable also to detect the location of the die attach problems also in the case of stacked die packages as well. The applicability can be proven first with the help of simulation tests. If these tests indicate the applicability of the method, we may start with experiments on real structures.

In order to calculate the structure function of the heat flow path by simulation first the thermal transient behavior has to be simulated, assuring one-dimensional heat flow through the structure. In the presented simulation experiments the simulation was done by the SUNRED program [8]. The one-dimensional heat flow was assured by supposing cold plate at the header of the simulated package, and constant and evenly distributed dissipation was switched on the top of the structure. The increasing temperature was observed on the location of the dissipation, on the top die.

The results of two different structures are presented: a 3D package with equal size dies, and a pyramidal structure, that is frequently used in case of stacked die structures in order to facilitate the bonding.

23.4.1. Simulation Tests Considering Stacked Dies of the Same Size

In the first simulation experiments 3-, 2- and 1-dimensional thermal simulations were applied for 3D stacked die structures of the same die size. Note, that 1-dimensional thermal simulation is very useful sometimes to get an immediate answer to some basic questions. 2- or 3-dimensional simulations need much more time to prepare, but to obtain accurate results we frequently need them. For all the simulation tests step function powering was applied on the top surface of the top layer silicon, and the resulting temperature transients were obtained in the structure by calculating the time dependent temperature in the middle of the top surface. In order to emulate measurement conditions random noise was added to the simulated results and the same resolution (bit-number) was used in the results that we use normally in the measurements.

From the results of the simulated and measurement emulated transient response function, we can calculate [3] the structure functions. This calculation involves supposing one-dimensional heat flow from the top surface of the stacked structure toward the cold plate. In the simulation experiments the dies were separated by die attach (glue) layers, having the same uniform thickness as the die thickness. The considered structure is presented in Figure 23.10.

From the simulated temperature transients the structure functions were created by the measurement evaluation software of the T3Ster tool [9]. In Figure 23.11 the differential structure function of the structure of Figure 23.10 is presented, calculated from the simulated time response. The arrows are pointing to the peaks, representing the silicon dies, their respective distances on the horizontal axis give the values of the respective thermal resistances of the die attach material between them.

In the presented first simulation experiment the thermal resistance of the die attach layer separating the 1st and the 2nd silicon dies was increased, modeling a die attach void in between these dies. In Figure 23.12 from the shift in the right hand side end of the second curve the increased total thermal resistance of the structure can be read, this is about 0.6 K/W.

In the second simulation experiment presented the same increased thermal resistance value was considered between the 2nd and the 3rd silicon dies. It is well observable from the comparison of the figures that the total thermal resistance of the structure is the same increased value in both cases, but the beginning of the displacement is different. In the

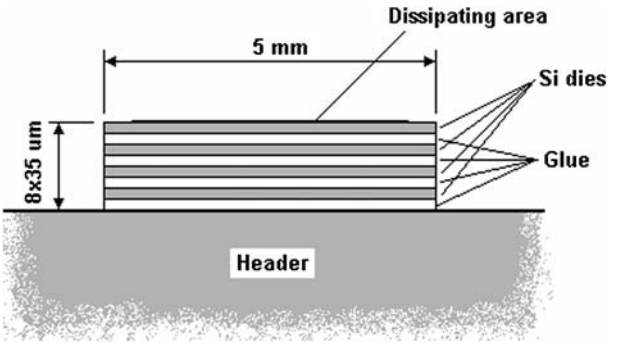


FIGURE 23.10. The simulated structure contained 4 silicon dies of 35 μm thickness, separated by die attach of the same thickness. The figure is not to scale. (© 2004 IEEE SEMI-THERM Proceedings, "Structure function evaluation of stacked dies," M. Rencz, V. Szekely, Fig. 2.)

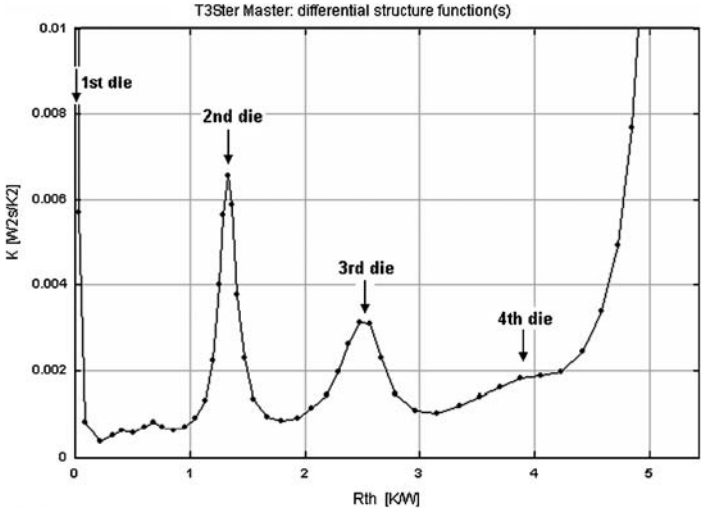


FIGURE 23.11. On the differential structure function of the structure the peaks represent the subsequent dies, their distances give the value of the die attach thermal resistance between them. (© 2004 IEEE SEMI-THERM Proceedings, "Structure function evaluation of stacked dies," M. Rencz, V. Szekely, Fig. 3.)

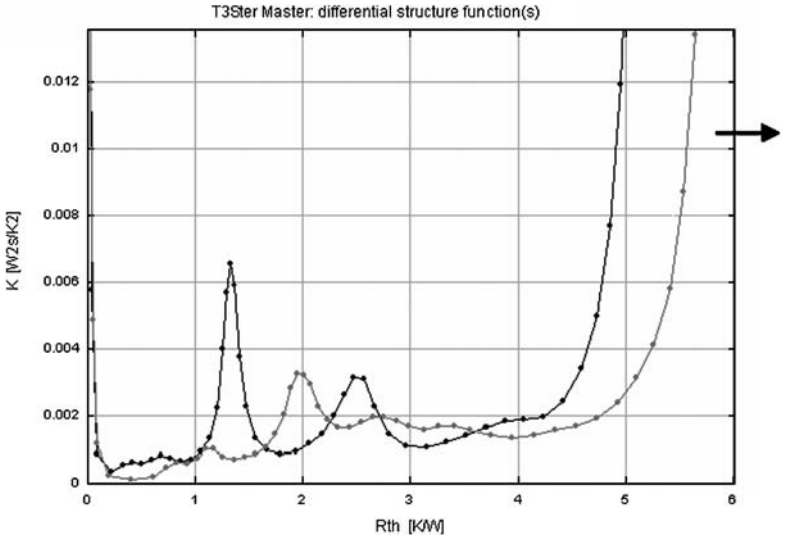


FIGURE 23.12. The increased thermal resistance of the die attach between the 1st and the 2nd silicon layer results in the displacement of the peak representing the 2nd die. (© 2004 IEEE SEMI-THERM Proceedings, "Structure function evaluation of stacked dies," M. Rencz, V. Szekely, Fig. 4.)

second experiment the source of the increased thermal resistance is the die attach between the 2nd and the 3rd dies, since the first two peaks are at the same location in the differential structure function as in the case of the nominal curve of Figure 23.13.

The simple simulation experiments presented demonstrate well the feasibility of the method for the detection of die attach problems in stacked die structures for at least 4 layers of dies.

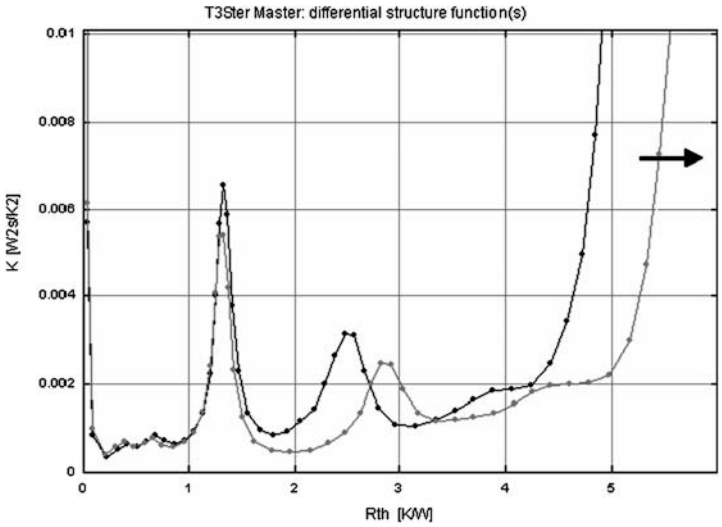


FIGURE 23.13. Inserting additional thermal resistance between the 2nd and the 3rd dies results in the shifting of the 3rd peak and from that on the whole right hand side of the curve. (© 2004 IEEE SEMI-THERM Proceedings, "Structure function evaluation of stacked dies," M. Rencz, V. Szekely, Fig. 6.)

23.4.2. Simulation Experiments on a Pyramidal Structure

Since today most of the stacked dies are of pyramidal structure to facilitate bonding it has to be examined if the method works appropriately also in such cases.

One quarter of the investigated pyramidal structure is presented in Figure 23.14. The dissipating area was supposed in the middle of the top layer. Note that the figure is not to scale.

The sizes of the simulated structure were as follows. The size of the bottom die was 14×14 mm (7×7 in the $1/4$ simulation), the die size in the middle was 12×12 mm, and the size of the top die was 10×10 mm. The thickness of all the die and the glue layers were 0.035 mm. The considered thermal conductivity and volumetric thermal capacity values were 156 W/mK and 1.6×10^6 W s/m^3 K for the silicon, 1 W/mK and 10^6 W s/m^3 K for the glue layers. The considered heat transfer coefficients were 6000 W/m^2 K on the bottom, and 100 W/m^2 K on the top.

The simulated time range was between 1 μs–2 s, logarithmic time steps were applied. The differential structure function of the faultless structure shows the expected peaks, see Figure 23.15. The three chips and the glue under the structure can be well identified.

The thermal resistance and thermal capacitance values may be directly read from the cumulative structure functions. In the step function like cumulative structure function of Figure 23.16 the height values of the individual steps give the thermal capacitances, while the width values of the steps provide the related thermal resistances of the subsequent layers, measured from the top die top surface ($R_{th} = 0$) toward the bottom die and finally to the cold plate. The values that can be read from Figure 23.16 are good approximations of the values that can be calculated for the steady state from the geometrical and material parameters, but do not equal them. The explanation of the small difference is that as the structure is very shallow, the heat is not spreading out in the whole silicon area at once, and the one-dimensional heat flow approximation is not correct in the whole structure. The heat is reaching the cold plate under the structure rapidly, and for this reason practically only the

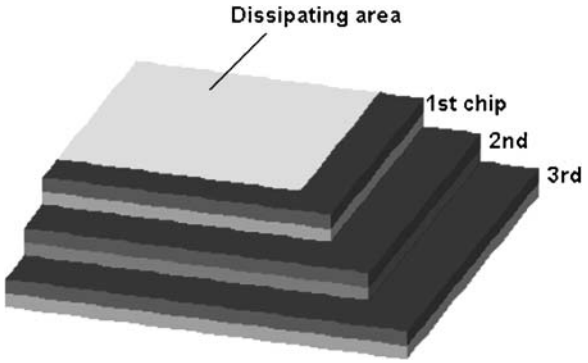


FIGURE 23.14. The investigated pyramidal structure. (© 2004 IEEE SEMI-THERM Proceedings, "Structure function evaluation of stacked dies," M. Rencz, V. Szekely, Fig. 7.)

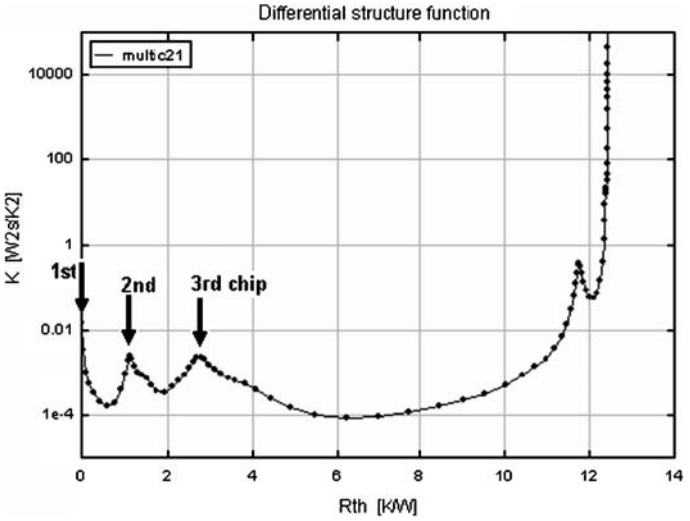


FIGURE 23.15. The structure function of the faultless pyramidal structure. (© 2004 IEEE SEMI-THERM Proceedings, "Structure function evaluation of stacked dies," M. Rencz, V. Szekely, Fig. 8.)

volume under the dissipating shape is participating in the heat conduction, showing smaller thermal capacitance values than expected in the steady state calculations.

Let us suppose now die-attach failure between the 2nd and the 3rd dies. This can be modeled with an increased thermal resistance of the die attachment between these dies. In the results of the simulation experiment presented in Figure 23.17 the thermal resistance of the die attach between the 2nd and 3rd dies was increased by 33%, that is, the thermal conductivity was decreased to 0.75 W/mK.

The curve denoted by ***A*** in Figure 23.17 refers to the faultless case, ***B*** refers to the case of die attach void between the 2nd and 3rd chips. As it is show in the figure, the peak referring to the 3rd chip is displaced now with the value of Δ, while the peak referring to the 2nd chip did not move from the original position. The value of Δ gives back the increased thermal resistance of the die attach layer, as expected.

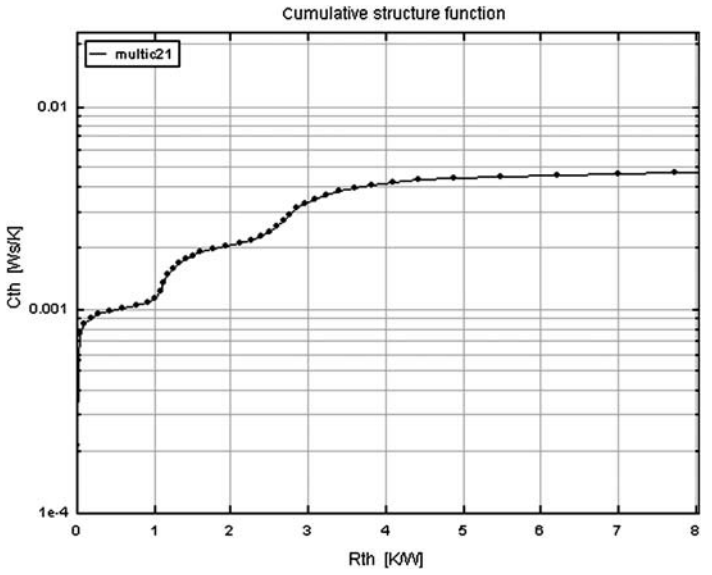


FIGURE 23.16. The cumulative structure function of the faultless structure. (© 2004 IEEE SEMI-THERM Proceedings, "Structure function evaluation of stacked dies," M. Rencz, V. Szekely, Fig. 9.)

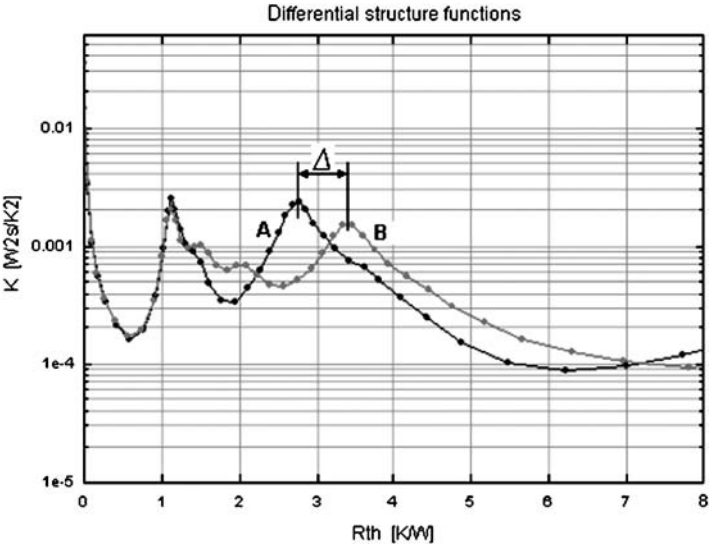


FIGURE 23.17. Die-attach voids are supposed between the 2nd and the 3rd chips. The displacement of the 3rd peak, that is, the distance between the A and B curves gives the increased thermal resistance. (© 2004 IEEE SEMI-THERM Proceedings, "Structure function evaluation of stacked dies," M. Rencz, V. Szekely, Fig. 10.)

In the next presented simulation experiment the die-attach failure was supposed to be between the 1st and the 2nd dies. This was modeled by an increased thermal resistance again, obtained by decreasing the value of the glue thermal conductivity to 0.75 W/mK. The new results are presented in Figure 23.18. In this figure ***A*** denotes the faultless curve,

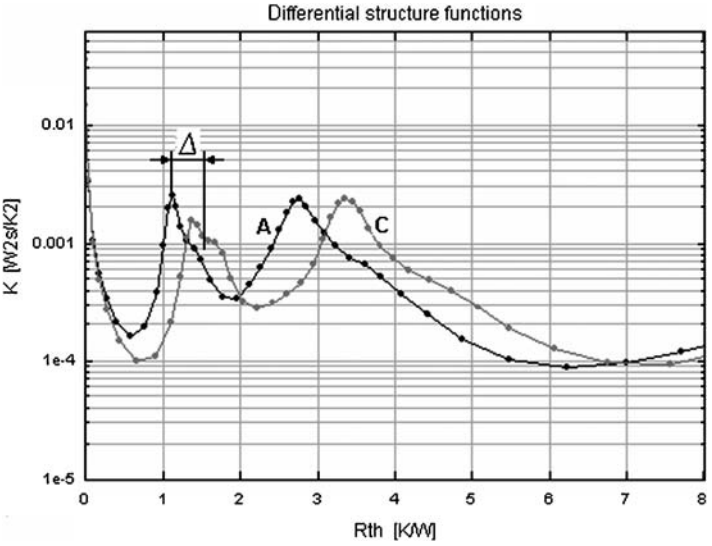


FIGURE 23.18. Die-attach voids are supposed between the 1st and the 2nd chips. The Δ displacement of the 2nd peak, that is, the distance on the A and C curves gives the increased thermal resistance. (© 2004 IEEE SEMI-THERM Proceedings, "Structure function evaluation of stacked dies," M. Rencz, V. Szekely, Fig. 11.)

C denotes the curve obtained by considering an increased thermal resistance between the 1st and 2nd dies. As it is shown in the figure, now the second peak is displaced, showing clearly the location of the increased glue thermal resistance, resulted from the die attach failure. The 3rd peak is displaced as well, with about the same value as the 2nd one.

With these simulation experiments it is demonstrated that the location of an increased thermal resistance, even if this increase is not larger than 20–30%, may be determined also in the case of pyramidal die structures.

23.5. VERIFICATION OF THE METHODOLOGY BY MEASUREMENTS

In order to verify the results obtained from simulations a large series of measurements were accomplished. In the presented first group of experiments the packages contained thermal test dies [10], enabling the verification of the measured results by steady state measurements as well. The thermal transient measurements and the subsequent evaluations were done with the T3Ster equipment [9].

23.5.1. Comparison of the Transient Behavior of Stacked Die Packages Containing Test Dies, Prior Subjected to Accelerated Moisture and Temperature Testing

In the first presented series of measurements we examined several samples that have been prior subjected to accelerated moisture and temperature shock testing, to induce different extent of interfacial integrity. The packages contained 2 test dies, and were of 44 lead QFN (Quad Flat No-lead) type [11].

The cross section of the pristine package is presented in Figure 23.19.

In the presented experiments the top die was used to heat and measure the temperature in the middle of the structure. Two samples were examined, denoted by EM2 and

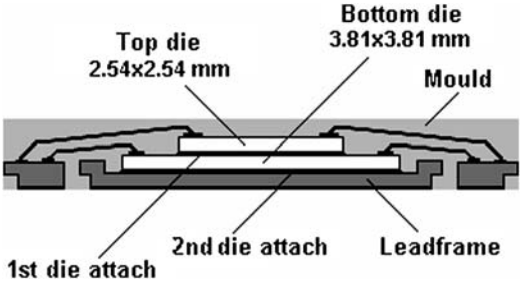


FIGURE 23.19. Cross section of the 44L QFN packages, denoted as EM2 series.

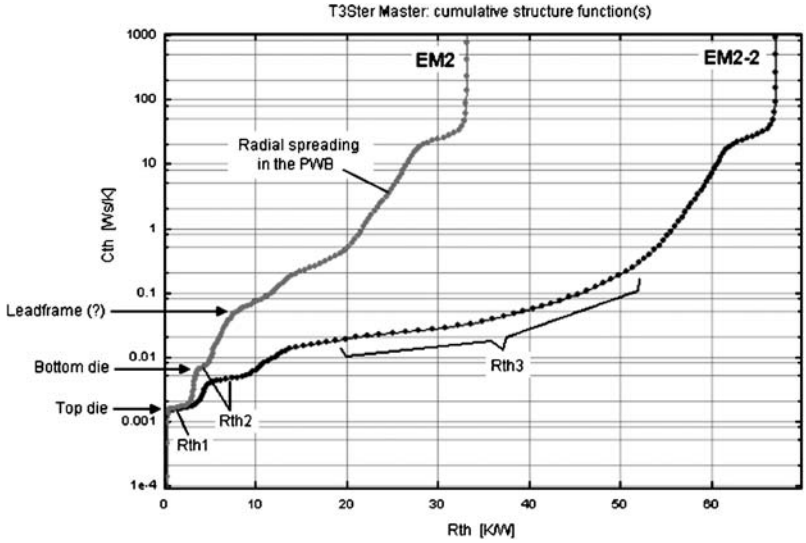


FIGURE 23.20. The cumulative structure functions of the two 44L QFN packages.

EM2-2. The cumulative structure functions calculated from the measured results are presented in Figure 23.20.

The first step in the structure function, denoted as "top die" thermal capacitance of Figure 23.20 indicates a 1.05 mm^3 volume of silicon. The volume of the top die calculated from the data of Figure 23.19 is $2.54 \times 2.54 \times 0.2 = 1.3$ mm^3. "Bottom die" in Figure 23.20 shows the thermal capacitance value of the bottom die. From this value the calculated volume of the bottom die is 3 mm^3, while $3.81 \times 3.81 \times 0.2 = 2.9$ mm^3 volume is expected from Figure 23.19. This difference between the real and the calculated volume is fairly small ($\sim$3%). In case of the package denoted as EM2-2 this thermal capacitance value is smaller, which is may be originating from the detected anomalies, most probably delamination in the further regions of the heat flow path. The almost horizontal section under the top die denoted by R_{th1} is the thermal resistance of the die attach under the top die. It is slightly increased for the sample denoted by EM2-2, the values are: $\sim$2.7 K/W and $\sim$3.8 K/W, suggesting a small delamination in the package EM2-2 in the upper die attach layer.

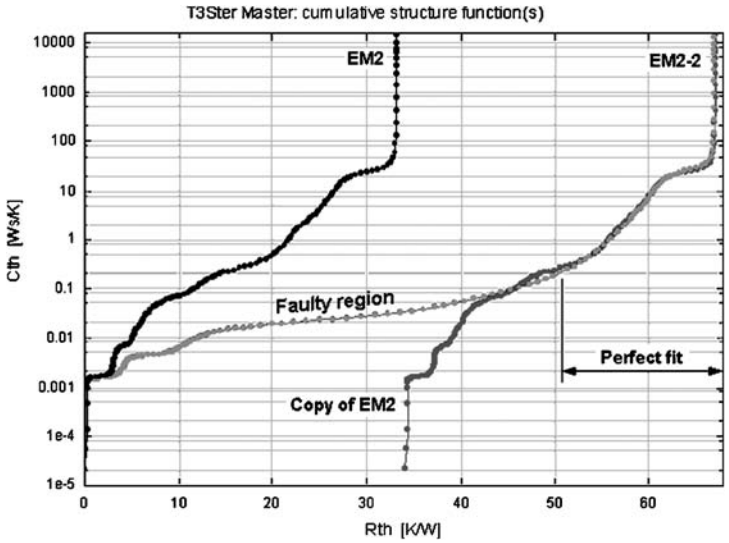


FIGURE 23.21. Cumulative structure functions of the two samples, a copy of the EM2 curve is shifted to match at the right end.

The second almost horizontal section in Figure 23.20, denoted as R_{th2} is the thermal resistance of the die attach under the bottom die. This is strongly different at the two samples: EM2 ∼1.6 K/W, EM2-2 ∼5 K/W, and before reaching the thermal capacitance of the leadframe we can identify a further large almost horizontal part, demonstrating a largely increased thermal resistance, shown by R_{th3} in Figure 23.20. This resistance can be even better seen from a figure where the structure functions of the two samples are drawn up in one figure, and the curve of the known good one is copied and shifted to demonstrate the fit on the right-hand side, see Figure 23.21. The section, where the curve denoted as EM2-2 travels alone, shows the increased thermal resistance on the EM2-2 sample. Since this section of the curve is running entirely under the capacitance value attributed to the leadframe this curve indicates the presence of an increased thermal resistance between the bottom die and the leadframe.

It can be stated from this large shift that there is probably a strong delamination at this sample between the lead-frame and the bottom die. This statement has been later verified also by C-mode Scanning Acoustic Microscope inspections, revealing in fact delamination of the bottom die. These pictures are so blurred however, that even the most experienced eyes have difficulties to evaluate them; this is why they are not presented here.

23.5.2. Comparison of the Transient Behavior of Stacked Die Packages Containing Real Functional Dies, Subjected Prior to Accelerated Moisture and Temperature Testing

The final test of the method is when the examined packages contain real functional stacked dies. In the presented experiments 48 lead QFN packages were examined, containing two stacked functional dies each. The measurements were done by using the substrate diode of the top die both for heating and temperature sensing. #789 and #630 are the codes of the two investigated sample packages. The measurements were blind, nothing was known about the reliability testing history of the packages. Figure 23.22 shows the

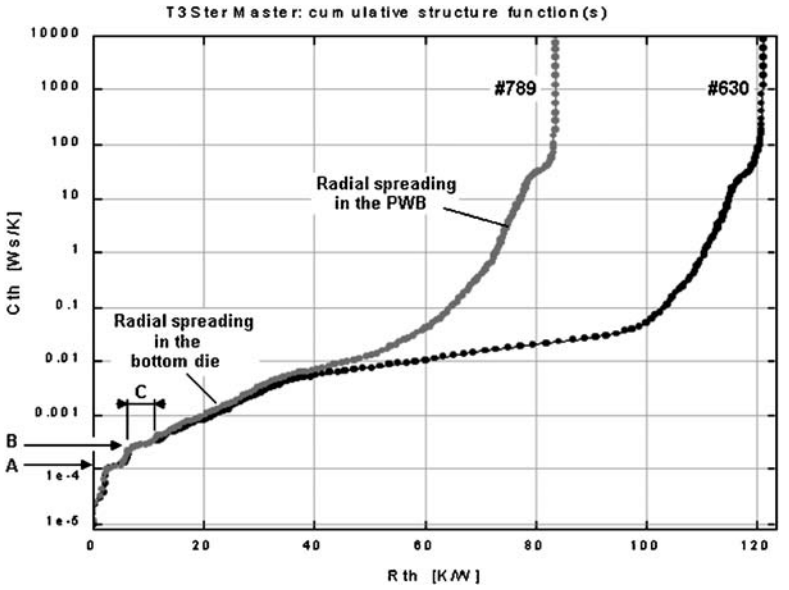


FIGURE 23.22. Cumulative structure functions of the measured two samples.

cumulative structure functions, calculated from the measured thermal transient results. The following train-of-thoughts shows how to draw inferences and conclusions by using the methodology.

The location of the top die in the structure function of Figure 23.22 needs some thinking. At the first glance the step marked with "***A***" with its value of $C_{th} = 1.3 \times 10^{-4}$ W s/K seems to be the thermal capacitance of the top die. Calculating however this C_{th} value from the geometrical data, using 1.6×10^{-3} W s/mm^3 K volumetric heat capacity of the Si we obtain $C_{th} = 1.555 \times 1.555 \times 0.076 \times 1.6 \times 10^{-3} \cong 2.9 \times 10^{-4}$ W s/K. This suggests that in fact the second step, marked with "***B***" corresponds to the top die, and the first one may be related to the transients inside the top die. The almost horizontal section denoted by "***C***" gives the ≈6 K/W thermal resistance of the die attach under the top die.

Ascendant linear regions of the structure function (in the lin/log representation) indicate always radial heat spreading. [12] Such a region is shown in Figure 23.22, denoted as "Radial spreading in the bottom die." This suggests that the heat moves from the top die first to the middle of the bottom die and continues to spread out laterally. From the slope of this region even the thickness of the bottom die may be deduced as follows:

$$w = \frac{1}{\lambda}\frac{1}{4\pi}\frac{\log(C_{th2}/C_{th1})}{R_{th2} - R_{th1}} = \frac{1}{120}\frac{1}{12.56}\frac{\log(10)}{30 - 11} = 83 \times 10^{-6}\ \text{m} \tag{23.3}$$

where R_{th1}, C_{th1} and R_{th2}, C_{th2} are representing two points in the linear section, the slope of which is calculated [12].

This calculation results in an about 83 μm thickness value for the bottom die, while the real thickness is 76.2 μm.

The "Radial spreading in the bottom die" section ends at $C_{th} \approx 0.002$ W s/K, the calculated heat capacity of the bottom die is $C_{th} = 3.467 \times 3.302 \times 0.1 \times 1.6 \times 10^{-3} = 0.00183$ W s/K, the fitting is acceptable. At the same location the differential structure

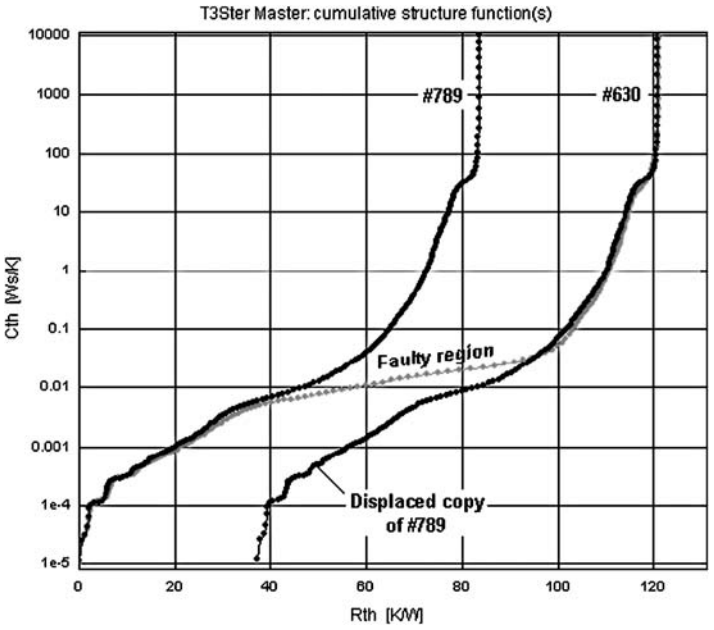


FIGURE 23.23. The copy of the structure function of the good device is displaced to fit at the ambient side to the structure function of the faulty device.

function (see Figure 23.24) gives $K = 3 \times 10^{-4}$ W s^2/K^2 value, from which, for Si material 1.25 mm^2 cross section is calculated by T3Ster. This is about the peripheral area of the bottom die. This means that at the right hand side of the "Radial spreading in the bottom die" section of Figure 23.22 the heat reaches exactly the periphery of the bottom die.

This calculation and the curves in Figure 23.23, where the copy of the structure function of the good device is displaced to fit at the ambient side to the structure function of the faulty device, suggest that the faulty region of the #630 sample is under the bottom die. If we calculate also the differential structure functions, we can identify also the location of the leadframe from the bump at the area referring to the leadframe.

Comparing Figure 23.23 and Figure 23.24 we can determine that the faulty section (from Figure 23.23) is divided by the "Cu leadframe" bump of Figure 23.24. This suggests that there is a strong delamination both in the die attach under the bottom die and in the soldering of the leadframe. In order to check this, the sample #630 was examined by using X-ray microscopy, see Figure 23.25. The X-ray images clearly show the deterioration of the soldering between the package and the PWB. The loss in the solder area is about 20%; this explains only partially the increase in Rth. The other part of the increase has to be attributed to the induced delamination of the die attach, that unfortunately can not be seen even by X-ray images.

The conclusions inferred from the structure functions may be cross checked with the help of the time constant density functions [3] and the complex loci of the $Z_{th}(\omega)$ curves as well [6]. These curves are generated by the same mathematical transformations that are used to create the structure functions and provided by the measuring equipment [9]. Figure 23.26 shows the calculated thermal time constant density of the measured samples, while Figure 23.27 shows the frequency dependence of their complex thermal impedances [6].

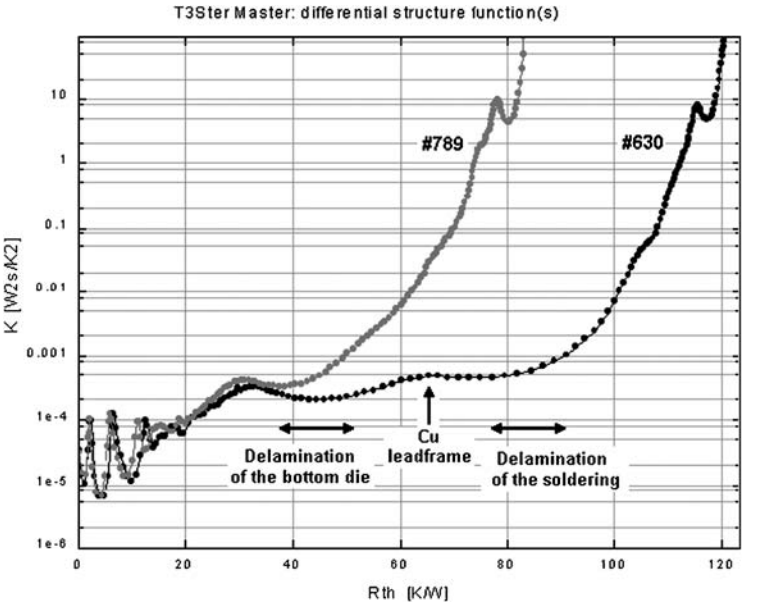


FIGURE 23.24. Differential structure functions calculated from the same measurements.

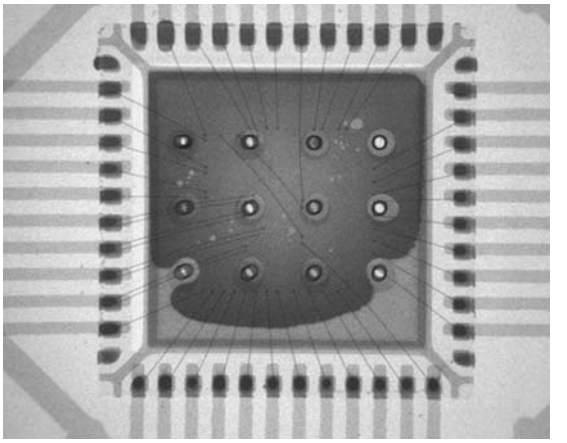


FIGURE 23.25. X-ray image shows the delamination in the soldering (dark area in the figure).

A characteristic feature of these curves is that the first parts of them show nice fitting, suggesting that there is no defect in the neighborhood of the top die, but the right hand side of the curves in both of the figures show large differences, suggesting a different material structure, that is delamination of the layers.

Although these functions are very useful for the experienced reliability engineer to locate die attach problems, we recommend to start the failure analysis with the structure function evaluation, since these functions can provide directly the data that can be used for characterizing the die attach problems.

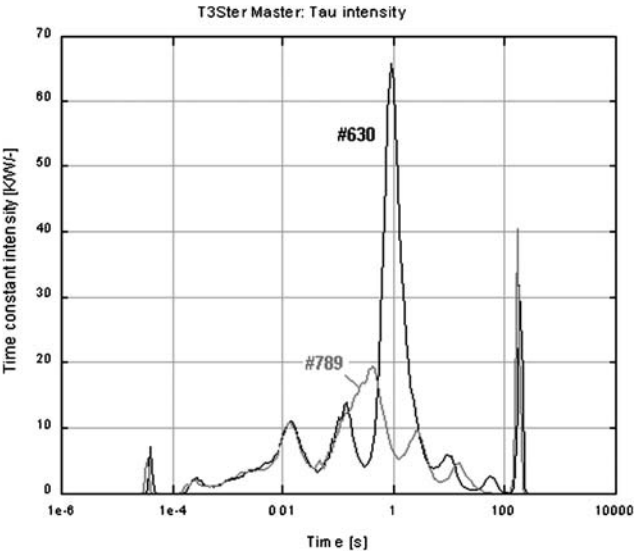


FIGURE 23.26. The calculated thermal time constant density of the measured samples.

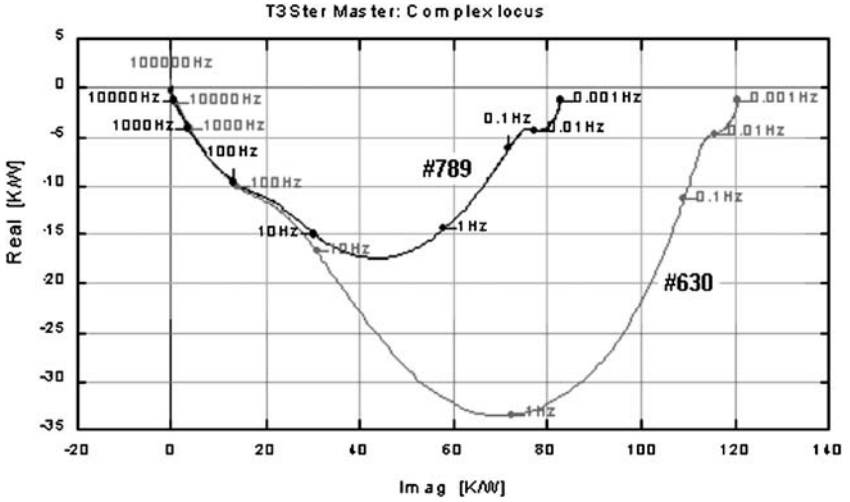


FIGURE 23.27. Frequency dependence of the complex thermal impedances of the measured samples.

A special advantage of the presented test technology is that thermal test dies are not needed to detect integrity problems in the packages. Usual working dies can be used both as heater and temperature sensor in the proposed methodology. For measuring the structure function we even do not need any extra elements on the dies if the substrate diodes of the individual dies are not connected electronically, since for the temperature sensing the substrate diode of the top die may be conveniently used. If for some reason the substrate diodes of the dies have to be connected electrically, other dedicated elements have to be selected on the dies for temperature sensing.

On the other hand, switching on either the whole dissipation on the top die, or just one dissipating element, may provide the constant source of heat during the time of the transient measurement. This means that obtaining the necessary transient curves is feasible normally without any special built in elements or test structures in any circuit. This way we may conclude that with the presented method the fast diagnosis of the die-attach problems of stacked dies is possible, and the method does not require any additional circuit elements in any of the stacked dies.

With the help of the method die attach problems affecting about 5% of a certain surface can be detected in the form of increased thermal resistance (shift in the structure functions). This value strongly depends on the actual structure. For usual package structures it can be even smaller, but sometimes it may be somewhat larger. The closer is the heat flow in the structure the 1-dimensional flow, the smaller is the die attach failure that can be detected.

Finding the exact locations of the material transitions is not always very simple. It strongly depends on the geometry and on the materials used in the structure. It can be stated however, that by finding the inflection points in the structure function, which can be easily automated in the structure function software, the material transition point can be easily approximated.

23.6. CONCLUSIONS

With the presented measurement experiments we have demonstrated the feasibility of locating die attach problems of packages containing either single or stacked dies. In the proposed methodology first the thermal transient response of the structure has to be obtained from measurement, and from this the structure functions have to be calculated. Comparing the structure functions of a measured sample with the appropriate structure functions of a known good stacked structure, which can be obtained, e.g., from the simulation of the structure, or from the measurement of a known good sample package, the differences can be easily observed. From the location of the shift in the structure function of the device under test, in case of structures with integrity problems, the location of the integrity problem can be determined.

The method works well both in the case of stacked dies of the same size, and in the case of pyramidal stacked die structures. The resolution of the method will be further investigated with different die sizes and layer numbers, but from the experiments up to now it can be expected that the method is well applicable for structures up to at least 3–4 layers of stacked dies.

The large number of experiments and cross verifications indicated that if we wish to use the methodology to determine the exact thermal resistance values of the individual die attach layers the measured values have to be corrected with the structure functions of the parallel heat flow paths [13]. The correction can be a built in function of the evaluation software of the thermal transient measurement equipment.

ACKNOWLEDGMENTS

This work was supported by the PATENT IST-2002-507255 Project of the EU and by the INFOTERM 2/018/2001 NKFP project of the Hungarian Government. The authors

wish to thank G. Farkas his contribution by carrying out some of the measurements, and for R. Kovács for taking the X-ray image.

REFERENCES

1. L. Zhang, N. Howard, V. Gumaste, A. Poddar, and L. Nguyen, Thermal characterization of stacked dies, Proceedings of the XXth SEMI-THERM Symposium, San Jose, CA, USA, March 9–11, 2004, pp. 55–63.
2. C. Lin, S. Chiang, and A. Yang, 3D stackable packages with bumpless interconnect technology, 5th Electronics Packaging Technology Conference, Singapore, 10–12 December 2003.
3. V. Székely and T. van Bien, Fine structure of heat flow path in semiconductor devices: a measurement and identification method, Solid-State Electronics, 31, pp. 1363–1368 (1988).
4. M. Rencz, V. Székely, A. Morelli, and C. Villa, Determining partial thermal resistances with transient measurements and using the method to detect die attach discontinuities, Proceedings of the XVIIIth SEMI-THERM Symposium, San Jose, CA, USA, March 1–14, 2002, pp. 15–20.
5. M. Rencz and V. Székely, Structure function evaluation of stacked dies, Proceedings of the XXth SEMI-THERM Symposium, San Jose, CA, USA, March 9–11, 2004, pp. 50–54.
6. V. Székely, Distributed RC networks, The circuits and Filters Handbook, CRC Press, USA, 2003, pp. 1202–1221.
7. E.N. Protonotarios and O. Wing, Theory of nonuniform RC lines, IEEE Trans. on Circuit Theory, 14(1), pp. 2–12 (1967).
8. http://www.micred.com.
9. http://www.micred.com/t3ster.html.
10. http://www.delphi.com/pdf/techpapers/2004-01-1681.pdf.
11. M. Rencz, V. Székely, B. Courtois, L. Zhang, N. Howard, and L. Nguyen, Die attach quality control of 3D stacked dies, IEMT, 2004, pp. 78–84.
12. V. Székely, M. Rencz, S. Török, and S. Ress, Calculating effective board thermal parameters from transient measurements, IEEE Transactions on Components and Packaging Technology, 24(4), pp. 605–610 (2001).
13. M. Rencz, A. Poppe, E. Kollár, S. Ress, V. Székely, and B. Courtois, A procedure to correct the error in the structure function based thermal measuring methods, Proceedings of the XXth SEMI-THERM Symposium, San Jose, CA, USA, March 9–11, 2004, pp. 92–98.

24

Mechanical Behavior of Flip Chip Packages under Thermal Loading

Enboa Wu[a,e], Shoulung Chen[a,b], C.Z. Tsai[a,c] and Nicholas Kao[a,d]

[a]*Institute of Applied Mechanics, National Taiwan University, Taipei 106, ROC*
[b]*Electronics Research and Service Organization, Industrial Technology Research Institute, Hsin-Chu 310, Taiwan, ROC*
[c]*Macronix International Co., Ltd., Hsin-Chu 300, Taiwan, ROC*
[d]*Siliconware Precision Industries Co., Ltd., Taichung 400, Taiwan, ROC*
[e]*Hong Kong Applied Science and Technology Research Institute, Hong Kong*

Abstract

A complete report on mechanical behavior of large flip chip plastic ball grid array (FC-PBGA) packages under reflow condition is presented in this chapter. The coefficients of thermal expansion (CTE) of BT substrates were also measured using electronic speckle pattern interferometry (ESPI) and were found to change significantly at different processing stages. Careful selection of a substrate CTE is needed for accurate warpage prediction using the finite element method. A linear relationship between the temperature and the FC-PBGA warpage was observed from the data measured using both the phase-shifted shadow moiré and the ESPI. Zero warpage was measured at approximately 150°C regardless of the size of the chip composed of the package. An optimal warpage design for FC-PBGA was also suggested.

24.1. INTRODUCTION

The mechanical behavior of a flip chip package under temperature change is always important as warpage and stresses are generated due to mismatch of the mechanical properties of the materials used to make the package. A flip chip package is usually composed of a silicon chip, solder joints, an organic substrate, and underfill. In order to understand the behavior of flip chip packages under thermal loading, numerous numerical and theoretical methods have been employed. For theoretical analysis, bi-material and tri-material models have been developed to predict stresses and failure of flip chip packages [1–3]. Simple structural mechanics models, such as those based on beam or plate theories, have frequently been adopted. These models have also been used to predict stresses on solder joints [4,5], underfills [6], or UBM [7] of flip chip packages, and could possibly be applied under different environmental conditions [8,9]. For finite element analysis, two-dimensional and three-dimensional finite element models have been constructed to study

packaging deformation [10,11], interfacial stresses [12], solder joint shapes and stresses [13–15], and underfill behaviors and stresses [16] to determine the reliability of packages [17–19]. Parametric studies have also been presented to determine the optimal properties and dimensions of the materials and devices in packages [10–22].

On the other hand, experimental tools have been adopted to directly measure the behavior of flip chip packages at different temperatures. Full field optics methods, such as the shadow moiré, moiré interferometry, reflection moiré, or electronic speckle pattern interferometry (ESPI) schemes, have frequently been adopted. The shadow moiré method has been employed to measure the out-of-plane deformation of flip chip packages [23–26], printed wiring boards [27,28], and organic substrates [29]. The resolution for shadow moiré can be enhanced by an order of magnitude if the phase shifting setup and subsequent algorithms for phase diagram construction are employed [30]. Moiré interferometry, on the other hand, has frequently been adopted to measure the in-plane deformation of the cross-sections of solder joints and flip chip packages after the temperature cools down [31]. Because a package subjected to moiré interferometry measurement has to be cross-sectioned, the deformation on the cross-section is disturbed. As a result, the measurement result may not reflect the true deformation conditions in the package. This disadvantage can be avoided by using ESPI, where the experimental setup can be arranged to measure either the in-plane or the out-of-plane deformation of the package [32–34]. On the other hand, as the sensitivity of ESPI is high, it is suitable only for deformation measurement at small scale. In addition, ESPI is fragile to the ambient environment. When the measurement is performed at elevated temperature, care must be taken to avoid drifting of the fringes due to the unstable hot air flow. When the measurement surface is reflective or shiny, such as a polished die surface in a flip chip package, reflection moiré is an ideal method for measuring the out-of-plane deformation [35]. The sensitivity is high because the slope of the out-of-plane deformation is directly measured.

In this chapter, we will discuss use of both the ESPI and phased shifted shadow moiré methods to measure the mechanical behavior of flip chip packages. We will also discuss how the coefficients of thermal expansion of flip chip BGA substrates at different fabrication stages are measured using ESPI, and how the out-of-plane deformation of different types of flip chip packages is measured using the phased-shifted shadow moiré and the ESPI methods. The use of two-dimensional and three-dimensional finite element models to perform parametric analyses of flip chip packages will also be discussed.

24.2. FLIP CHIP PACKAGES

A typical picture of a flip chip package is shown in Figure 24.1. The flip chip packaging process generally includes wafer bumping and flip chip assembly. In the wafer bumping process, the peripheral pads on each chip in a wafer are redistributed to form area array pads. Under bump metallurgy and solder bumps are then deposited on the redistributed pads. In the flip chip assembly process, bumped chips, after being diced from a wafer, are placed on a substrate and undergo a subsequent reflow process. Underfill is then deposited to reinforce the solder bumps and enhance the reliability of the flip chip package.

In this chapter, we will focus on the mechanical behavior of flip chip packages when they are subjected to different temperature conditions. Except for the flip chip packages discussed in Section 24.5.3 to study the underfill effect, flip chip packages used for warpage measurement can be divided into four groups, as listed in Table 24.1. The chip sizes used

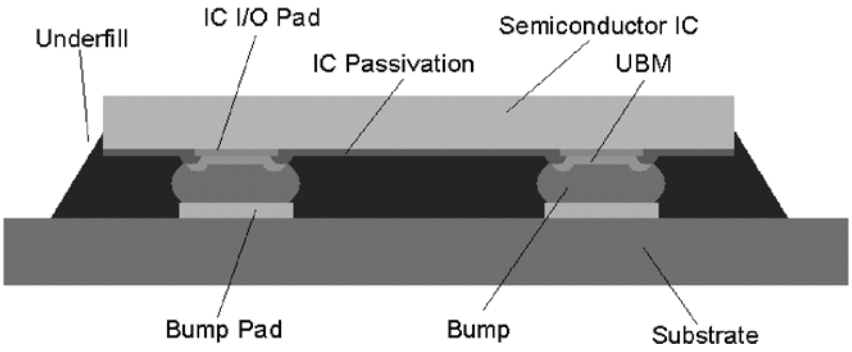


FIGURE 24.1. Schematic diagram of a typical flip-chip package.

TABLE 24.1.
Dimensions in mm of flip chip packages and the number of samples used.

Group	1	2	3	4
Chip size	10 × 10	20 × 20	26 × 26	26 × 26
Chip thickness	0.73	0.73	0.73	0.40
Solder bump pitch	0.5	0.5	0.5	0.5
Bump height	0.07	0.07	0.07	0.07
Substrate size	40 × 40	40 × 40	40 × 40	40 × 40
Substrate thickness	1.12	1.12	1.12	1.12
Sample No.	3	3	2	2

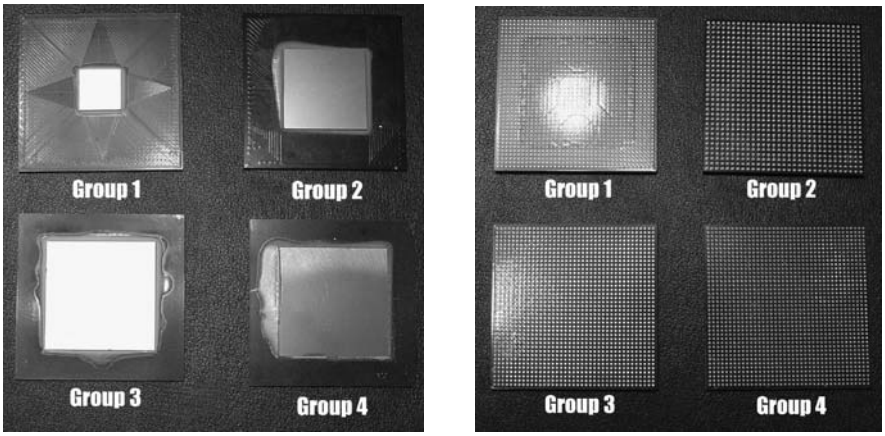
TABLE 24.2.
Mechanical properties of chip, solder bump and underfill.

	E (GPa)	CTE (ppm/°C)	Poisson ratio
Chip	156	2.7	0.278
Solder bump	57.4	24.5	0.4
Underfill	5.5	45	0.35

are 10 × 10 mm, 20 × 20 mm, and 26 × 26 mm. The thicknesses of the chips are 0.73 mm and 0.40 mm. The solder bump height and the pitch are 70 μm and 500 μm, respectively. The mechanical properties of the silicon chip, the solder bump and the underfill are listed in Table 24.2. For these flip chip packages, the build-up substrates used are 40 × 40 × 1.12 mm in size and are made by deposition of one top and one bottom build-up layers onto two 2-metal-layer cores (1 + 4 + 1). The material data are listed in Table 24.3 and were provided by manufacturers. The coefficient of thermal expansion (CTE) will be verified in Section 24.4. Figure 24.2 shows pictures of Group 1, 2, 3, and 4 flip chip packages. In this chapter, we will also discuss the use of other types of substrates for CTE measurement. The results will be given in Section 24.5.

TABLE 24.3.
Mechanical properties of the substrate used for the flip chip packages listed in Table 24.1.

	E (GPa)		CTE (ppm/°C)		Poisson ratio	
	In-plane	Out-of-plane	In-plane	Out-of-plane	In-plane	Out-of-plane
Substrate	24.5	9.81	12	50	0.143	0.02



(a) (b)

FIGURE 24.2. Photographs of Groups 1, 2, 3, and 4 flip chip packages. (a) Chip side; and (b) substrate side.

24.3. MEASUREMENT METHODS

24.3.1. Phase Shifted Shadow Moiré Method

For shadow moiré [23–29], the fringe patterns are formed through superposition of the reference gratings and the object gratings. The reference gratings are placed in front of a specimen whose out-of-place displacement or surface profile is to be measured. The object gratings are the shadows of the reference gratings, which are produced by shining parallel light onto the specimen surface through the reference gratings. The out-of-plane displacement can be expressed as

$$w = \frac{Np}{\tan\alpha + \tan\beta}, \tag{24.1}$$

where N is the fringe order, p is the pitch of the object grating, α is the angle between the projected parallel light and the normal of the reference grating plane, and β is the angle between the axis of the CCD camera and the normal of the reference grating plane. In this chapter, α will always be equal to 45°, β will always be equal to 0°, and $p = 50$ μm, i.e., 20 lines/mm.

In order to enhance the sensitivity to warpage in shadow moiré measurement, the phase of the recorded moiré fringes is used instead of the intensity of the fringe patterns. The general expression for the fringe intensity can be expressed as in Equation (24.2) when the reference plane moves a distance equal to p/n along its normal direction. If we take $n = 4$, i.e., adopt the four-step phase shifting method, the phase of the recorded light intensity

can be extracted by using Equation (24.3) [36]:

$$I_k(x, y) = I_0(x, y)\{1 + \gamma(x, y)\cos[\phi(x, y) + \alpha_k]\},$$
$$\alpha_k = (k-1)\alpha_r,$$
$$k = 1, 2, \ldots, n, \tag{24.2}$$

$$\phi(x, y) = \tan^{-1}\left(\sqrt{3}\frac{I_1 - I_3}{2I_2 - I_1 - I_3}\right) + \frac{2\pi}{3}. \tag{24.3}$$

All the moiré measurements discussed in this chapter are performed using the four-step phase shifting method. Measurement, which is frequently performed at an elevated temperature, can thus be accomplished in a relatively short period of time. Movement of the reference grating plane is controlled by a stepping motor.

24.3.2. Electronic Speckle Pattern Interferometry (ESPI) Method

When laser light is directed onto the surface of an object, speckles are formed because of the roughness of the surface. These speckles move as the object deforms. As a result, fringes are formed, and the corresponding deformation can be measured. In this study, ESPI was adopted for both in-plane and out-of-plane deformation measurements of flip chip packages [32–34].

The experimental set up for in-plane deformation measurement performed using ESPI is depicted in Figure 24.3. Laser light was split into two beams of equal intensity. These two beams traveled the same distance and, by means of two spatial filters, were uniformly distributed onto the surface of the object whose deformation was to be measured. The speckle patterns were recorded by a CCD camera when the surface of the body to be measured moved.

Because the two laser beams are symmetric along the CCD axis, any movement in the out-of-plane direction of the object will not change the traveling distance of the two optical paths. As a result, no fringe is formed by this out-of-plane movement. On the other hand, when the object moves in the in-plane direction, the difference in the optical paths of the

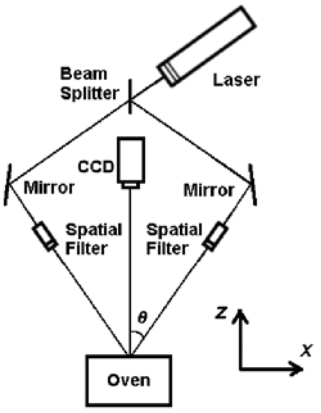


FIGURE 24.3. Schematic diagram of the experimental setup for in-plane ESPI displacement measurement.

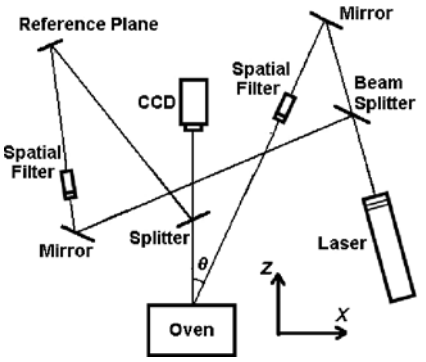


FIGURE 24.4. Schematic diagram of the ESPI experimental setup for out-of-plane deformation measurement.

two laser beams that hit a specific speckle changes. As a result, the in-plane displacement can be expressed as

$$U_x = \frac{n\lambda}{2\sin\theta}. \tag{24.4}$$

The ESPI experimental set up for out-of-plane displacement measurement is shown in Figure 24.4. A laser beam was split into object and reference beams by a beam splitter. Spatial filters were employed to produce uniformly distributed scattering light projected onto the object and reference planes. The path distances between these two beams were adjusted so as to be equal. A CCD camera was employed to record the fringe patterns when the object inside the oven deformed due to temperature change. The incident angle, θ, was kept as small as possible to reduce the effect of in-plane displacement on the out-of-plane displacement measurement. The out-of-plane displacement corresponding to the nth dark fringe is

$$U_z = \frac{n\lambda}{1+\cos\theta}. \tag{24.5}$$

In this study, the surface of the electronic package to be measured using either the shadow moiré or ESPI methods was always sprayed with white paint in order to enhance contrast. During the measurement process, the flip chip package was mounted on a tripod inside an oven. Because of the high sensitivity of the ESPI method, the temperature loading applied to the specimens was caused by a 5 to 10°C temperature difference, and the accuracy of the temperature adjustment was controlled so as to be ±0.1°C during the measurement process.

24.4. SUBSTRATE CTE MEASUREMENT

The in-plane ESPI experimental setup shown in Figure 24.3 was employed to perform in-plane deformation measurement of substrates. The specimen to be measured was supported by a small tripod inside an oven. Once the in-plane deformation had been recorded, with knowledge of the temperature change, the equivalent CTE value was calculated. In this experimental setup, the incident angle, θ, that is the angle between the laser beam

TABLE 24.4.
Specimens used for CTE measurement.

Specimen		Dimensions (mm)
A	Aluminum	$31.0 \times 30.1 \times 1.0$
B	Copper	$30.6 \times 29.7 \times 1.5$
C	BT Laminates	$35.0 \times 35.0 \times 0.45$
D	BT Laminates with via	$35.0 \times 35.0 \times 0.45$
E	BT Laminates with via and copper trace	$35.0 \times 35.0 \times 0.50$
F	BT substrate	$35.0 \times 35.0 \times 0.55$

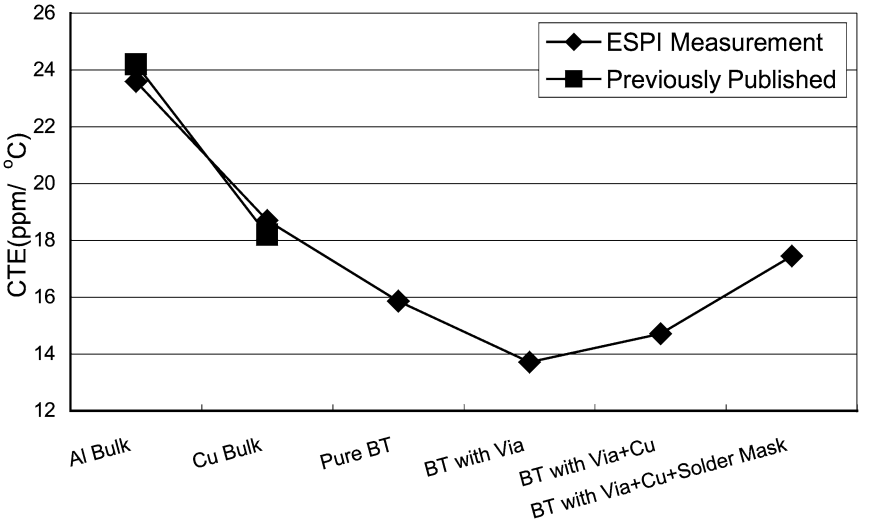


FIGURE 24.5. Measured CTE values for aluminum, copper, and BT substrates at different fabrication stages.

and the axis of the CCD camera shown in Figure 24.3, was 30°. Therefore, the in-plane displacement of the specimen was equal to $n\lambda$ and could be determined directly from the speckle fringe number.

In this study, four types of BT substrates were used to perform CTE measurement. In the first group of CTE measurements, six specimens were employed. The purposes of this group of measurements were (a) to verify the accuracy of the ESPI method by measuring the CTE values of aluminum and copper; and (b) to study the change of CTE values for BT substrates in different fabrication stages. Table 24.4 lists the specimens and their dimensions.

To fabricate a BT substrate for use in electronic packages, glass/BT composite laminates need to undergo drilling, via filling, trace forming, and solder mask forming processes. In this part of study, the BT substrate used in Table 24.4 was for PBGA packaging. The final thickness of the BT substrate was 0.55 mm. The in-plane dimensions of the BT substrate were 35×35 mm.

For each measurement, the temperature range was approximately 10.0°C. For each specimen, the temperature varied within the range of 50°C to 70°C. Four measurements were performed for each specimen. The results are plotted in Figure 24.5. The recorded values for aluminum and copper were 24.2 ppm/°C and 18.2 ppm/°C, respectively. The

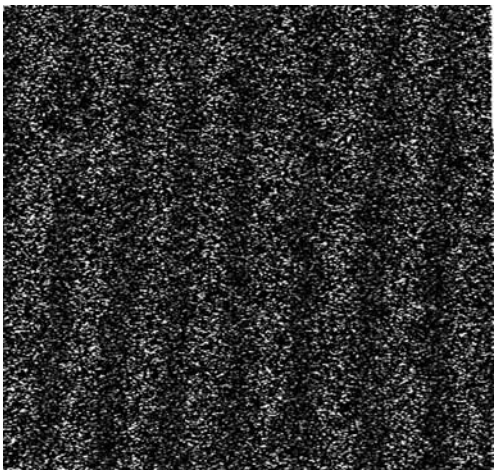


FIGURE 24.6. A typical ESPI fringe pattern for a BT laminate without via holes.

previously published values are 23.6 ppm/°C and 18.7 ppm/°C, respectively [37]. The deviation is less than 3%, which shows that the measurement method is accurate.

The measured average CTE for the BT/glass laminates was 15.86 ppm/°C. This value is very close to the value provided by the manufacturer, which is 15.0 ppm/°C (Mitsubishi Gas Chemical Co.). After the BT/glass laminates were drilled to form vias, the CTE value became 13.71 ppm/°C, a reduction of approximately 14%. This is because the drilled vias provide additional flexibility for BT/glass composite laminates to expand onto the neighborhood of these vias when the temperature is increased. As a result, the overall elongation for the laminates with vias becomes smaller as compared to the overall elongation for the laminates without these vias when the temperature is increased. Once the copper foil was plated, the CTE value increased to 14.71 ppm/°C, approximately a 7% increase. This increase of the CTE value occurred because the CTE value of the copper was approximately 18.2 ppm/°C, i.e., 33% higher than the recorded CTE of the BT/glass laminates. The last process in substrate fabrication was deposition of the solder mask. The CTE value was found to jump to 17.45 ppm/°C, a 19% increase. This significant increase of the CTE value indicated that the CTE value of the solder mask is much higher than that of the BT/glass laminates. As a result, significant stresses were induced at the interfaces among the BT/glass laminates, copper foil, and solder mask. The results for CTE measurement performed at different manufacturing stages for BT substrates are shown in Figure 24.5. For illustration purposes, in Figure 24.6 it shows a typical fringe pattern for a BT/glass composite laminate without via holes. The total pixel number for this 35 mm wide laminate was 286, and there existed six complete fringes and two additional fringes at two edges of the laminates that were not completely shown. For these six complete fringes the pixel number was 218, and the total spacing was 26.68 mm. In this measurement set-up each fringe is equal to an in-plane deformation of 0.6328 μm, and during the measurement the temperature was increased from 65.0°C to 73.9°C. As a result, the CTE is calculated as $6 \times 0.6328 \times 1000/(26.68 \times (73.9 - 65.0)) = 15.99$ ppm/°C, which is very closed to the data of 15.86 ppm/°C shown in Figure 24.5. The slight deviation is that in Figure 24.5 the CTE value was averaged from multiple measurements at different temperatures.

The second type of specimen used for CTE measurement was a BT substrate used in a two-chip PBGA module, i.e., an MCM module. The dimensions were $30 \times 12 \times 0.56$ mm.

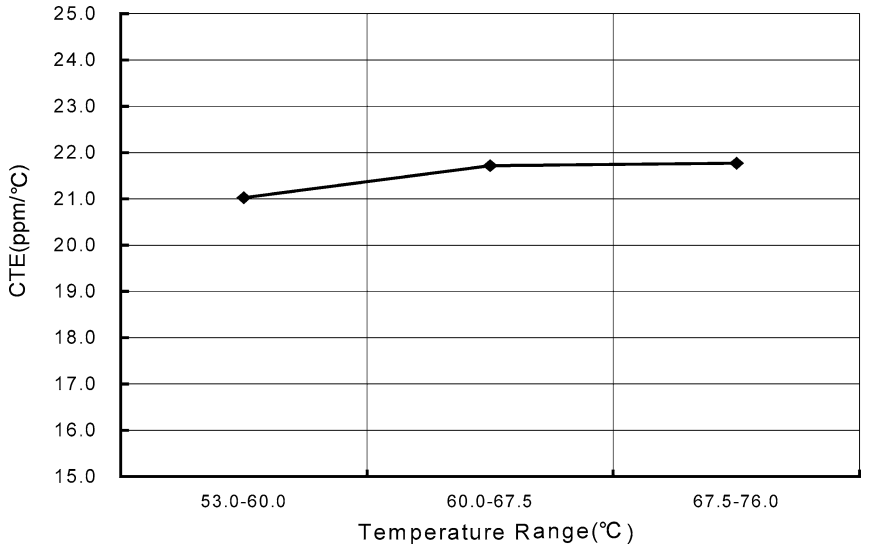


FIGURE 24.7. CTE values for 0.56 mm thick BT substrate used for MCM.

The average CTE value measured within three temperature ranges, 53°C to 60°C, 60°C to 67.5°C, and 67.5°C to 76°C, was 21.25 ppm/°C. The results are shown in Figure 24.7. This value is much higher than the recorded CTE values for the single chip PBGA substrate, as shown in Figure 24.5. The significant increase of the CTE value for this substrate may be due to the thicker layer of the solder mask used for MCM purposes.

The third type of specimen used for CTE measurement was a FC BGA substrate. The dimensions were $27 \times 27 \times 0.55$ mm. The average CTE value obtained from four measurements within the temperature range of 50.8°C to 82.5°C was 18.85 ppm/°C. This value is close to 17.45 ppm/°C, the measured CTE value for the PBGA substrate shown in Figure 24.5. The measurement results obtained within different temperature ranges are shown in Figure 24.8.

The CTE value for the 40×40 mm FC BGA substrate used for warpage measurement as described in Section 24.5 was measured indirectly. This substrate is classified as a type 4 substrate. Instead of using the substrate directly for CTE measurement, the FC BGA package was used. The chip in it had dimensions of $20 \times 20 \times 0.73$ mm; i.e., it was a Group 2 package as shown in Table 24.1. The in-plane deformation of the specimen was measured using the developed in-plane ESPI method (Figure 24.3). The advantage of ESPI over traditional measurement methods, such as using strain gauges, is its high sensitivity. In this case, two measurement runs, at ΔT of 1.0°C and 1.3°C, were performed. The latter run was then normalized to 1.0°C for comparison purposes. The results are plotted in Figure 24.9. The good correlation between these two measurements demonstrates the repeatability of the experiment. The largest elongation value is found to be 3.92 μm, which was obtained between two sides of the substrate. Therefore, the equivalent CTE value, which was calculated by dividing the measured elongation of the package by its in-plane dimension, 40 mm, was 9.79 ppm/°C.

In order to obtain the CTE value of the build-up substrate in this measured flip chip package, 3D finite element analysis was performed. The material data listed in Tables 24.2 and 24.3 were used as the input data except for the CTE value of the substrate, which was

obtained using this finite element model. For a 1°C change, the elongation was calculated by plugging in different CTE values of the substrate. The equivalent CTE value was then obtained by dividing the packaging elongation by its in-plane dimension, 40 mm. It was found that the optimal CTE value of the substrate was equal to 11.9 ppm/°C. The value was be very close to the CTE value of 12 ppm/°C provided by the manufacturer. This excellent agreement also indicates that the material data provided by the manufacturer are reliable.

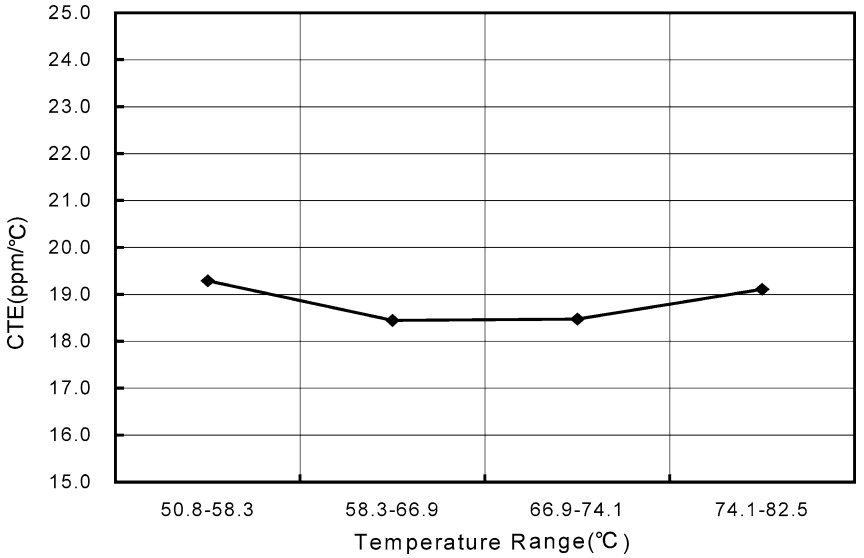


FIGURE 24.8. CTE values for 0.56 mm thick FC substrate.

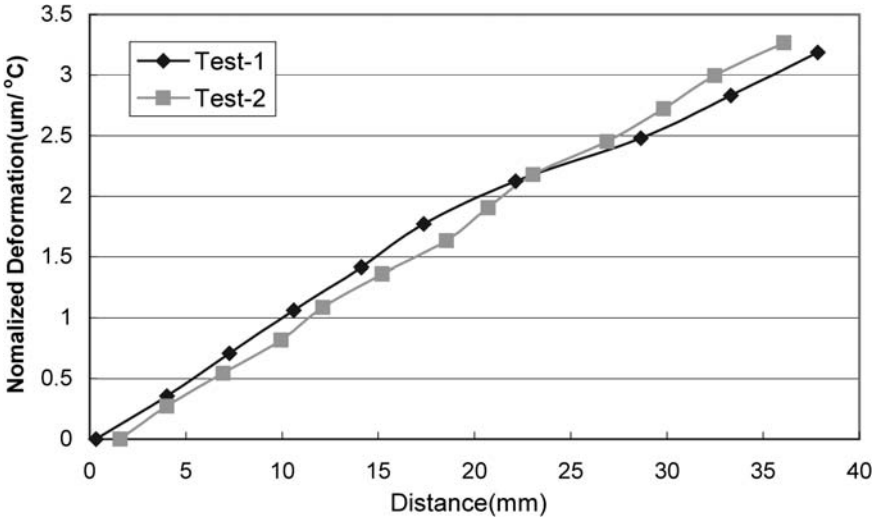


FIGURE 24.9. In-plane deformation of flip chip package under a 1.0°C temperature change.

24.5. BEHAVIOR OF FLIP CHIP PACKAGES UNDER THERMAL LOADING

24.5.1. Warpage at Room Temperature

Figure 24.10 shows the warpage measurement results obtained using the shadow moiré method described in Section 24.3. The flip chip packages were divided into four groups. The dimensions of the each group of flip chip packages are listed in Table 24.1. The plots in Figures 24.10(a) are the warpage profiles measured on the chip side. Therefore, the linear dimensions of Groups 1, 2, 3, and 4 were 10 mm, 20 mm, 26 mm, and 26 mm, respectively. On the other hand, the plots in Figures 24.10(b) are the warpage profiles measured on the substrate side. Therefore, the linear dimension was always 40 mm. The warpage profiles of the chips in all four groups show an axisymmetrical pattern regardless of the difference for the chip size used. This axisymmetrical pattern was not observed in the warpage profile measured on the substrate side [Figure 24.10(b)]. In addition, when the chip size became larger, such as 20 × 20 mm and 26 × 26 mm, the warpage profiles had rhombus shapes.

Table 24.5 lists the warpage data for the four groups of flip chip packages obtained at room temperature. The chip center was the reference point used to calculate the warpage values. Warpage 1 refers to the average warpage measured in the horizontal and vertical directions, and warpage 2 refers to the average warpage measured in the ±45 degree directions. The warpage values of the chips were smaller than the warpage values of the substrates, and the values corresponding to warpage 1 were always smaller than the values corresponding to warpage 2, as expected. On the other hand, a comparison of the Groups 1, 2, and 3 results reveals that the amount of warpage increased as the chip size increased. A comparison of the Group 3 and Group 4 results also shows that warpage increased as the chip thickness decreased.

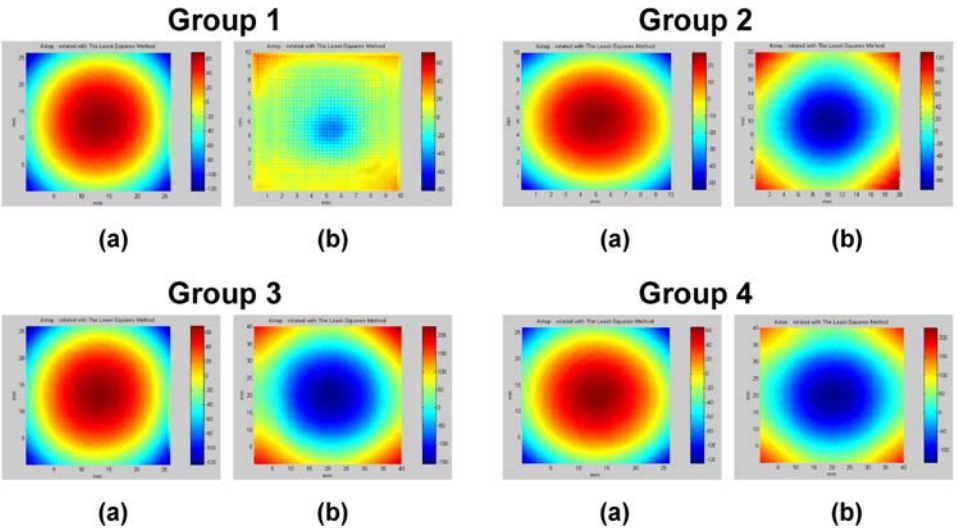


FIGURE 24.10. Warpage of flip chip packages measured at room temperature. (a) Measured on the chip side; and (b) measured on the substrate side.

TABLE 24.5.
Warpage measured at room temperature for chip and substrate of flip chip packages.

	Chip		Substrate	
	Warp. 1	Warp. 2	Warp. 1	Warp. 2
Group 1	12.6	21.2	48.4	65.0
Group 2	41.9	79.1	118	205
Group 3	75.2	143	159	271
Group 4	98.6	188	193	336

24.5.2. Warpage at Elevated Temperatures

The warpage profiles for a flip chip package having a 20×20 mm chip were then obtained at different temperatures. Figures 24.11 plots the warpage profiles measured on the chip side at temperatures of 26°C (room temperature), 50.7°C, 73.8°C, 101°C, 121°C, and 145°C. The warpage of the chip became smaller as the temperature increased. When the temperature approached 145°C, no warpage of the chip was observed. Wrinkles observed in the warpage profile when the warpage values were small were due to variation of the stepping motor, which was used to control the motion of the reference gratings for phase shifting purposes.

The warpage of the same flip chip package was then measured on the substrate side. The results are shown in Figures 24.12. The measurements were performed at 25.7°C, 48.0°C, 72.0°C, 97.0°C, 127°C, 151°C, 176°C, 201°C, and 228°C. Similar to the warpage results obtained from the chip side, the warpage values decreased as the temperature increased. When the temperature was approximately 150°C, the package became essentially flat (with a maximum warpage of 47.7 μm), which implied that this was the temperature at which the package was nearly stress free. As the temperature continued to increase from 176°C to 201°C, the warpage profile changed from concave upward to concave downward, although the warpage values were smaller and the profile shapes became more irregular. This is because the temperature had exceeded the glass transition temperature of approximately 180°C. As a result, the stiffness of the substrate dropped to a value too small to support its own weight. This phenomenon was found to be even more significant at 228°C. Consequently, we conclude that it is important to not keep a package at elevated temperatures for a long period of time, as always suggested in the reflow profile for a PBGA package in the assembly and packaging industry. On the other hand, by investigating the warpage profile carefully, it was found that there existed a slight profile change at the edge of the chip. This is considered to be due to the reinforcement imposed by the fillet of the underfill. Due to the large CTE of the underfill, tensile force was induced in the fillet when the temperature dropped below 150°C, i.e., the stress free temperature. As a result, the tensile force pulled on the substrate and induced this slight profile change.

The maximum warpage value at each temperature for each flip chip package in Groups, 1, 2, 3, and 4 are shown in Figures 24.13, 24.14, 24.15, and 24.16, respectively. The warpage measurements on the chip side were performed at temperatures lower than 150°C, mainly to keep the package from being damaged due to excessive heating at elevated temperatures. A linear relationship between the warpage measured on the chip side and the corresponding temperatures was observed in all four groups of packages. Furthermore, the chip warpage was found to approach to zero at approximately 150°C. Comparing the results for Groups 1, 2, and 3, one can see that the warpage increased as the

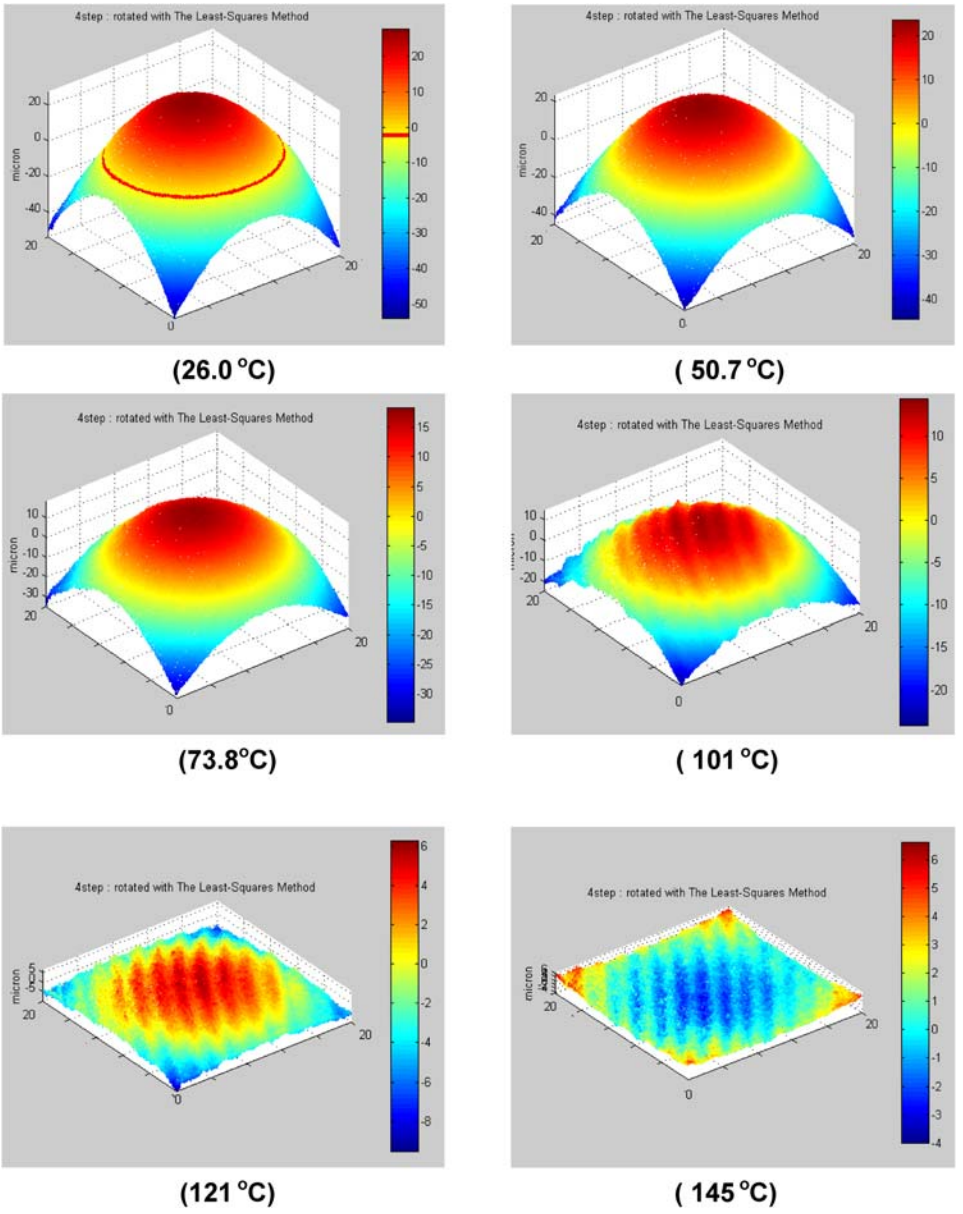


FIGURE 24.11. Recorded warpage profiles measured on the chip side of the FC package at 26.0°C, 50.7°C, 73.8°C, 101°C, 121°C, and 145°C. The chip size was 20 × 20 mm.

chip size increased. On the other hand, a comparison of the results for Groups 3 and 4 reveals that the warpage remained the same regardless of the change of the chip thickness.

A linear relationship was also observed between the warpage measured on the substrate side and the corresponding temperatures. For the flip chip packages with 10 × 10 mm and 20 × 20 mm chips (Groups 1 and 2), this linear relationship was found for a temperature

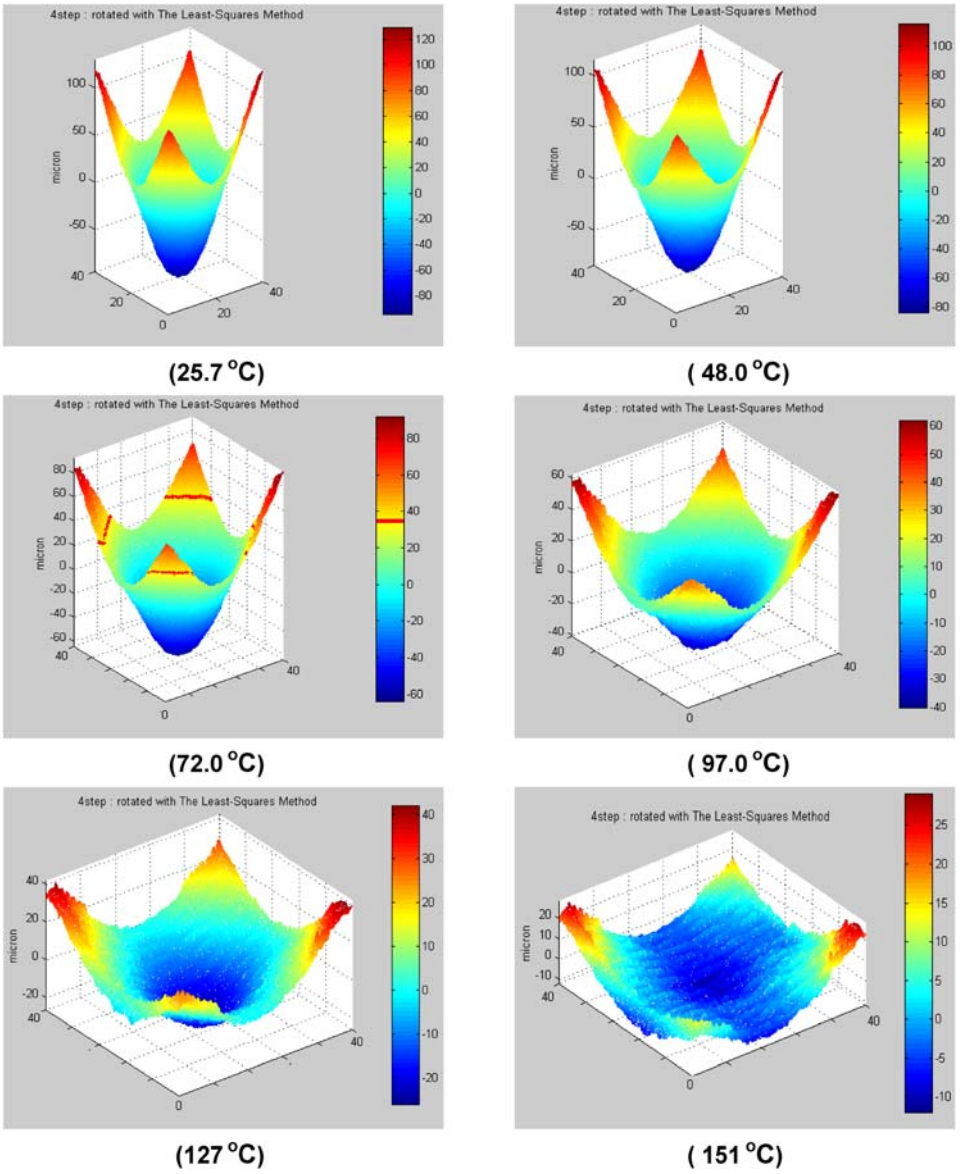


FIGURE 24.12. Recorded package warpage profiles at 25.7°C, 48.0°C, 72.0°C, 97.0°C, 127°C, 151°C, 176°C, 201°C, and 228°C. The chip size was 20 × 20 mm. Warpage measurements were performed on the substrate side.

range of from room temperature to 225°C. On the other hand, due to the constraint from the 26 × 26 mm chips on the flip chip packages in Groups 3 and 4, the substrate warpage remained unchanged as the temperature passed the glass transition temperature of the BT substrate, as shown in Figures 24.15 and 24.16. When the temperature dropped to room temperature, it was found that the warpage values measured on the substrate side became larger when the chip size increased and the chip thickness decreased.

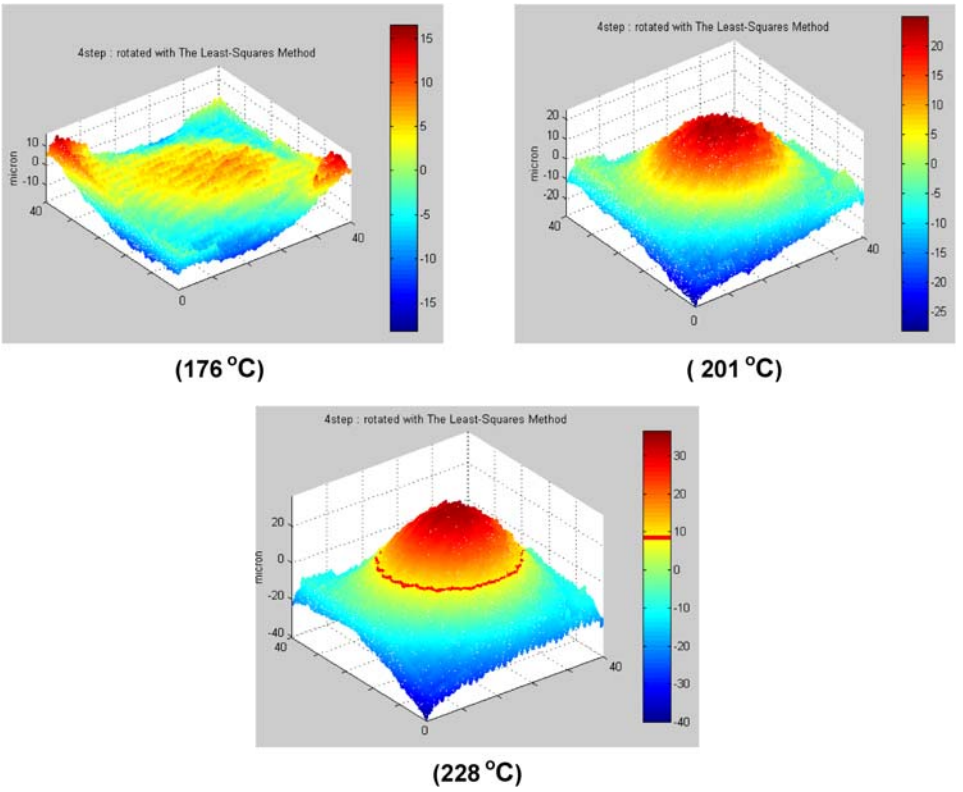


FIGURE 24.12. (Continued.)

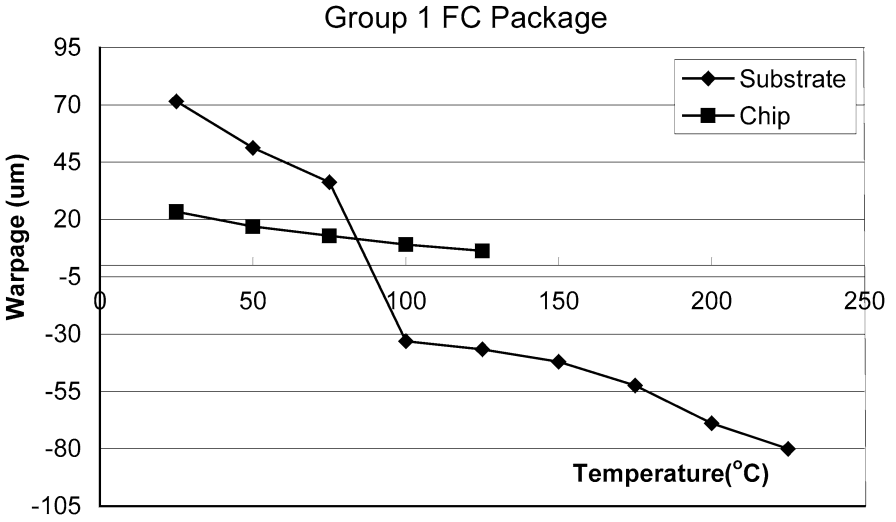


FIGURE 24.13. Warpage of the chip and substrate in the flip chip packages in Group 1.

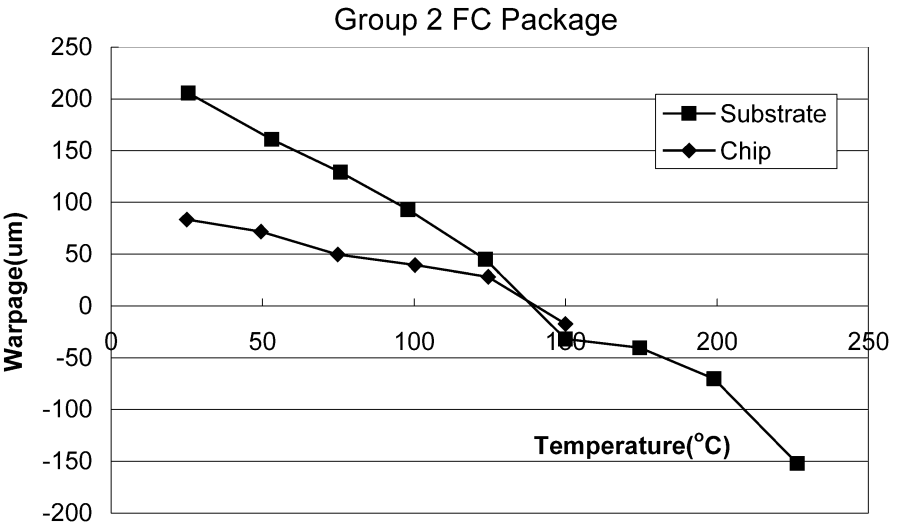


FIGURE 24.14. Warpage of the chip and substrate in the flip chip packages in Group 2.

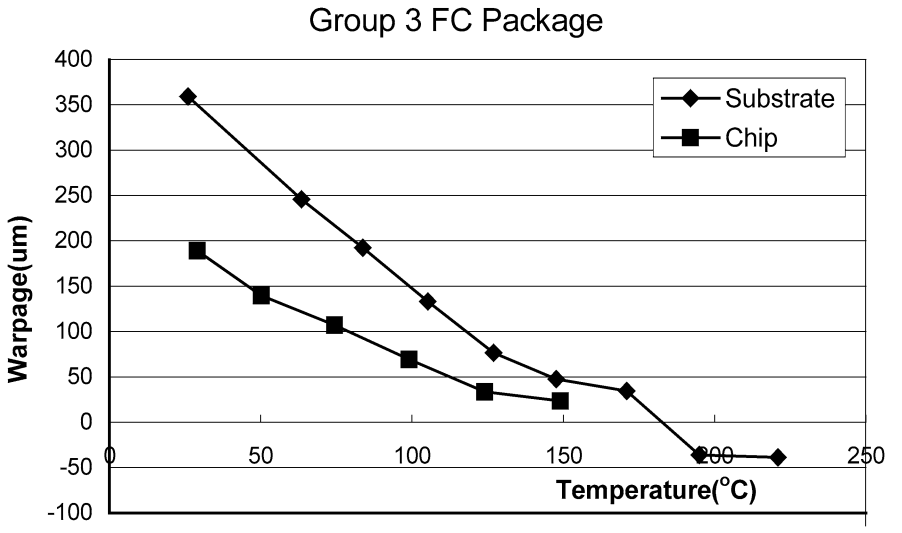


FIGURE 24.15. Warpage of the chip and substrate in the flip chip packages in Group 3.

24.5.3. Effect of Underfill on Warpage

In order to investigate the effect of underfill on the warpage behavior of flip chip packages, warpage in two 27 × 27 mm FC packages with 10 × 10 mm chips was measured using the developed out-of-plane ESPI method (Figure 24.4). The results are shown in Figures 24.17 and 24.18 for flip chip packages without and with underfill reinforcement, respectively. In both figures, two measurement results for the same packages are plotted. It was found that the relationship between the package warpage and the corresponding temperatures was not linear when the package was not reinforced by underfill. When the packaged was reinforced by underfill, the warpage to temperature relationship became linear,

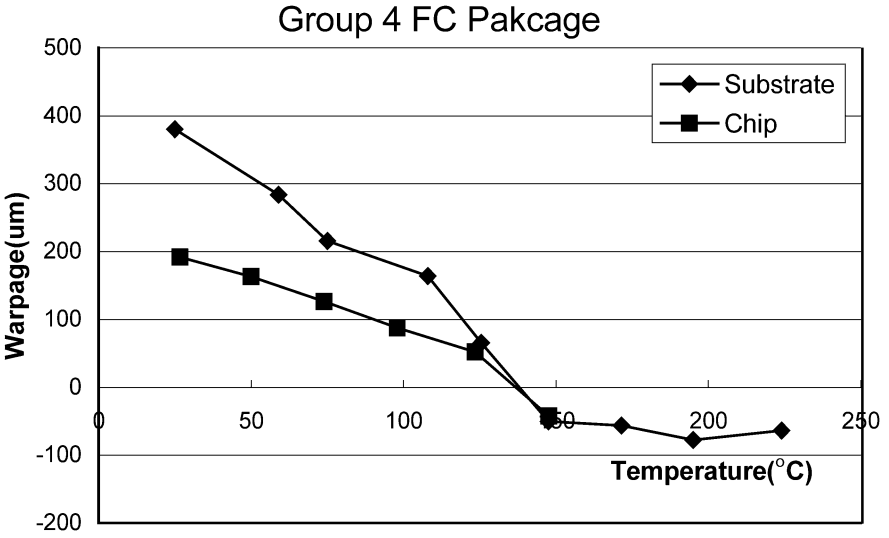


FIGURE 24.16. Warpage of the chip and substrate in the flip chip packages in Group 4.

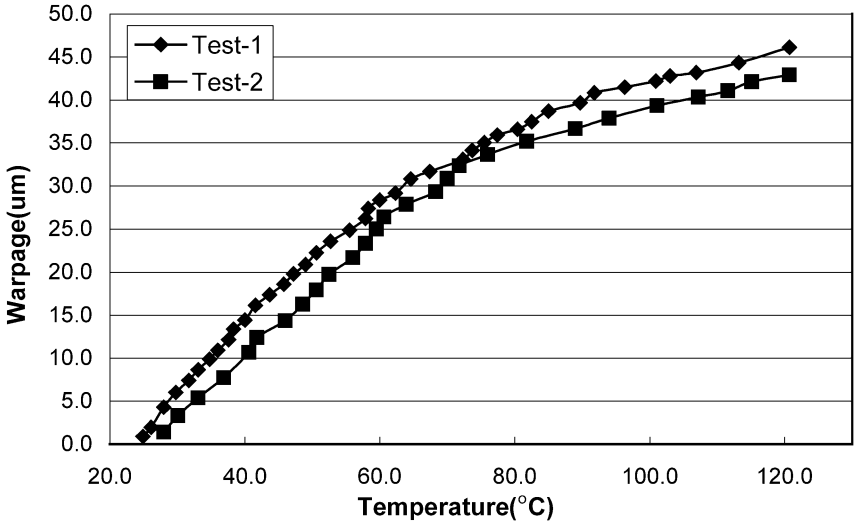


FIGURE 24.17. Warpage versus temperature for a 27 × 27 mm FC package without underfill reinforcement.

and the values also became larger over the same temperature range. For example, when the temperature increased from room temperature to 60°C, a linear relationship was observed between warpage and temperature for both packages, but the warpage in the package with underfill was 1.1 times larger than that in the package without underfill. When the temperature reached 120°C, this ratio increased to 2.3.

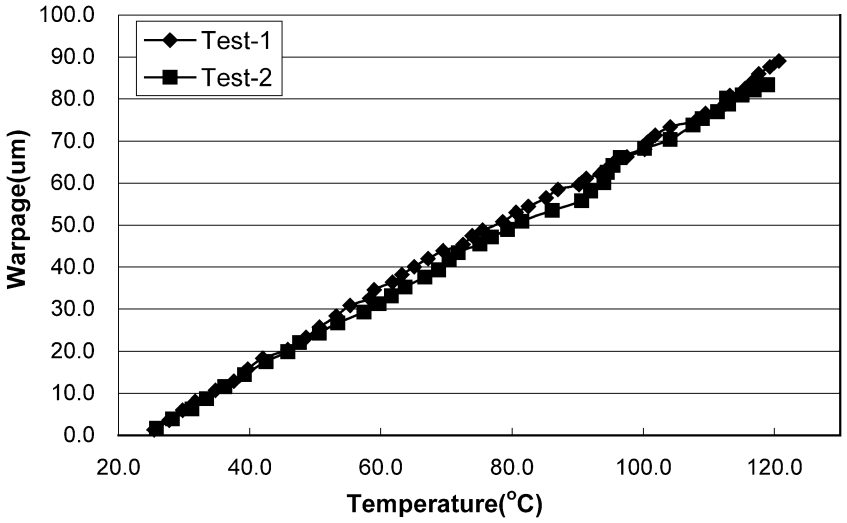


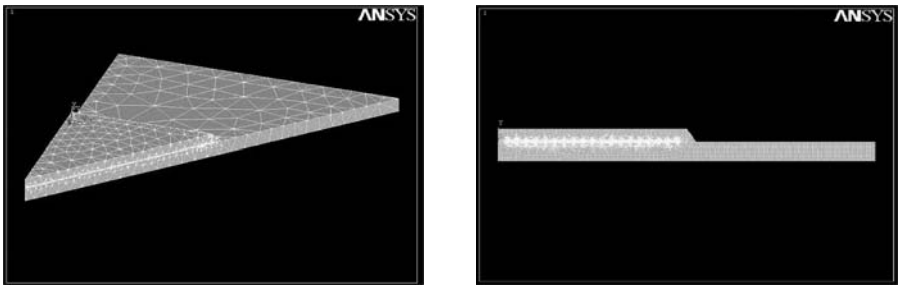
FIGURE 24.18. Warpage versus temperature for a 27 × 27 mm FC package with underfill reinforcement.

24.6. FINITE ELEMENT ANALYSIS OF FLIP CHIP PACKAGES UNDER THERMAL LOADING

Both two-dimensional and three-dimensional finite element models were constructed using ANSYS for warpage analysis of a flip chip package. The material properties adopted in these models are listed in Table 24.2 and Table 24.3. The CTE value and the Young's modulus for the $1+4+1$ build-up substrate were the two most sensitive properties used in analysis. The CTE value was verified by measurement, as described in Section 24.4. The Young's modulus of the substrate was verified in [38]. For 3D analyses, a $1/8$ model was constructed using a 10-node tetrahedral element (SOLID45) to avoid erroneous spurious shear strain [39]. The total number of elements used was 69,659. For 2D analyses, a $1/2$ model was constructed. The element used was PLANE82. The total number of elements was 11,571. The finite element models discussed in this section were constructed based on the geometry of a flip chip package assembled using a 20×20 mm chip. The underfill at the edge of the chip, called the fillet, was modeled so that the height would be equal to a chip thickness of 730 μm and a tail length of 500 μm. According to the experimental results discussed in Section 24.5, the stress free temperature was 150°C. Therefore, the thermal loading applied in the finite element models was based on a temperature change from 150°C to 26°C. The material properties were assumed to remain constant during the change of temperature.

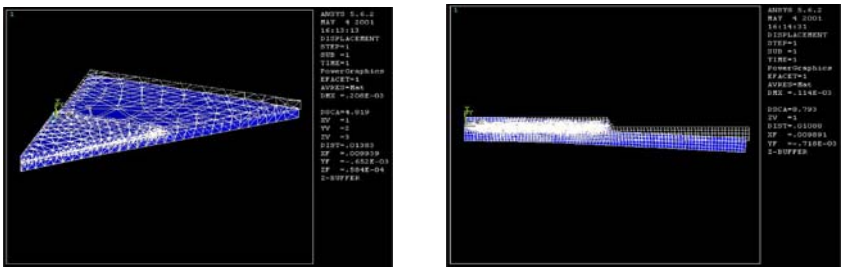
Figure 24.19(a) and (b) plot the constructed 3D and 2D finite element models, respectively. The corresponding warpage profiles of the two models are depicted in Figure 24.20(a) and (b).

A comparison of the experimental and numerical data for the maximum warpage values, i.e., at warpage 1 and warpage 2, is shown in Table 24.6. The definitions of warpage 1 and warpage 2 are the same as those given in Section 24.5.1. The correlation is considered to be satisfactory. Therefore, both the 2D and 3D finite element models are considered to be accurate when adopted for warpage analysis purposes.



(a) (b)

FIGURE 24.19. Plots of finite element models for (a) 3D and (b) 2D analyses. The flip chip package had a 20×20 mm chip.



(a) (b)

FIGURE 24.20. Warpage profiles of the models plotted in Figure 24.19 for (a) 3D and (b) 2D analyses when the temperature changed from 150°C to 26°C.

TABLE 24.6.
Comparison of the measured data and numerical warpage predictions obtained using 2D and 3D finite element models.

		Shadow moiré	3D FEM		2D FEM	
		Result	Result	Difference	Result	Difference
Chip	Warp. 1	41.9	45.7	8.84%	43.5	3.60%
	Warp. 2	79.0	87.5	10.7%	N/A	N/A
Substrate	Warp. 1	118	118	0%	109	−7.63%
	Warp. 2	205	203	−0.86%	N/A	N/A

24.7. PARAMETRIC STUDY OF WARPAGE FOR FLIP CHIP PACKAGES

In this section, the results of the parametric analysis will be used to evaluate the effects of the (a) chip thickness, (b) substrate thickness, (c) Young's modulus of the underfill, (d) CTE of the underfill, and (e) geometry of the underfill fillet on the warpage behavior of flip chip packages. The studied chip thickness ranged from 0.119 mm to 1.793 mm. The substrate thickness ranged from 0.3 mm to 1.5 mm. The Young's modulus and CTE of the underfill ranged from 5.0 GPa to 30 GPa and from 15.0 ppm/°C and 52.5 ppm/°C, respectively. The 2D finite element model described in Section 24.6 was employed in this study. The analyzed chip and the substrate had square in-plane dimensions, which were

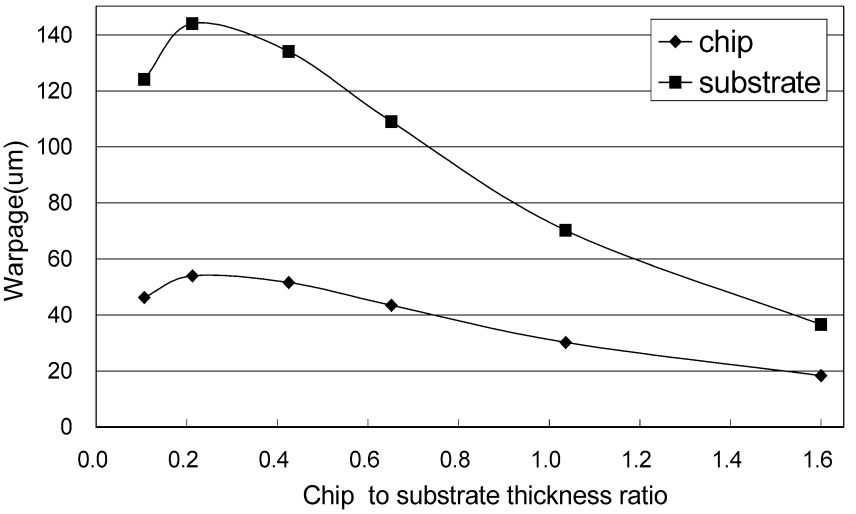


FIGURE 24.21. Effect of chip thickness change on flip chip warpage.

20 × 20 mm and 40 × 40 mm, respectively. The analyses were performed while the temperature of the flip chip package changed from 150°C to 26°C. The baseline model had chip and substrate thicknesses of 0.73 mm and 1.12 mm, respectively, and the Young's modulus and CTE of the underfill were 5 GPa and 45 ppm/°C, respectively. For each parametric analysis, only one parameter was changed.

24.7.1. Change of the Chip Thickness

The results of the warpage analyses of the chip and substrate in a flip chip package are plotted in Figure 24.21. The chip thickness for the baseline analysis was 0.73 mm. The analyses covered change of the chip thickness from 0.16 to 2.46 mm. When the chip to substrate ratio was reduced to 0.2, the maximum warpage values for both the chip and substrate were obtained. When the chip thickness increased, the warpage values for both the chip and substrate decreased due to the increase of the stiffness of the chip. On the other hand, when the ratio became smaller, the warpage values for both the chip and substrate dropped again. This was because the stiffness of the chip became small enough to comply with the deformation of the substrate.

24.7.2. Change of the Substrate Thickness

The substrate thickness for baseline analysis was 1.12 mm. The substrate thickness changed from 0.3 mm to 1.5 mm, which corresponded to ratios of 0.3 to 1.34, compared to the baseline substrate. The results are plotted in Figure 24.22. The change of warpage of the chip was found to be insignificant. On the substrate side, the maximum warpage occurred when the substrate to chip thickness ratio was 1.25. As the substrate thickness decreased from 1.12 mm to 0.3 mm, the warpage decreased progressively. This was because the stiffness of the substrate became so small that its influence on the change of warpage became progressively smaller. On the other hand, when the thickness of the substrate increased, the effect of the increase of the substrate stiffness on the decrease in the warpage was smaller

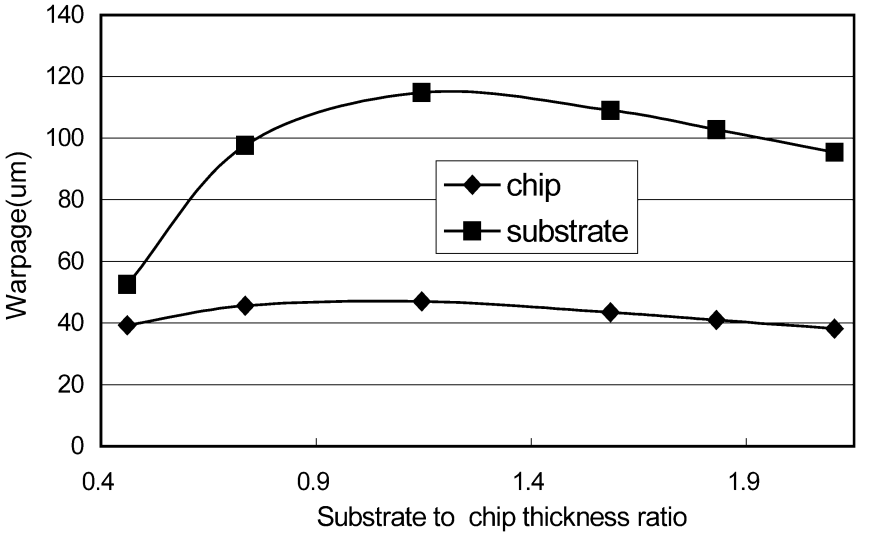


FIGURE 24.22. Effect of the substrate thickness change on the flip chip warpage.

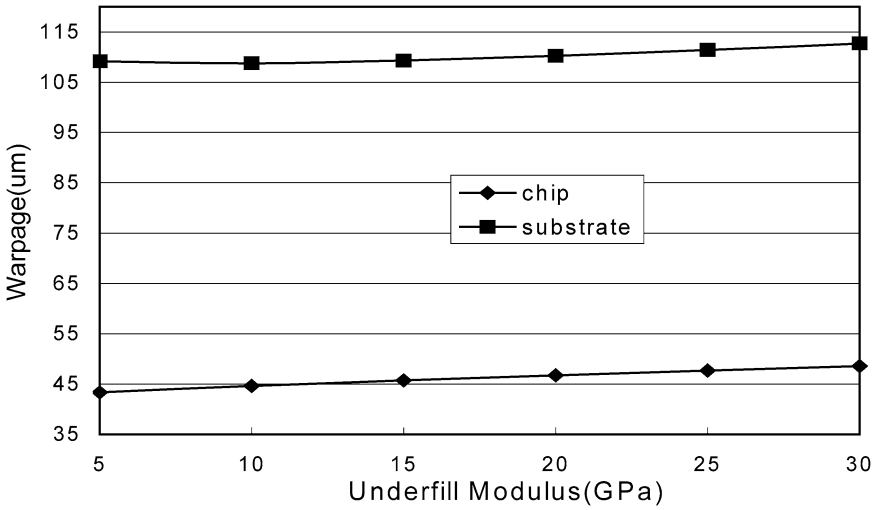


FIGURE 24.23. Effect of the change of the Young's modulus of the underfill on the flip chip warpage.

than the effect of the increase of the chip thickness (Figure 24.21). This was because the stiffness of the chip was much higher than that of the substrate.

24.7.3. Change of the Young's Modulus of the Underfill

The Young's modulus of the underfill used in the baseline analysis was 5 GPa. In this part of the study, the Young's modulus changed from 5 GPa to 30 GPa, i.e., a six-fold increase. The results are plotted in Figure 24.23. The warpage on both the chip and substrate sides increased almost linearly with the increase of the Young's modulus of the underfill. The experimental results plotted in Figures 24.17 and 24.18 confirmed this prediction. As

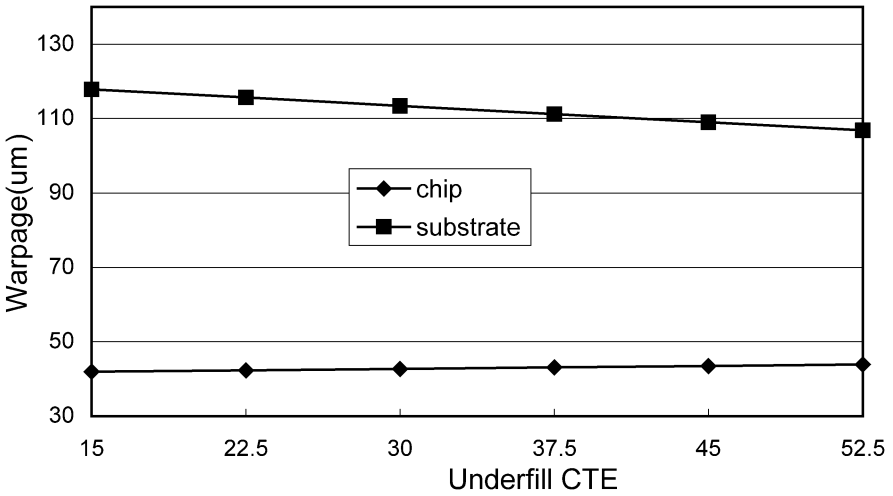


FIGURE 24.24. Effect of the change of the CTE of the underfill on the flip chip warpage.

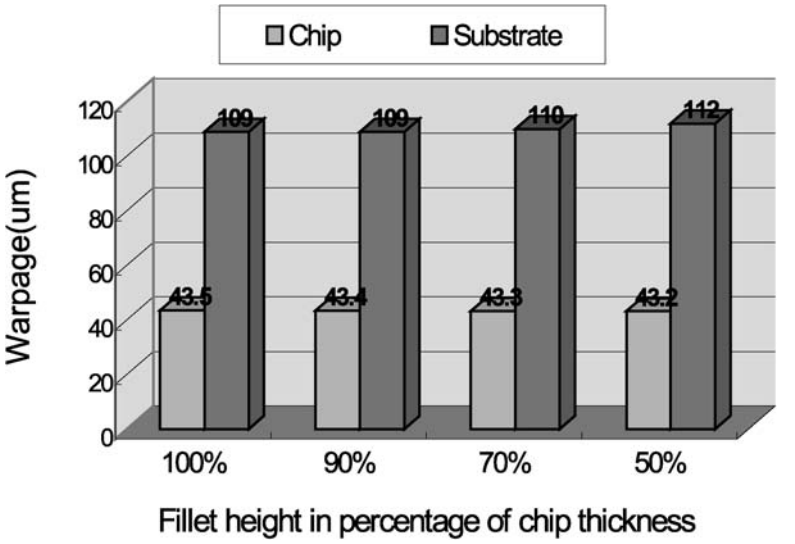
shown in Figure 24.18, when the temperature increased, the warpage of the package with underfill increase more than the warpage of the package without underfill. This was because the solder joints functioned as cushions, absorbing the deformation due to the CTE mismatch between the substrate and chip. Once the underfill filled the gap between the chip and substrate, the deformation of the solder joints was constrained. As a result, the deformation induced by the CTE mismatch between the chip and substrate resulted in global warpage of the package because the deformation of the solder joints was now constrained by the surrounding underfill. Therefore, if the Young's modulus of the underfill increased, the warpage of both the chip and substrate in the flip chip package also had to increase, as evidenced by the results shown in Figure 24.23.

24.7.4. Change of the CTE of the Underfill

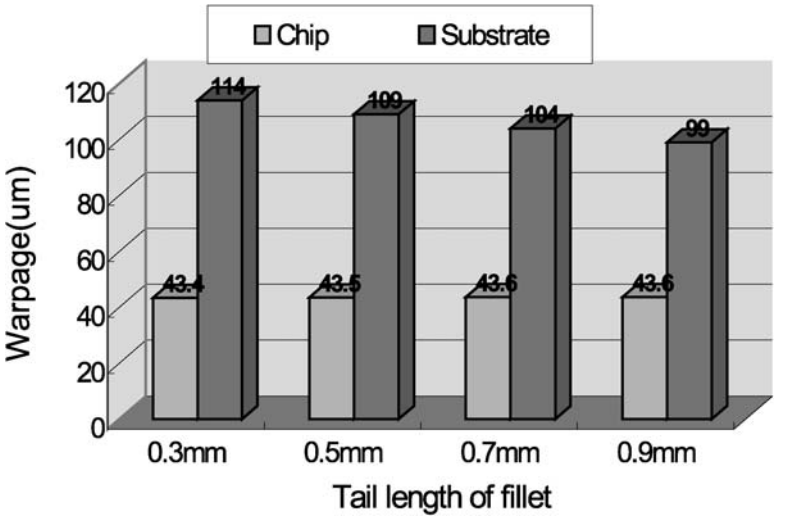
The CTE value for baseline analysis was 45 ppm/°C. The analysis covered CTE values ranging from 15 ppm/°C to 52.5 ppm/°C. The results are plotted in Figure 24.24. The change of the flip chip warpage was not significant. However, it is interesting that the warpage change trends of the chip and substrate were different. This was due to the underfill at the edge of the chip, i.e., at the fillet location. The fillet was modeled so as to have an inclined surface. The height and tail length of the fillet were 0.732 mm and 0.5 mm, respectively. When the CTE of the underfill increased, due to the temperature change from 150°C to room temperature, the tensile force induced by contraction at the fillet increased. As a result, the substrate warpage decreased and the chip warpage increased.

24.7.5. Effect of the Geometry of the Underfill Fillet

The height and tail of the fillet used in the baseline analysis were 0.732 mm and 0.5 mm, that is, 100% and 68.5% of the thickness of the chip, respectively. In this part of parametric analysis, the heights of the underfill fillet were selected to be 100%, 90%, 70%, and 50% of the flip chip thickness. On the other hand, the lengths of the underfill tail were selected to be 0.3 mm, 0.5 mm, 0.7 mm, and 0.9 mm. The results are plotted in



(a)



(b)

FIGURE 24.25. Effect of the change of the geometry of the underfill fillet on the flip chip warpage. (a) Change of the fillet height; and (b) change fillet the tail length.

Figure 24.25(a) and (b). It is found that the chip warpage remained essentially the same regardless of the geometric change of the fillet. On the other hand, when the fillet height increased from 50% to 100% of the chip thickness, the substrate warpage decreased only 3%. Also, when the tail length increased from 0.3 mm to 0.9 mm, the substrate warpage decreased by 15%. These results are considered to be reasonable because, when the height and tail length increased, the tensile force induced by fillet contraction, which, in turn, was induced by the decrease in temperature, also increased. As a result, the substrate warpage

decreased. The insignificant change in chip warpage was mainly due to the fact that the chip stiffness was much larger than the substrate stiffness.

24.8. SUMMARY

In this chapter, the mechanical behavior of flip chip packages has been reported in detail. The phase shifted shadow moiré approach was adopted to measure the warpage of flip chip packages, and both the in-plane and out-of-plane ESPI methods were adopted to measure the CTE values of the substrate and the warpage of flip chip packages. Meanwhile, two-dimensional and three-dimensional finite elements models were constructed for numerical analysis of flip chip packages. It was found that:

(a) The substrate CTE values varied significantly due to the use of different BT cores, numbers of vias drilled in the substrate, and thicknesses of the copper and the solder mask. Therefore, the CTE of each substrate needs to be carefully measured so that the mechanical behavior of the corresponding flip chip package can be accurately determined.
(b) At room temperature, warpage of flip chip packages increased with increasing chip size and decreasing chip thickness. The warpage of the flip chip package also increased after the underfill reinforcement, and the relationship between the package warpage and the temperature changed from nonlinear to linear.
(c) The flip chip warpage was found to decrease to zero when the temperature increased to around 150°C. Linear relationships between the warpage values, for both the chip and the substrate, and the corresponding temperatures were observed. When the temperature increased further, the amount of substrate warpage gradually approached to the amount of chip warpage, especially when the chip size became large, as in the packages with 20×20 mm and 26×26 mm chips.
(d) The parametric analysis showed that a minimum warpage value for a flip chip package could be obtained by selecting a chip thickness that was much smaller than the thickness of the substrate, by choosing underfill materials with small Young's modulus and large CTE values, and by selecting an underfill fillet with a long tail.

REFERENCES

1. K. Wang, et al., Interfacial shear stress, peeling stress, and die cracking stress in trilayer electronics assembles, 2000 Inter Society Conference on Thermal Phenomena, 2000, pp. 56–64.
2. H.B. Fan, M.M.F. Yuen, and E. Suhir, Prediction of delamination in a bi-material system based on free-edge energy evaluation, IEEE Conference on Electronic Components and Technology Conference, 2003, pp. 1160–1164.
3. E. Suhir and J.D. Weld, Application of a surrogate layer for lower bending stress in a tri-material body, IEEE Conference on Electronic Components and Technology Conference, 1996, pp. 435–439.
4. T.M. Robert, Shear and normal stresses in adhesive joints, J. Eng. Mech., 115(11), pp. 2460–2479 (1989).
5. E. Suhir, Flip-chip solder joint interconnections and encapsulants in silicon-on-silicon MCM technology: thermally induced stresses and mechanical reliability, Multi-Chip Module Conference, 1993, pp. 92–99.
6. P. Palaniappan, et al., Correlation of flip chip underfill process parameters and material properties with in-process stress generation, Electronic Component and Technology Conference, 1998, pp. 838–847.
7. Y. Guo, S.M. Kuo, and L. Mercado, Stress/strength and reliability evaluations on UBM in different solder systems, 2000 Inter Society Conference on Thermal Phenomena, 2000, pp. 193–199.

8. W. Engelmaier, The use environments of electronic assembles and their impact on surface mount solder attachment reliability, IEEE Transaction on Components, Hybrids, and Manufacturing Technology, 13(4), pp. 903–908 (1990).
9. E. Suhir, Predicted failure criterion (von-Mises stress) for moisture-sensitive plastic packages, IEEE Conference on Electronic Components and Technology Conference, 1995, pp. 266–284.
10. J.H. Zhao, X. Dai, and P.S. Ho, Analysis and modeling verification for thermal-mechanical deformation in flip-chip package, Electronic Components and Technology Conference, 1998, pp. 336–344.
11. K.S. Beh, A. Ourdjini, V.C. Venkatesh, and Y.L. Khong, Finite element analysis of substrate warpage during die attach process, Electronic Materials and Packaging, 2002, pp. 94–98.
12. A.O. Ayhan and H.F. Nied, Finite element modeling of interface fracture in semiconductor packages: issue and applications, 1998 Inter Society Conference on Phenomena, 1998, pp. 185–192.
13. T. Lee, J. Lee, and I. Jung, Finite element analysis for solder ball failures in chip scale package, Microelectronics Reliability, 38, pp. 1941–1947 (1998).
14. K.N. Chiang, Y.T. Lin, and H.C. Cheng, On enhancing eutectic solder joint reliability using a second-reflow-process approach, IEEE Transactions on Advanced Packaging, 23(1), pp. 9–14 (2000).
15. M.J. Pfeifer, Solder bump size and shape modeling and experimental validation, IEEE Transaction on Components, Packaging, and Manufacturing Technology, Part B, 20(4), pp. 452–457 (1997).
16. J.H. Lau, R. Lee, and C. Chang, Effect of underfill material properties on the reliability of solder bumped flip chip on board with imperfect underfill encapsulants, IEEE Transactions on Components and Packaging Technologies, 23(2), pp. 323–333 (2000).
17. C. Bailey, H. Lu, and D. Wheeler, Computer based modeling for predicting reliability of flip-chip components on printed circuit boards, 1999 IEEE/CPMT Int. Electronics Manufacturing Technology Symposium, 1999, pp. 42–49.
18. E.C. Ahn, T.J. Cho, J.B. Shim, H.J. Moon, J.H. Lyu, K.W. Choi, S.Y. Kang, and S.Y. Oh, Reliability of flip chip BGA package on organic substrate, Electronic Components and Technology Conference, 2000, pp. 1215–1220.
19. L. Leicht and A. Skipor, Mechanical cycling fatigue of PBGA package interconnects, Microelectronics Reliability, 40, pp. 1129–1133 (2000).
20. A. Mertol, Application of the Taguchi method to chip scale package (CSP) design, IEEE Transaction on Advanced Packaging, 23(2), pp. 266–276 (2000).
21. V.V. Calmidi and R.L. Mahajan, Optimization for thermal and electrical performance for a flip-chip package using physical-neural network modeling, Electronic Components and Technology Conference, 1997, pp. 1163–1169.
22. A.O. Cifuentes and I.A. Shareef, Modeling of multilevel structures: a general method for analyzing stress evolution during processing, IEEE Transaction on Semiconductor Manufacturing, 5(2), pp. 128–137 (1992).
23. K.S. Chen, T.Y.F. Chen, C.C. Chuang, and I.K. Lin, Full-field wafer level thin film stress measurement by phase-stepping shadow moiré, IEEE Transactions on Components and Packaging Technologies, 27(3), pp. 594–601 (2004).
24. Y.Y. Wang and P. Hassell, Measurement of thermally induced warpage of BGA packages/substrates using phase-stepping shadow moiré, Electronic Packaging Technology Conference, 1997, pp. 283–289.
25. S. Chen, C.Z. Tsai, E. Wu, and C.A. Shao, A study on effect of flip-chip BGA warpage and stresses using the different chip sizes, The 25th Conference on Theoretical and Applied Mechanics, 2001 (in Chinese).
26. Z.S. Chen, C.Z. Tsai, E. Wu, and C.A. Shao, Stress measurement and analysis of flip-chip BGA, SEMICON Taiwan 2001, Packaging & Testing Seminar, pp. 245–254.
27. M.R. Stiteler, I.C. Ume, and B. Leutz, In-process board warpage measurement in a lab scale wave soldering oven, IEEE Transaction on Components, Packaging, and Manufacturing Technology, Part A, 19(4), pp. 562–569 (1996).
28. D. Zwemer, et al., PWB warpage analysis and verification using an AP210 standards-based engineering framework and shadow moiré, Thermal and Mechanical Simulation and Experiments in Microelectronics and Microsystems, EuroSimE 2004, 2004, pp. 121–131.
29. K.S. Beh, A. Ourdjini, V.C. Venkatesh, and Y.L. Khong, Finite element analysis of substrate warpage during die attach process, Electronic Materials and Packaging, 2002, pp. 94–98.
30. I. Tsai, C.Z. Tsai, E. Wu, C.A. Shao, On accurate measurement of warpage for electronic packages, Proceedings of IMAPS Taiwan Technical Symposium, 2001, pp. 290–297.
31. Y. Guo and J.H. Zhao, A practical die stress model and its applications in flip-chip packages, 2000 Inter Society Conference on Thermal Phenomena, 2000, pp. 393–399.
32. J. Woosoon, et al., Evaluation of thermal shear strains in flip-chip package by electronic speckle pattern interferometry (ESPI), Electronic Materials and Packaging, 2001, pp. 310–314.

33. J.R. Huang, H.D. Ford, and R.P. Tatam, Speckle techniques for material testing, Optical Techniques for Structural Monitoring, IEE Colloquium, 1995, pp. 6/1–6/6.
34. N. Kao, Z.C. Tsai, and E. Wu, The application of electronic speckle pattern interferometry (ESPI) on the PBGA package, The Sixteenth National Conference on Mechanical Engineering, The Chinese Society of Mechanical Engineers, 1999, Vol. 5, pp. 254–261 (in Chinese).
35. E. Wu and A.J.D. Yang, Simultaneously determining Young's modulus, coefficient of thermal expansion, Poisson ratio and thickness of mult-layered thin films on silicon wafer, Electronic Components and Technology Conference, 2004, Vol. 1, pp. 901–905.
36. S. Liu, J. Wang, D. Zou, X. He, Z. Qian, and Y. Guo, Resolving displacement field of solder ball in flip-chip package by both phase shifting moire interferometry and FEM modeling, Electronic Components and Technology Conference, 1998, pp. 1345–1353.
37. F.F. Beer and E.R. Johnston, Mechanics of Materials, second edition, McGraw-Hill Publ., New York, 1992.
38. Y.K. Chung, G.Y. Tzeng, and B.C. Chen, Determining the Mechanical Properties of PBGA Substrates using Inverse Method, Report NSC-88-2815-002-038-E, National Taiwan University, 1999 (in Chinese).
39. R.D. Cook, D.S. Malkus, and M.E. Plesha, Concepts and Applications of Finite Element Analysis, John Wiley and Sons, 1981.

25

Stress Analysis for Processed Silicon Wafers and Packaged Micro-devices

Li Li[a], Yifan Guo[b], and Dawei Zheng[c]

[a]*Cisco Systems, Inc., USA*
[b]*Skyworks Solutions, Inc., USA*
[c]*Kotura, Inc., USA*

25.1. INTRINSIC STRESS DUE TO SEMICONDUCTOR WAFER PROCESSING

For quality control and reliability analysis in semiconductor product development, it is important to be able to determine the intrinsic stresses in the devices and interconnects due to various wafer processes. These processing steps include thin film deposition, plating, patterning, etching and heat treatment [1]. The local intrinsic stress built up in the devices, which is closely related to the wafer processes, is one of the major causes of low manufacturing yield and early failures in semiconductor products.

Currently, there are no effective tools that can characterize and monitor the local intrinsic stress induced by wafer processes. Average intrinsic stresses in thin films on a wafer are measured indirectly by the FLEXUS machine (KLA-Tencor Corporation, San Jose, CA) based on the Stoney's equation [5–7]. The FLEXUS machine is widely used in thin film stress analysis treating whole wafers as substrates (the thin film carrier). This approach can only provide information about the intrinsic stress in thin films of large dimensions. It is not capable of determining the intrinsic stresses in the thin films of small dimensions, such as bond pads and UBMs (under bump metallurgy). Other existing methods for intrinsic stress measurements are the X-ray diffraction and Raman microscopy [8–11]. The best spatial resolution of the X-ray diffraction is about 30 μm, which is not enough to determine the stress distribution at a local area especially when the strain gradient is high. The Raman microscopy has a spatial resolution of about 1 μm. However, it is only effective for single crystal materials. And it does not provide whole-field stress maps, which are very important in the analysis of stress distributions. Both methods need a stress free state as reference. This is not always available under actual manufacturing conditions. In addition, both methods require sophisticated measuring instruments and typically can not be used for on-site monitoring and measuring during wafer processing.

In this section, a novel testing methodology to evaluate the whole-field intrinsic stress is presented. Using this methodology, a comprehensive study of a Ni plating process is conducted. The testing structure consists of a Si membrane with SiO_2 and $Si_xN_yH_z$ buffer

layer on top, which is used in the pressure sensor applications. The thin film material and the plating are processed on the top of the Si membrane using the regular device processing. The Si membrane deformation caused by the intrinsic stresses from the plating process is measured by an optical method. A finite element model is then used to calculate the intrinsic stress and the resulting full field stress distribution in the thin film and the Si membrane using the input from the measured membrane deformation.

25.1.1. Testing Device Structure

As shown in Figure 25.1, a sensing device wafer is, essentially, a pressure sensor wafer of specified membrane thickness. Either the whole pressure sensor wafer or a single sensing chip can be used for the monitoring and testing functions. The sensing wafer (or chip) is put into the processes which are under investigation or being monitored. During each material deposition or other process steps in the wafer process, the membrane will deform due to the process induced intrinsic stress.

In this study, an n-type (100) Si of 22 μm thick was grown onto a p-type (100) Si wafer of 380 μm thick as the starting substrate. 3500 Å thick SiO_2 and 4000 Å thick $Si_xN_yH_z$ were deposited on top of the Si wafer as insulating layers. The Si wafer was patterned using Si_3N_4 as the etching mask. The etching stopped automatically at the n-type Si layer where a Si membrane window of 2650×2650 μm^2 was formed at the center of a die of 4214×4214 μm^2 area. This process formed a membrane with a thickness of about 23 μm. This sensing device was used to determine the intrinsic stress developed during an electroless Ni plating process. In the process, Ni films with different sizes and thickness were plated at the central portion of the sensor membranes.

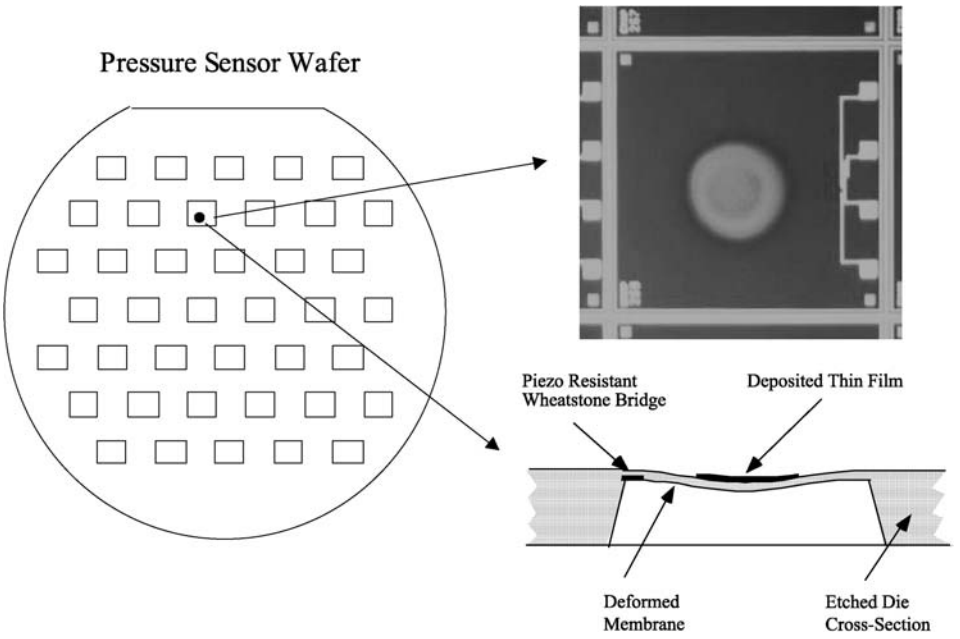


FIGURE 25.1. A sensing device wafer is essentially a Si wafer of specified membrane thickness. The membrane deformations are determined and monitored by an optical measurement technique.

25.1.2. Membrane Deformations

The deformations of the sensor membranes were determined using the Twyman-Green Interferometry technique as shown in Figure 25.2. The technique is a whole-field optical method with very high displacement sensitivity. It provides fringe patterns which are contour maps of out-of-plane displacements. The standard displacement resolution is 0.3165 μm which is one half of the wavelength of the laser used in the measurement. This optical technique has been used to measure static or dynamic deformations of Si membranes as a function of time and during progress of processes [12].

The membrane deformations were measured before and after the Ni plating process. Figure 25.3 shows the front and back view images of a membrane prior to the electroless Ni plating. The front side of the membrane is very flat. The back surface of the membrane has little variations caused by the thickness changes of the membrane as a result of the Si etching process. The Ni films are plated at the center of the Si membrane. In the fringe patterns, the out-of-plane displacement is determined from the fringe orders with a resolu-

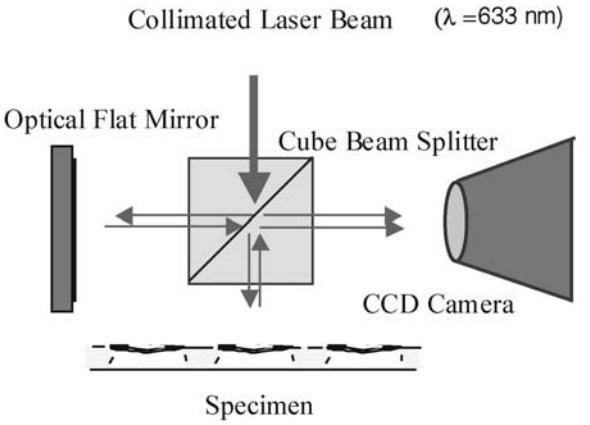


FIGURE 25.2. The deformation in the membranes of the sensor wafer is measured using the Twyman-Green interferometer.

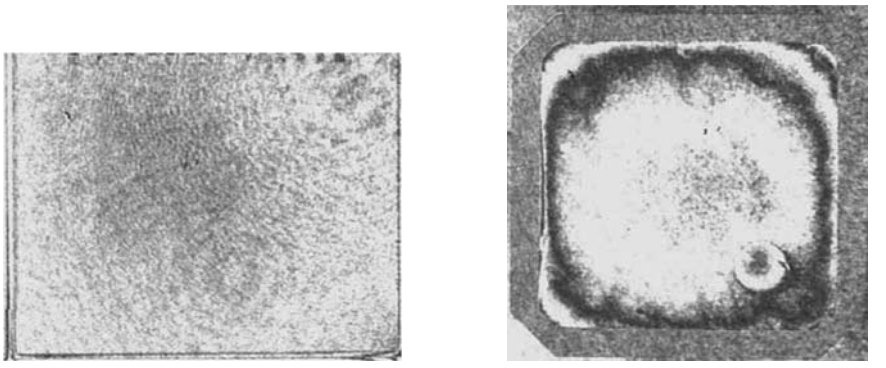
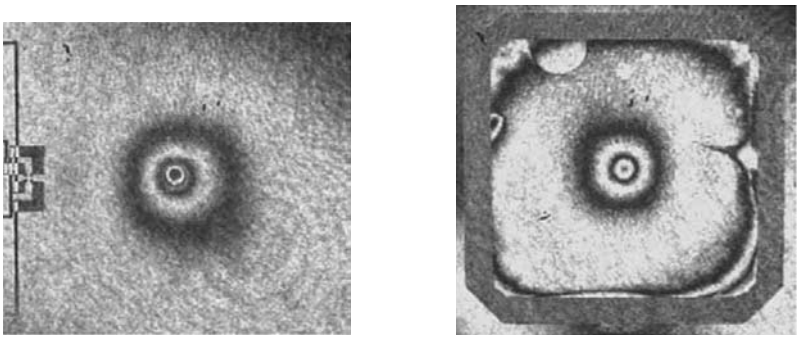


FIGURE 25.3. The front (a) and back side (b) images of a Si membrane before the electroless Ni plating.



(a) (b)

FIGURE 25.4. The membrane deformation produced by the Ni plating. Membrane deformations are measured from the front (a) and back side (b) of the membrane.

tion of 0.3165 μm. However, the resolution of this technique is not limited to 0.3165 μm. A phase-shift mechanism can be used to achieve a resolution of 1/10 of the fringe order.

The back surface of the membrane is optically reflective after the etching process. In practice, the membrane deformation can be measured from the backside of the membrane. The variation in the membrane thickness, which is measured before the plating process, as shown in Figure 25.3(b), is subtracted from the final deformation measured after the plating process. Thus, the final result is the deformation caused by the plating only. The measurement provides the whole-field deformation of the membrane including the area directly under the Ni film. However, sometimes, the etching process results in a rough surface with optical noises. In that case, the front side deformation can be used for the analysis. If the front side is used for deformation measurements, the thickness of the Ni film should be taken into consideration. If the data is taken from the backside of the membrane, the membrane thickness variation should be subtracted.

Figure 25.4 shows the typical interferometric fringe patterns obtained after the Ni plating process. The measurements were conducted at room temperature. In the experiment, membrane deformations were measured from the front and backsides of the membrane. The membrane size is 2600 × 2600 μm; membrane thickness is 23 μm; Ni film diameter is 200 μm; and Ni film thickness is 6 μm.

Since the plating temperature of the electroless Ni is at 85°C, in addition to the intrinsic stress, the measurements conducted at the room temperature will include the thermal stress caused by the Coefficient of Thermal Expansion (CTE) mismatch between the Ni and Si membrane. In order to determine the deformation induced by the intrinsic stress only, the measurements were also conducted at the plating temperature, 85°C. A hot plate with an enclosure and a glass window was used to heat the wafer and keep it at 85°C. Figure 25.5 shows the results of the measurements at the plating temperature. The images are membranes with the plated Ni films of 200, 500, 1000 and 1500 μm in diameters, respectively. The Ni thickness is 6 μm.

From the comparison of the membrane deformations obtained at the room temperature and those at the plating temperature, it was found that the deformation directions of the membranes were the same. When the measurements were conducted at room temperature, the thermal stress added into the intrinsic stress and the measurements showed that the magnitude of the deformation increased, but the deformation shape remained the same.

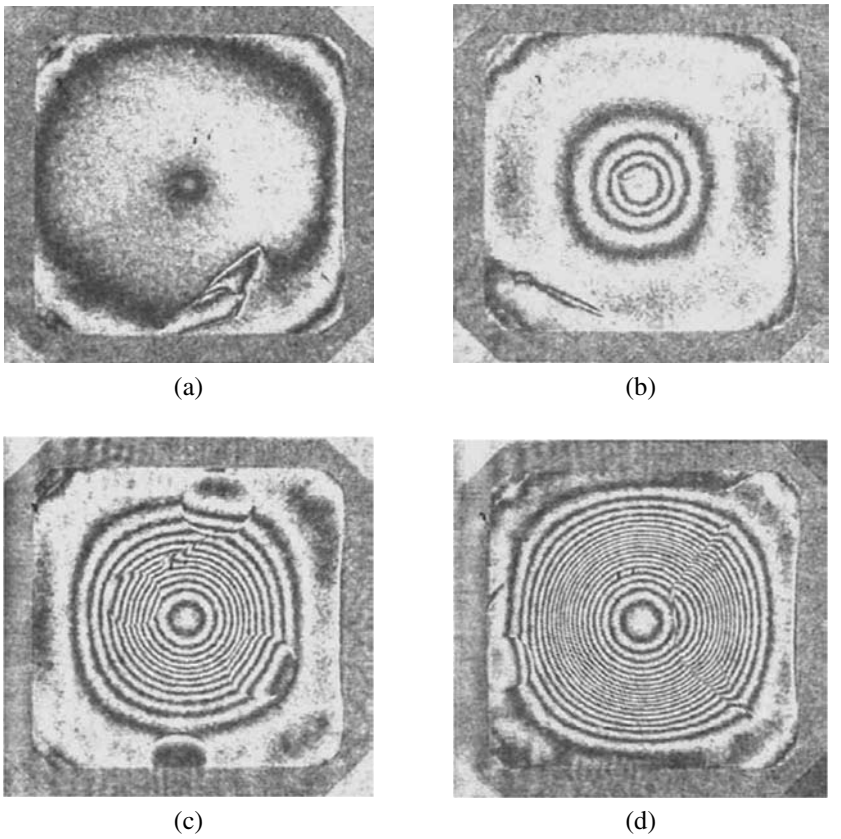


FIGURE 25.5. The membrane deformation produced by the Ni plating. The diameters of the Ni films are (a) 200 μm, (b) 500 μm, (c) 1000 μm, and (d) 1500 μm. Deformations are measured on the back surface of the membrane at 85°C, which is the plating temperature.

It means that the intrinsic stress in the Ni film has the same sign as the thermal stress in the Ni film caused by the negative ΔT (temperature decrease), which was a tensile stress.

25.1.3. Intrinsic Stress

In order to obtain the actual values of the intrinsic stresses, a finite element model (FEM) was used (Figure 25.6). With a parametric FEM model, we can model the geometry of the membrane sensing device exactly and can cover a range of wafer designs. In the process of the stress calculation, a pre-assumed intrinsic stress value was applied to the Ni film in the FEM model and the resulting membrane deformation was calculated. The membrane deformation from the FEM model was then compared with the experimentally measured membrane deformation. According to the comparison, the intrinsic stress value was adjusted in the FEM model until the deformation output from the model matched the experimental results. Figure 25.7 shows the final output of the FEM and the experimental results on the membrane deformations for the four Ni film sizes. With the right amount of intrinsic stress input in the model, the deformations output can be closely matched with the experimental results. Since all the boundary conditions are exact, the intrinsic stress in the FEM model, therefore, represents the intrinsic stress value in the Ni film.

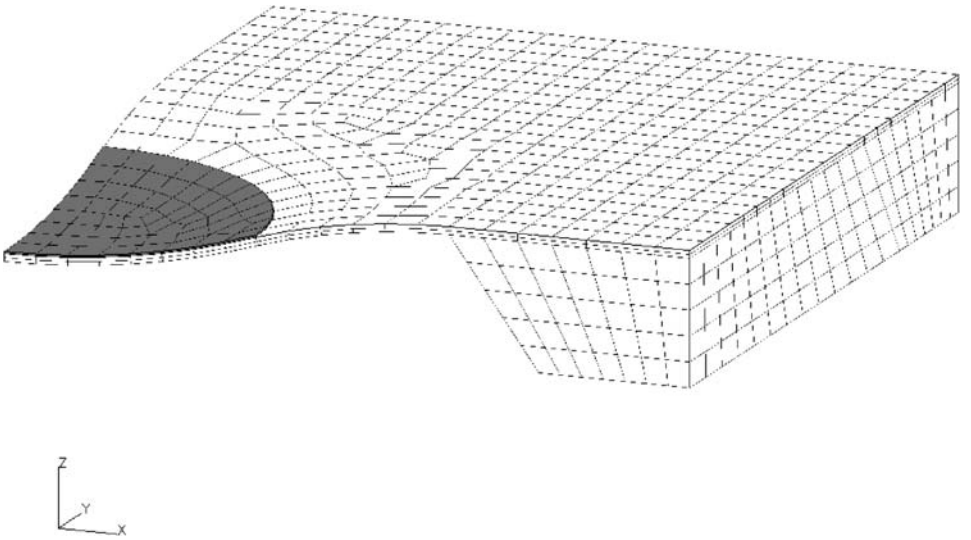


FIGURE 25.6. The FEM axisymmetric model simulated the membrane deformation with intrinsic stress input in the Ni film.

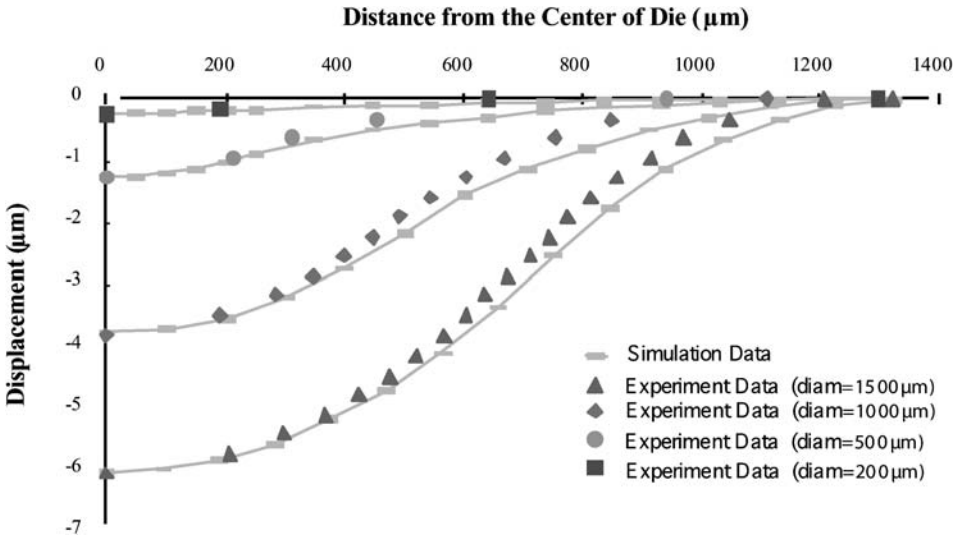


FIGURE 25.7. Experimental and FEM results on the membrane deformations for the four Ni film sizes at plating temperature.

Figure 25.8 gives the intrinsic stress values in the four Ni films with diameters of 200, 500, 1000 and 1500 μm. It is clear that the intrinsic stresses are dimensional sensitive. The intrinsic stresses increase as the Ni film dimension increases. In the real UBM, the Ni pad size is 120 μm, and the Ni intrinsic stress is around 50 MPa.

The resulting stress distributions in the Ni and the Si membrane can also be calculated from the FEM model. Figure 25.9 shows the normal and shear stress in the Si membrane at the edges of the Ni film. The resulting stress distributions are obtained based on the

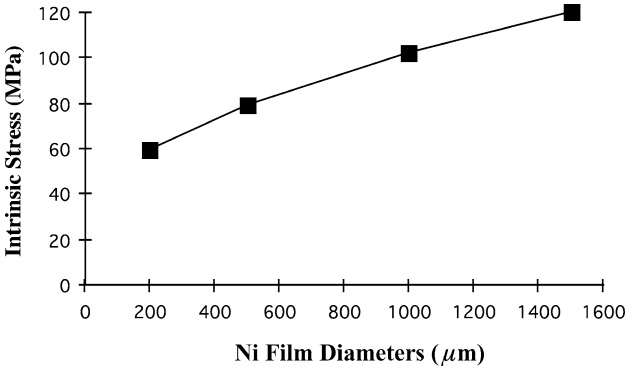


FIGURE 25.8. The intrinsic stress values in the four Ni films with different diameters. The intrinsic stress is a function of the Ni film size.

assumption that there is no Si membrane deformation before the metal was deposited. In practice, this assumption is not necessarily needed since the process induced Si membrane deformation can be obtained by subtracting the initial deformation from the final deformation after each process step.

25.1.4. Intrinsic Stress in Processed Wafer: Summary

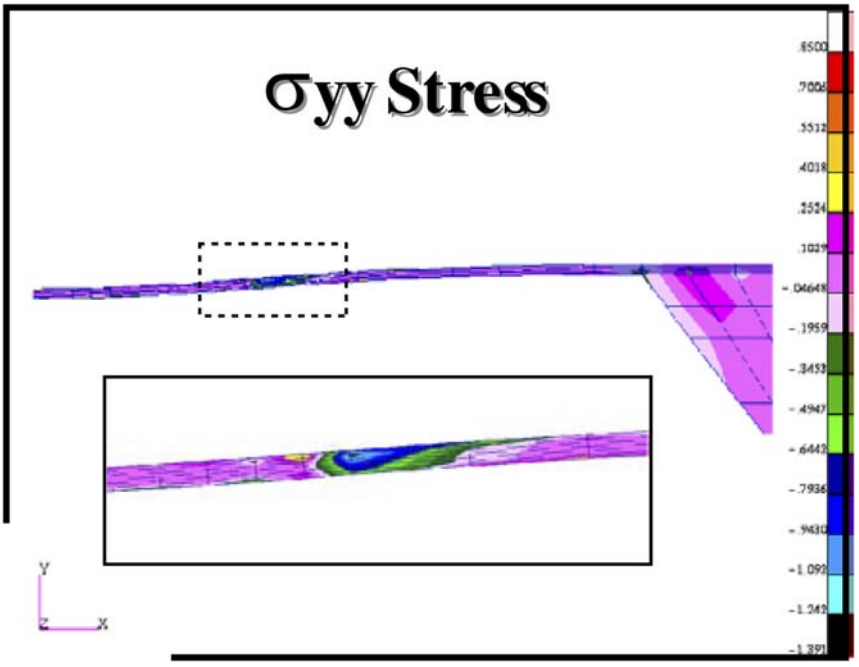
The proposed technique uses a silicon membrane to detect and display the intrinsic stress, which is the stress accumulated during the plating processes. Deformations in thin reflective Si membrane are measured accurately with laser interferometry and the stress calculations are completed by finite element method (FEM).

In the study of the electroless Ni plating process, it is found that the intrinsic stress value in the Ni film is a function of the film dimension, and the dimension dependency is very strong. As the film dimension increases, the intrinsic stress should approach the value measured by the FLEXUS machine which provides the intrinsic stress values in a thin film with a dimension close to infinity. This hypothesis is currently being verified.

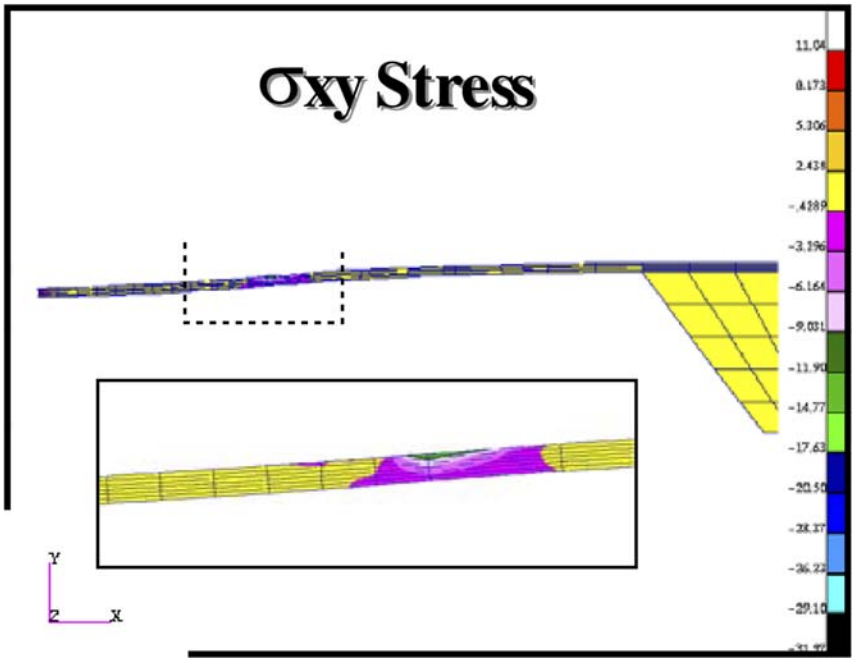
The intrinsic stress from the wafer process is a critical parameter affecting the reliability of the device, such as the bump strength, UBM (under bump metalization) strength, thin film adhesion, failure in the passivation layer and die cracking. The technique experimented should have impact on the following areas:

(a) Intrinsic stress measurements of different UBM.
(b) Process characterization, a simple technique to determine the local thin-film stress induced by metalization, deposition, plating, etching and heat treatment.
(c) In-situ monitoring of stress change in interconnect lines during electromigration and thermal migration in the accelerated testing.
(d) Local stress in active region after ion implantation (doping) and subsequent annealing.

The current technique used in the intrinsic stress measurement is the FLEXUS machine. Very often, it does not have enough sensitivity to measure the deformation if the thin film is patterned on a regular Si substrate. In practice, the Si substrates have to be thinned down to get enough sensitivity. In the FLEXUS machine, the wafer deformation is measured by the laser reflection, which can only obtain the average thin-film stress but not the



(a)



(b)

FIGURE 25.9. The normal (a) and shear stresses (b) in the silicon membrane at the interface with the Ni film by FEM analysis. High stresses are shown in the silicon membrane next to the edge of the Ni layer.

local stress variations. The proposed device with a thin membrane can provide extremely high sensitivity. By using a full field optical method, local deformations from small and various thin film geometry can be determined with very high resolution. With the aid of the FEM method, the intrinsic stress and the resulting stress field can be determined. This technique has its advantages to complement the FLEXUS machine for determinations of local stress fields caused by the intrinsic stress.

There is no practical limitation regarding the dimensions and shapes of the thin films studied. Currently, a great interest is in the area of determining the intrinsic stress in the UBM processes. In this study, the dimensional effect on intrinsic stress in the electroless Ni UBMs is analyzed. This Si membrane based testing sensor has great potentials in the wafer process characterization with high accuracy solution and a low cost. This new technology could also promise the capability to study the stress in much finer structures, such as interconnect line, stress change due to electromigration and thermal migration in interconnect lines and, more importantly, the stress at gate SiO_2/Si interface.

25.2. DIE STRESS RESULT FROM FLIP-CHIP ASSEMBLY

In flip chip plastic ball grid array (FC-PBGA) packages, silicon die is attached on a laminate substrate by solder joints. Underfill material is filled in the gap between the die and the substrate to protect the solder joints for better reliability. After the underfill process, the die and the substrate are rigidly bonded together and no interface delamination and separation should be present. A silicon die, with a coefficient of thermal expansion (CTE) of 2.6 ppm/°C, and a laminate substrate, with a CTE from 15 to 25 ppm/°C, are connected by underfill material. As a result of CTE mismatch, significant thermal stress occurs in the die and the substrate during thermal cycles. In component level reliability testing, this thermal stress is the major cause of many failure modes including die cracking. Lately in the development of FC-PBGA packages, reducing die stress and improving reliability has been a very serious issue.

25.2.1. Consistent Composite Plate Model

The flip chip package can be treated as a multilayer composite system. The consistent plate model treats the chip, the underfill and the carrier substrates, i.e., the chip/carrier module, as plies of an integrated laminated plate. The only assumptions made are those normal to thin and classical laminate plate theory [2].

The term consistent plate model comes from the assumption of consistent deformations in the chip and substrate. The strain in the system will therefore depend only on the strain on the reference plane somewhere in the middle of the plate, $\boldsymbol{\varepsilon}^0$, and the curvature of the laminated plate, $\boldsymbol{\kappa}$,

$$\begin{Bmatrix} \varepsilon_x \\ \varepsilon_y \\ \varepsilon_{xy} \end{Bmatrix} = \begin{Bmatrix} \varepsilon_x^0 \\ \varepsilon_y^0 \\ \varepsilon_{xy}^0 \end{Bmatrix} + z \begin{Bmatrix} \kappa_x \\ \kappa_y \\ \kappa_{xy} \end{Bmatrix}. \tag{25.1}$$

The constitutive relation for any ply of a laminated plate is

$$\boldsymbol{\sigma} = \mathbf{Q}(\boldsymbol{\varepsilon} - \boldsymbol{\Lambda}), \tag{25.2}$$

where the stress vector $\boldsymbol{\sigma}$ and induced strain vector $\boldsymbol{\Lambda}$ are

$$\boldsymbol{\sigma} = \begin{Bmatrix} \sigma_x \\ \sigma_y \\ \sigma_{xy} \end{Bmatrix}, \quad \boldsymbol{\Lambda} = \begin{Bmatrix} \Lambda_x \\ \Lambda_y \\ \Lambda_{xy} \end{Bmatrix} = \begin{Bmatrix} \alpha_x \Delta T \\ \alpha_y \Delta T \\ \alpha_{xy} \Delta T \end{Bmatrix}. \tag{25.3}$$

Here α is the coefficient of thermal expansion and ΔT is the temperature change.

The matrix $\mathbf{Q}$ is the transformed reduced stiffness of the lamina and is given by Ref. [4] as following

$$\mathbf{Q} = \begin{Bmatrix} \overline{Q}_{11} & \overline{Q}_{12} & \overline{Q}_{16} \\ \overline{Q}_{12} & \overline{Q}_{22} & \overline{Q}_{16} \\ \overline{Q}_{16} & \overline{Q}_{16} & \overline{Q}_{66} \end{Bmatrix}. \tag{25.4}$$

The load-deformation relationship for the consistent plate is given by

$$\begin{bmatrix} \mathbf{N} \\ \mathbf{M} \end{bmatrix} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B} & \mathbf{D} \end{bmatrix} \begin{Bmatrix} \boldsymbol{\varepsilon}^0 \\ \boldsymbol{\kappa} \end{Bmatrix} - \begin{bmatrix} \mathbf{N}^\Lambda \\ \mathbf{M}^\Lambda \end{bmatrix}, \tag{25.5}$$

where the conventional mechanical stress resultants, the mechanical forces and moments, are

$$\mathbf{N} = \int_t \boldsymbol{\sigma} dz, \tag{25.6}$$

$$\mathbf{M} = \int_t \boldsymbol{\sigma} z dz. \tag{25.7}$$

The matrices $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{D}$ are the usual extensional stiffness, bending-stretching coupling stiffness and bending stiffness of the plate [4]. The integrations in the above equations are carried out through the composite plate thickness. The equivalent thermal forces and moments are

$$\mathbf{N}^\Lambda = \int_t \mathbf{Q}\boldsymbol{\Lambda} dz, \tag{25.8}$$

$$\mathbf{M}^\Lambda = \int_t \mathbf{Q}\boldsymbol{\Lambda} z dz. \tag{25.9}$$

The total potential energy stored in the plate is given by

$$U = \frac{1}{2} \iint_\Omega \begin{Bmatrix} \boldsymbol{\varepsilon}^0 \\ \boldsymbol{\kappa} \end{Bmatrix}^T \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B} & \mathbf{D} \end{bmatrix} \begin{Bmatrix} \boldsymbol{\varepsilon}^0 \\ \boldsymbol{\kappa} \end{Bmatrix} d\Omega - \iint_\Omega \begin{bmatrix} \mathbf{N}^\Lambda \\ \mathbf{M}^\Lambda \end{bmatrix}^T \begin{Bmatrix} \boldsymbol{\varepsilon}^0 \\ \boldsymbol{\kappa} \end{Bmatrix} d\Omega. \tag{25.10}$$

This strain energy equation together with a Ritz approximate solution method can be used to solve for the approximate strains and curvatures in the multilayer system.

25.2.2. Free Thermal Deformation

When there is no external mechanical load, Equation (25.5) can be reduced to

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B} & \mathbf{D} \end{bmatrix} \begin{Bmatrix} \boldsymbol{\varepsilon}^0 \\ \boldsymbol{\kappa} \end{Bmatrix} = \begin{bmatrix} \mathbf{N}^\Lambda \\ \mathbf{M}^\Lambda \end{bmatrix}. \tag{25.11}$$

Equation (25.11) can be solved directly for $\boldsymbol{\varepsilon}^0$ and $\boldsymbol{\kappa}$.

As shown in Figure 25.10, three special cases are of particular interests for electronics packaging analysis, namely, the beam bending, the cylindrical bending and the axisymmetrical bending. When material anisotropy for each layer is small, Equation (25.4) can be written as

$$\mathbf{Q} = \begin{Bmatrix} \dfrac{E_i}{1-\nu_i^2} & \dfrac{\nu_i E_i}{1-\nu_i^2} & 0 \\ \dfrac{\nu_i E_i}{1-\nu_i^2} & \dfrac{E_i}{1-\nu_i^2} & 0 \\ 0 & 0 & \dfrac{E_i}{2(1+\nu_i)} \end{Bmatrix}. \tag{25.12}$$

Here E_i and ν_i are the Young's modulus and Poisson's ratio for the ith layer lamina. Note Equation (25.12) is given for the axisymmetrical case. For the cylindrical and beam bending cases, we can simply replace the term $E_i/(1-\nu_i)$ with E_i.

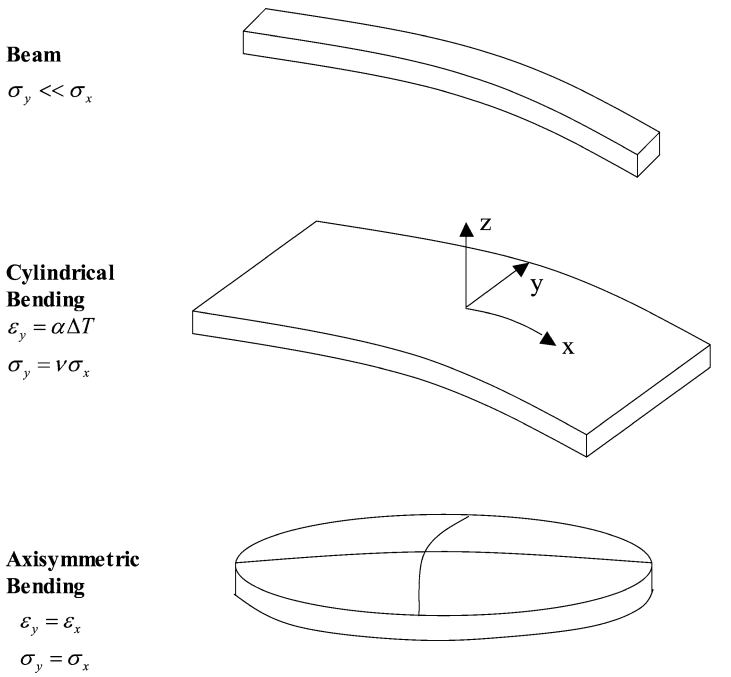


FIGURE 25.10. Three possible deformation modes for multilayered structures.

The deformation of the composite plate can be further simplified such that for beam bending,

$$\sigma_y \ll \sigma_x;$$

for cylindrical bending,

$$\begin{Bmatrix} \kappa_x \\ \kappa_y \\ \kappa_{xy} \end{Bmatrix} = \begin{Bmatrix} \kappa_x \\ 0 \\ 0 \end{Bmatrix}, \begin{Bmatrix} \varepsilon_x^0 \\ \varepsilon_y^0 \\ \varepsilon_{xy}^0 \end{Bmatrix} = \begin{Bmatrix} \varepsilon_x^0 \\ \varepsilon_y^0 \\ 0 \end{Bmatrix},$$

for axisymmetrical bending,

$$\begin{Bmatrix} \kappa_x \\ \kappa_y \\ \kappa_{xy} \end{Bmatrix} = \begin{Bmatrix} \kappa_x \\ \kappa_x \\ 0 \end{Bmatrix}, \begin{Bmatrix} \varepsilon_x^0 \\ \varepsilon_y^0 \\ \varepsilon_{xy}^0 \end{Bmatrix} = \begin{Bmatrix} \varepsilon_x^0 \\ \varepsilon_x^0 \\ 0 \end{Bmatrix}.$$

The force and the moment equations for the above three cases become

$$\begin{bmatrix} A_x & B_x \\ B_x & D_x \end{bmatrix} \begin{Bmatrix} \varepsilon_x^0 \\ \kappa_x \end{Bmatrix} = \begin{bmatrix} N_x^\Lambda \\ M_x^\Lambda \end{bmatrix}. \tag{25.13}$$

Here coefficients A_x, B_x, D_x, N_x^Λ and M_x^Λ are given by the following equations [2,3]

$$\begin{cases} A_x = \displaystyle\int_t \frac{E_i}{1-\nu_i} dz \\ B_x = \displaystyle\int_t \frac{E_i}{1-\nu_i} z dz \\ D_x = \displaystyle\int_t \frac{E_i}{1-\nu_i} z^2 dz \\ N_x^\Lambda = \displaystyle\int_t \frac{E_i \alpha_i \Delta T}{1-\nu_i} z dz \\ M_x^\Lambda = \displaystyle\int_t \frac{E_i \alpha_i \Delta T}{1-\nu_i} z dz. \end{cases} \tag{25.14}$$

Again, Equation (25.14) is given for the axisymmetrical case. For cylindrical and beam bending cases, we simply replace the term $E_i/(1-\nu_i)$ with E_i in the above equation.

25.2.3. Bimaterial Plate (BMP) Case

As an example, we consider a flip chip package shown in Figure 25.11. Using the consistent composite plate model, the flip-chip package can be treated as a system of two plates bonded together if only the die bending curvature and the stress on top of the die is interested. The underfill and solder joins can be treated as part of the substrate in this scenario. The Young's modulus, Poisson's ratio, CTE and thickness of upper plate (plate 1 or die) are E_1, ν_1, α_1, h_1, respectively (Figure 25.12). The corresponding quantities for

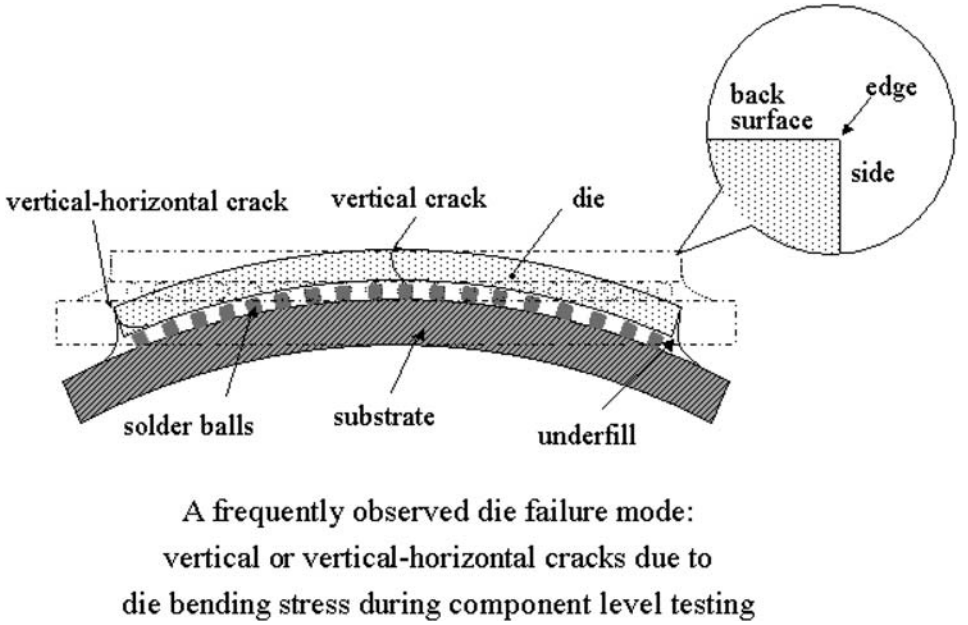


FIGURE 25.11. A schematic diagram of a flip-chip plastic ball grid array package and die failure modes due to die bending. The dashed lines represent the package at underfill curing temperature, which is stress free. The solid lines represent the package at room temperature, at which the package bends due to the CTE mismatch between the die and the substrate.

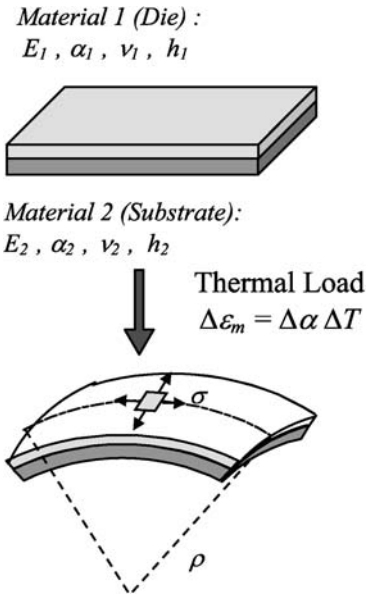


FIGURE 25.12. Bimaterial plate model for a flip chip module.

the lower plate (plate 2 or substrate) are E_2, ν_2, α_2, h_2, respectively. In the flip chip package, the two plates were bonded together at a certain temperature T_0. The bimaterial plate (BMP) system will bend into a surface which is part of a sphere when temperature is reduced to T (Figure 25.12). Solving Equation (25.13) for the BMP system, we have

$$\begin{cases} \varepsilon_x^0 = \dfrac{D_x N_x^\Lambda - B_x M_x^\Lambda}{A_x D_x - B_x^2} \\ \kappa_x = \dfrac{A_x M_x^\Lambda - B_x N_x^\Lambda}{A_x D_x - B_x^2}. \end{cases} \tag{25.15}$$

Please note that ε_x^0 is dependent on the location of x–y plane while κ_x is consistent through the plate thickness. Substituting Equation (25.14) into Equation (25.15) and after some mathematical manipulations we have

$$\begin{cases} \varepsilon_x^0 = \dfrac{hm(4+3h+h^3m)(\alpha_2+\alpha_1)\Delta T}{(1+hm(4+6h+4h^2+h^3m))} \\ \kappa_x = \dfrac{6\varepsilon_m hm(1+h)}{h_1(1+hm(4+6h+4h^2+h^3m))}, \end{cases} \tag{25.16}$$

where $h = h_2/h_1$ is the thickness ratio of the lower plate to the upper plate, $m = M_2/M_1$ is the ratio of the biaxial modulus of the lower plate $[M_2 = E_2/(1-\nu_2)]$ respect to the upper plate $[M_1 = E_1/(1-\nu_1)]$, and $\varepsilon_m = (T - T_0)(\alpha_1 - \alpha_2)$ is the thermal mismatch strain between the two plates.

In practice, it is convenient to express the curvature in a dimensionless quantity or, so called, characteristic curvature:

$$\overline{\kappa} = \frac{h_1}{6\varepsilon_m}\kappa_x = \frac{hm(1+h)}{1+hm(4+6h+4h^2+h^3m)}. \tag{25.17}$$

The characteristic curvature is a function of only the thickness ratio h and the biaxial modulus ratio m.

From Equation (25.2) and Equation (25.16), the stress on the top surface of plate 1, the die, can be expressed as

$$\sigma_{top} = \frac{\varepsilon_m M_1 hm(2+3h-h^3m)}{1+hm(4+6h+4h^2+h^3m)}. \tag{25.18}$$

The stress is uniform at any point on the top of the die which is a direct result of neglecting the edge effect. The dimensionless stress or characteristic stress is defined as

$$\overline{\sigma} = \frac{\sigma_{top}}{\varepsilon_m M_1} = \frac{hm(2+3h-h^3m)}{1+hm(4+6h+4h^2+h^3m)}. \tag{25.19}$$

It is instructive to express the stress on top of the die in terms of the curvature or characteristic curvature.

$$\sigma = \kappa_x \frac{h_1 M_1(2+3h-h^3m)}{6(1+h)}, \tag{25.20}$$

$$\overline{\sigma} = \overline{\kappa}_x \frac{2 + 3h - h^3 m}{1 + h}. \tag{25.21}$$

According to Equations (25.20) and (25.21), the stress contains the curvature κ_x. Therefore, the curvature κ_x can provide direct information of the stress.

25.2.4. Validation of the Bimaterial Model

The mechanical behavior of the FC-PBGA package is similar to a mechanical system with two bonded plates. A bimaterial laminate model can be used to estimate the stress and bending curvature in each plate. The bimaterial plate model (BMP) has been widely used in the stress determinations of the wafer level process with thin films deposited on silicon wafers. With the thin film approximation, which is known as Stoney's equation [13], it provides an effective method to determine the stresses in thin films deposited on wafers [14]. However, this model can not be directly applied to the FC-PBGA packages because of the assumptions in the model that the film is much thinner than the wafer. In practice, the die stress in the FC-PBGA packages is usually determined from finite element method (FEM) models which require simulation tools and are comparatively time consuming when packages with different geometry are analyzed.

A typical FC-PBGA package is shown in Figure 25.11. In the assembly process, the die is attached to the substrate through solder joints. The die attach process does not produce high die stress because of the creep and relaxation in the solder joints. Most of the die stress is produced by the underfill process. When the underfill material is cured at high temperature (usually 150 to 170°C), the die and the substrate is connected. As the package cools down to room temperature, the thermal stress accumulates because of the CTE mismatch between the die and the substrate. A bending deformation is developed by cooling the package from the underfill curing temperature to room temperature. Further cooling to lower temperature in the reliability test will induce more package bending and higher die stress. During the package qualification, thermal cycles from −55 to 125°C are used for the package level test.

In a real package, the underfill material is adhered to the die and substrate, the degree of the creep and relaxation is not significant in the underfill material when it is fully cured. The stress free temperature of the system is near the underfill curing temperature. When the package is thermally cycled, the thermal stress in the die can cause several failure modes in die cracking at the low temperature. Figure 25.11 also shows the failure due to die cracking that have been identified during the reliability tests: (a) vertical crack, the crack initiated from the die back side and propagated vertically down into the die, (b) vertical-horizontal crack, the crack initiated from the back side propagated vertically first then turned to horizontal. The BMP model developed can be applied in the study of these two types of die cracks.

The tensile stress at the backside of the die is obviously the cause of the vertical cracks initiated from the backside. It is clear that the back surface of the die is in tension. The cracks are initiated from the back surface or the edges of the back surface and across the die from edge to edge. The depth of the crack depends on the stress level in the package. Simulation shows that the tensile stress at the backside of the die is distributed almost over entire back surface.

The BMP model is validated by experiments and finite element method (FEM). The curvature was directly validated by curvature measurement using an optical method. The stress calculated from the BMP model was compared with FEM results.

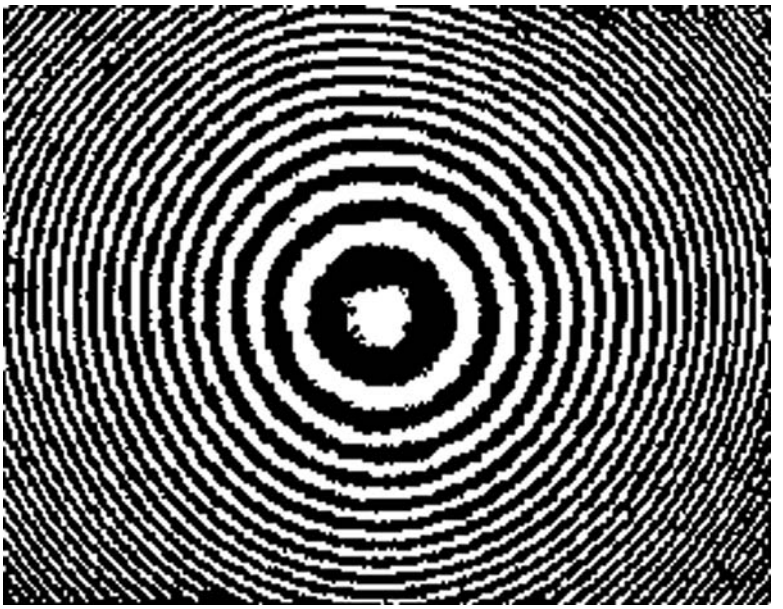


FIGURE 25.13. A typical interference fringe pattern obtained from the Si chip surface with the Twyman-Green interferometer.

In order to validate the curvature model for different thickness ratio and different materials, two sets of sample were prepared. The dice chosen in this study are single Si with (100) orientation. The dimensions are 6.78 mm × 5.50 mm × 0.54 mm. The dice are both-side polished. Substrates are printed circuit board (PCB) and Cu substrates. The dimension of the dice is kept unchanged for different samples. The thickness of the substrates varies from 0.25 times to 2 times of the die thickness for both PCB and Cu substrates. The length and the width of the substrates are kept the same as those of the dies. The die and substrate are bonded together by a thin layer of epoxy. After about 1 hour curing at 83°C, the epoxy layer firmly bonds the die and substrate. As the assemblies cool down to room temperature, bending and stress are generated due to the CTE mismatch. The thickness of the epoxy layer is less than 10 μm, which is negligible in the die bending and die stress calculation.

The curvature measurement was carried out using a Twyman-Green interferometer [15] with an *in situ* heating chamber. A typical fringe pattern of the die bending is shown in Figure 25.13. It is the interference pattern of a Si die (6.78 mm × 5.50 mm × 0.54 mm) on a 0.77 mm thick Cu substrate at 23.5°C. The package is flat at 78.5°C. The fringe pattern is a displacement contour of the sample surface. Each circular fringe corresponds to 316.4 nm out-of-plane deformation. Figure 25.13 clearly shows that the fringe pattern is consisted of concentric circles. The bimaterial plate theory treats the two plates as an axisymmetric mechanical system. The experiment proves that the deformation of the sample is indeed an axisymmetric spherical surface. It is noteworthy to point out that the zero bending temperature (78.5°C) is lower than the epoxy curing temperature 83°C in this case. The reason is that the thin epoxy layer had experienced a small stress relaxation before the test. It is crucial to check the zero bending temperature for each sample in order to get the accurate data.

The FEM calculation is carried out using the commercial software ANSYS 5.5. Two-dimensional (2D) 8-node axisymmetric solid elements were used to calculate the die bend-

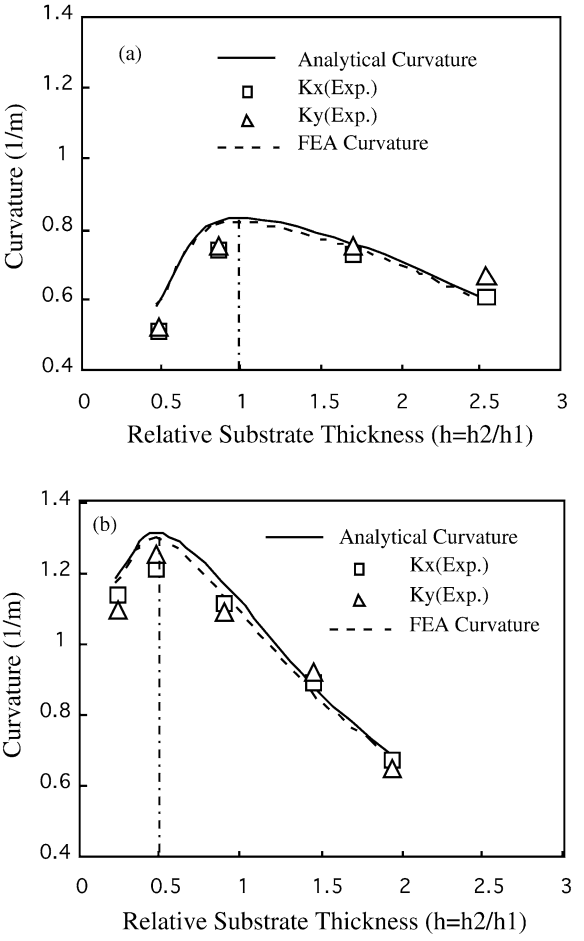


FIGURE 25.14. Bending curvatures (a) Si die on PCB substrate and (b) Si die on Cu substrate.

ing and the stress on top of the die. The FEM model treats the bimaterial plates as circular plates in order to be consistent with the analytical BMP model. Linear elastic deformation is assumed in the FEM model.

The bending curvatures from experiment measurement, BMP model and FEM are plotted in Figure 25.14. Figure 25.14(a) is the result of die on PCB substrate, and Figure 25.14(b) is the result of die on Cu substrate. The Young's modulus, Poison's ratio and CTE of Si used in the calculation are 131 GPa, 0.28 and 2.5 ppm/°C, respectively. The corresponding values are 21 GPa, 0.33, and 20 ppm/°C respectively, for PCB substrate, and 117 GPa, 0.35, and 17 ppm/°C, respectively, for Cu substrate. The experimental results of the curvature are in two orthogonal directions. The x-direction is along the length of the samples and the y-direction is along the width of the samples. The horizontal axis of the figure is the thickness ratio $h = h_2/h_1$, with the die thickness of $h_1 = 0.54$ mm fixed. For each substrate thickness, at least two samples were tested; the curvatures shown in the figure are the averaged values. Both Figure 25.14(a) and Figure 25.14(b) clearly show that the calculated curvatures from both BMP and FEM agree well with the experimentally mea-

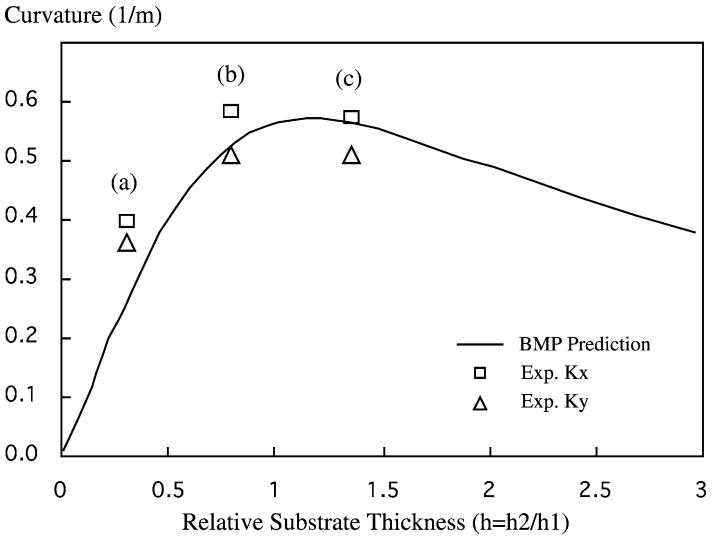


FIGURE 25.15. Curvature measured by Twyman-Green interferometer and calculated by BMP model for flip-chip packages.

sured curvatures. The curvatures of the two orthogonal directions are the same from the experimental measurement.

A very interesting feature shown in Figure 25.14(a) and Figure 25.14(b) is that there is a maximum bending curvature in both substrate systems. For PCB substrates, the maximum bending curvature occurs when the substrate to die thickness ratio equals to one, and the corresponding value for the Cu substrate occurs when the substrate to die thickness ratio equals to 0.5. This maximum bending curvature should be avoided in package design in order to reduce the package bending and die stress.

A set of real FC-PBGA packages was also tested. The die dimension is 7.0 mm $\times$ 6.5 mm $\times$ 0.61 mm. The underfill has a thickness of 0.13 mm. Figure 25.15 shows the curvature of this set of packages measured by Twyman-Green interferometer. The two orthogonal curvatures are plotted in the same figure. The curvature predicted by the BMP theory is plotted for comparison. An approximation made in the calculation of the curvature is that the underfill layer is treated as part of the substrate since the underfill layer has similar material properties as the substrate. It is clear that the predicted curvature by the BMP model is very close to the average curvature from experimental measurements in real packages.

The stresses on the top of the dies were calculated by BMP model and FEM for different samples. The stresses are shown in Figure 25.16. Although the BMP is based on plate approximation and the FEM uses 2D axisymmetric solid elements, the stress results are very close. The FEM calculation does show some edge effect at the die outer peripheral, this effect only affects the local area of about the dimension of order of the die thickness. The stress maintains constant in the rest part of the die. The FEM calculated stress shown in Figure 25.16 is the stress near the die center. The good agreement enables us to use BMP model to have a quick estimate of the stress on top of the die.

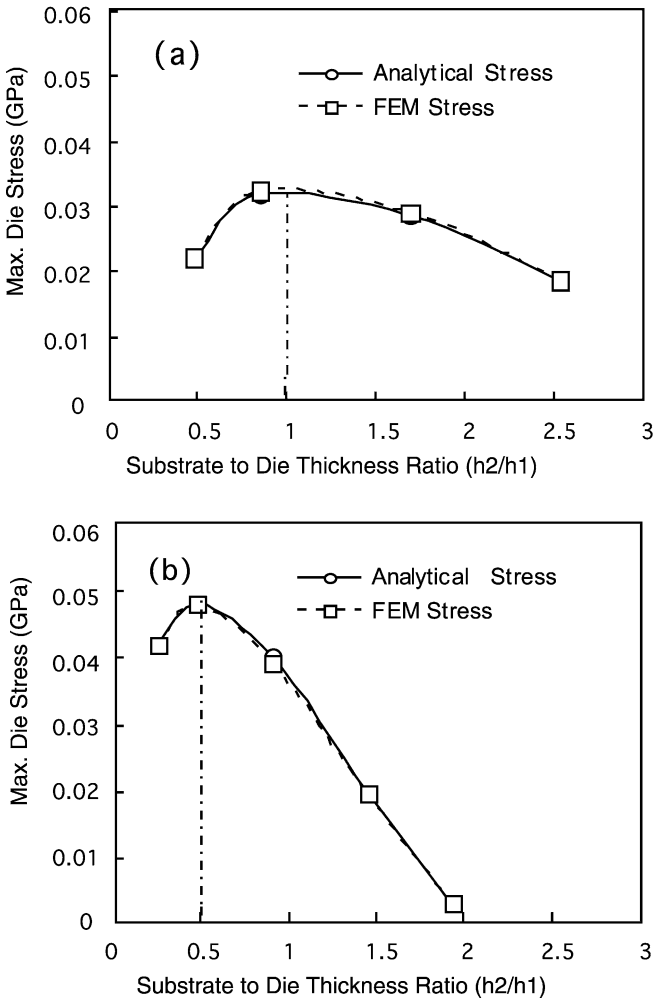


FIGURE 25.16. The stress on top of the die calculated from BMP model and FEM for (a) Si die on PCB substrate and (b) Si die on Cu substrate.

25.2.5. Flip-Chip Package Design

Since the BMP model gives closed-form solutions for bending curvature and die stress, it is very convenient to use the model to do parametric study in package design. The data can be tabulated or plotted as a handy reference. Figure 25.17 plots the characteristic curvature and characteristic die stress as a function of thickness ratio for different biaxial modulus ratios of the substrate to die. The reason to choose the characteristic curvature defined in Equation (25.17) is that it is dimensionless, thermal load independent and absolute thickness independent. This ensures that Figure 25.17 is universally applicable for arbitrary material combination, arbitrary geometry combination, and arbitrary thermal load. It is clearly shown that for every different material combination there is a maximum stress. As the ratio of the biaxial modulus of the substrate to die (m) increases, the maximum die stress (σ) increases and the position of the maximum die stress in terms of thickness

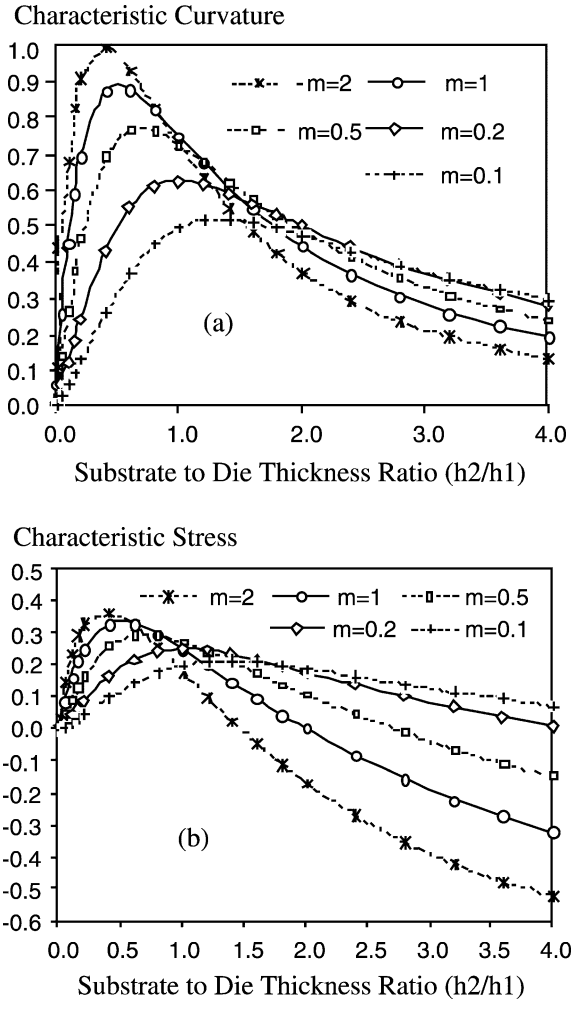


FIGURE 25.17. Dimensionless or characteristic curvature and stress in a bimaterial plate system as a function of thickness ratio and biaxial modulus ratio: (a) characteristic curvature; (b) characteristic stress.

ratio (h) tends to be smaller simultaneously. The position of maximum curvature is dependent on the material properties. Most of the FC-PBGA packages have the maximum curvature at $h \approx 1$, which is a coincident of the fact that most of packages are Si on PCB with $m \approx 0.2$. For Si on Cu ($m \approx 1$), the maximum curvature is located at $h \approx 0.5$. The characteristic curvature is approaching to zero as relative die thickness approaching zero or infinite. It is a reflection that the either the die or the substrate is too weak to let the system bend at these two extremes.

Figure 25.17(b) plots the stress on the top of the die predicted by the BMP theory. Although the die stress is related to the die bending, it is not simply related. The behaviors of the two quantities are very similar in certain range of thickness ratio h. This can be clearly seen in Figure 25.17(a) and Figure 25.17(b). Actually, if $h \to 0$, from Equation (25.21), $\overline{\sigma} \to 2\overline{\kappa}$. The behavior of stress is totally different from the curvature if h is large. From Equation (25.17), $\lim_{h\to\infty} \overline{\kappa} = 0$. On the other hand, Equation (25.19) leads to $\lim_{h\to\infty} \overline{\sigma} = -1$.

The physical meaning is that the system does not bend too much as the substrate is relatively thick or in other words the die is relatively thin. In this case, the curvature approaches to zero. The stress comes from two competing factors. One is the bending, and the other is the thermal mismatch strain. If the substrate is much thicker than the die, the bending of the bimaterial plate system is not significant and the die stress mainly comes from the mismatches in thermal deformations. Therefore, the die will be under a compressive stress, which is not critical to die cracking. If we are only interested in the lower thickness ratio range, a simple relation of the curvature and the die stress can be established. From Equation (25.21), one can have a relation $\overline{\sigma} = C\overline{\kappa}$ where $C = \overline{\kappa}/\overline{\sigma} = (1+h)/(2+3h-mh^3)$. Generally, C is not a constant. For Si on PCB with $h = 1$, $C(m = 0.2, h = 1) = 0.41$. If we use this $C = 0.41$ for thickness ratio from 0 to 1.5, the correlation between the curvature and stress is very good.

Although the BMP model neglects the edge effect, it does not introduce large errors in the estimation the curvature and the die stress. Three dimensional FEM calculation results show that the package size effect and aspect ratio effect only make 5% difference [16]. Furthermore, the BMP model provides us more insight in the packaging design when the die size is becoming very large. From Equations (25.16) and (25.18), it can be clearly seen that the curvature and die stress are independent of the package size. The curvature is only the function of modulus ratio (m), thickness ratio (h), thermal load (ε_m), and die thickness (h_1), and the die stress is only the function of modulus ratio (m), thickness ratio (h), thermal load (ε_m), and die biaxial modulus (M_1). This indicates that the increase of die size will not cause intrinsic limiting factor coming from die stress as long as material combination and thickness ratio is properly chosen. In real packaging design, one does experience lower yield when package size increases. The reason is that the number of defects per die increases for larger package due to the larger area of die (assuming defect density is a constant).

Another important observation is that the die stress is independent of absolute die thickness when the thickness ratio is fixed. Equation (25.18) does not contain the factor of absolute die thickness. In real flip-chip packaging design, it was found that thinning and polishing the die can significantly enhance the yield. This can be understood by looking at the Figure 25.17. Thinning the die and keeping the substrate thickness unchanged is equivalent to increasing the thickness ratio (h). This actually offsets the maximum stress location to get a good yield if the die is originally thicker than the substrate. The other advantage is that polishing the die can reduce the defect density, which means higher die strength, and as a consequence, the yield is improved significantly.

25.2.6. Die Stress in Flip Chip Assembly: Summary

An analytical model and governing equations are established for describing the intrinsic behavior of flip-chip packages. The model is validated by experiment and FEM calculation. The curvature and the bending stress are independent of the die size. The bending stress is independent of absolute die thickness if substrate to die thickness ratio (h) is kept the same. Both the curvature and the die stress have maximum values as the thickness changes. For FC-PBGA packages, which correspond to biaxial modulus ratio ($m = 0.2$), the die stress is maximum when the thickness ratio (h) equals to one. The die curvature and the bending stress are approximately proportional in certain range of thickness ratio.

25.3. THERMAL STRESS DUE TO TEMPERATURE CYCLING

Due to thermal expansion mismatch between the microelectronics package and the printed circuit board on which the package is mounted, thermal stress is introduced in the cooling step of the industry standard solder assembly process. Dynamic stress is further produced in the completed package-board assembly, especially in the solder joints, when the assembly is subjected to power cycling, thermal cycling, or thermal shock. The dynamic stress/strain has a great impact on the long-term reliability of the solder joint.

25.3.1. Finite Element Analysis

Non-linear finite element simulation methodologies have been adopted to investigate the time-dependent thermal stress/strain in the solder joints. As an example, the finite element modeling of a FC-PBGA assembly under accelerated temperature cycling conditions is given here. ANSYS 6.0 is used for all aspects of the study: pre-processing, solution, and post-processing.

Due to symmetry, only one-quarter of the module-board assembly is modeled. Figure 25.18 shows one quarter of the module-board assembly for the FC-PBGA. The solder material is modeled as a viscoplastic solid. The PBGA substrate along with the printed circuit board are modeled as composite materials with their copper power and ground planes, and their orthotropic, temperature dependent dielectric materials explicitly modeled. The physical and mechanical properties of the silicon chip, underfill material and board are summarized in Table 25.1.

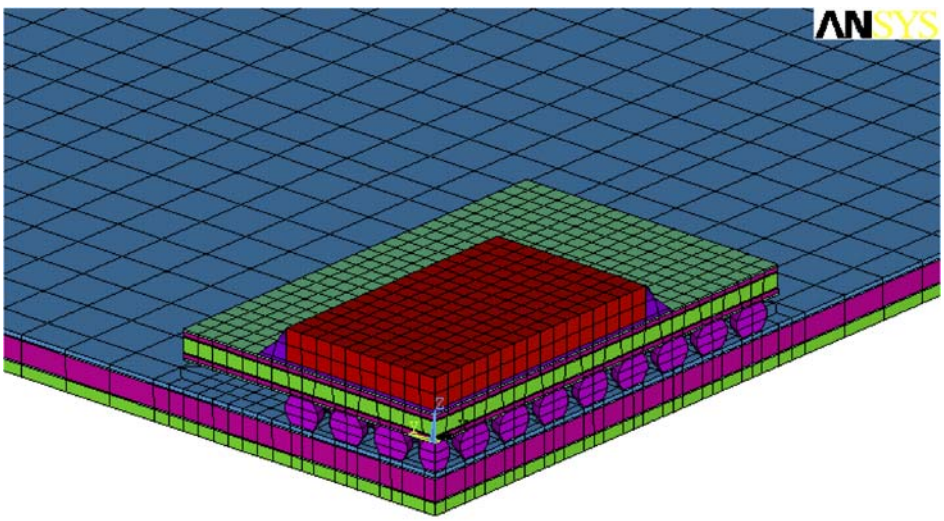


FIGURE 25.18. Finite element model of a FC-PBGA assembled on a board.

TABLE 25.1.
Material properties used in modeling of an FC-PBGA assembly.

Material	Modulus @ 23°C (GPa)	Poisson ratio @ 23°C	CTE (10^{-6}/°C) @ 23°C	T_g (°C)
Si Chip	186	0.28	3.2	NA
Underfill	3.4	0.33	24	145
PBGA substrate (Power Plane)	45.6	0.34	17.6	NA
PBGA substrate (Dielectric)	16.4 (x, y) 2.6 (z)	0.48 (x, y) 0.16 (z)	14.8 (x, y) 67 (z)	170
SnPb solder	Viscoplastic, see Ref. [21]	0.29	22	NA
Printed circuit board (Power Plane)	45.6	0.34	17.6	NA
Printed circuit board (Dielectric)	88.6 (x, y) 2.7 (z)	0.48 (x, y) 0.16 (z)	22.4 (x, y) 67 (z)	170

25.3.2. Constitutive Equation for Solder

Accurate constitutive modeling of the solder plays an important role in the simulation of solder joint reliability. ANSYS does have viscoplastic elements as a standard option, but they use Anand's constitutive model [17,18]. The use of these elements is convenient since the user does not have to modify the source code. Anand's model was developed for hot working metals, which unifying plasticity and creep via a set of flow and evolutionary equations:

Flow Equation

$$\frac{d\varepsilon_p}{dt} = A[\sinh(\xi\sigma/s)]^{1/m} \exp\left(-\frac{Q}{kT}\right). \tag{25.22}$$

Evolution Equations

$$\frac{ds}{dt} = \left[h_0(|B|)^a \frac{B}{|B|}\right]\frac{d\varepsilon_p}{dt}, \tag{25.23}$$

$$B = 1 - \frac{s}{s^*}, \tag{25.24}$$

$$s^* = \hat{s}\left[\frac{\dfrac{d\varepsilon_p}{dt}}{A}\exp\left(-\frac{Q}{kT}\right)\right]^n. \tag{25.25}$$

Darveaux et al. [19] were the first to modify the constants in Anand's constitutive relation to account for both time-dependent and time-independent phenomenon. Parameters for near-eutectic 62Sn/36Pb/2Ag are given in Table 25.2 (see Refs. [19,21]).

TABLE 25.2.
Anand's constants for 62Sn36Pb2Ag solder in ANSYS.

ANSYS	Parameter	Value	Definition
C1	S_o (psi)	1800	Initial value of deformation resistance
C2	Q/k (1/K)	9400	Activation energy/ Boltzmann's constant
C3	A (1/sec)	4.0×10^6	Pre-exponential factor
C4	ξ	1.5	Multiplier of stress
C5	m	0.303	Stain rate sensitivity of stress
C6	h_o (psi)	2.0×10^5	Hardening constant
C7	$s\hat{}$ (psi)	2.0×10^3	coefficient for deformation resistance saturation value
C8	n	0.07	Strain rate sensitivity of saturation (deformation resistance) value
C9	a	1.3	Strain rate sensitivity of hardening

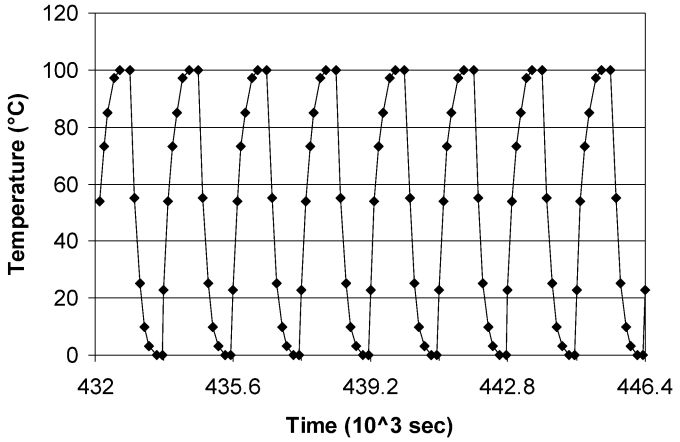


FIGURE 25.19. Cyclic temperature loading used in finite element modeling.

25.3.3. Time-Dependent Thermal Stresses of Solder Joint

The cyclic temperature loading imposed on the FC-PBGA/PCB assembly is closely matched to the temperature profile measured in accelerated temperature cycling test and is shown in Figure 25.19. It can be seen that for each cycle (30 minutes) the temperature is between 0 and +100°C, with 10 minutes ramp, 5 minutes hold at the two temperature extremes. In addition, pre-cycling, temperature-time history which corresponds to 5 days at 23°C is also included in the simulation.

Residual stresses in the solder joint after the FC-PBGA module was assembled to the board, cooled down to room temperature but before the temperature cycling test are shown in Figure 25.20. As expected the residual stresses relaxed very rapidly in the first hour after the assembly process. The stresses shown in Figure 25.20 are for the solder ball which is

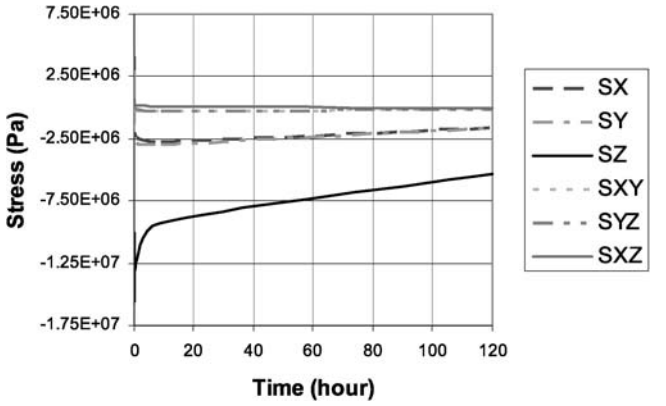


FIGURE 25.20. Residual stresses in the solder joint after assembly process.

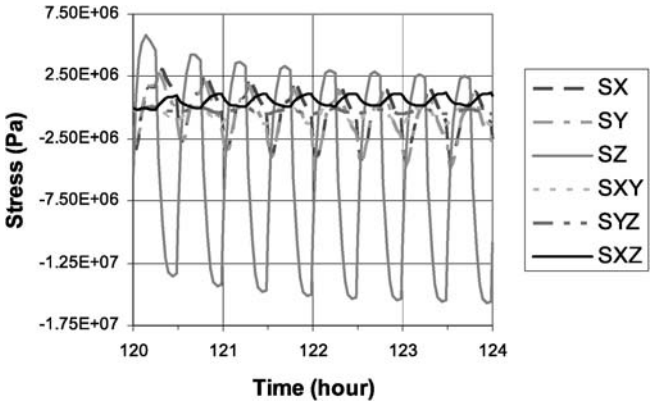


FIGURE 25.21. Time-dependent stresses in the solder joint due to cyclic temperatures.

under the corner of the chip as indicated in Figure 25.24. The dynamic stresses and strains in the same solder ball due to cyclic temperature variation are shown in Figure 25.21 and Figure 25.22, respectively. It can be seen from Figure 25.21 and Figure 25.22 the stresses and strains oscillate within certain ranges in accordance with the cyclic temperature load. Figure 25.23 shows the normal stress (σ_z) and the normal viscoplastic strain hysteresis loops for multiple cycles. For viscoplastic analysis, it is important to study the stress-strain responses for multiple cycles until the hysteresis loops become stabilized. Figure 25.23 indicates that the plastic strain is quite stabilized after the third cycle and the creep response converged after the fourth cycle.

25.3.4. Solder Joint Reliability Estimation

In addition to time-dependent thermal stress modeling, another goal of the simulation is to calculate the plastic work per unit volume (or viscoplastic strain energy density) accumulated per thermal cycle. Over the years, an energy based metric for predicting crack initiation and growth in solder joints was developed and refined [19–21]. The plastic work accumulated during the last cycle is used for crack growth and solder joint fatigue life cor-

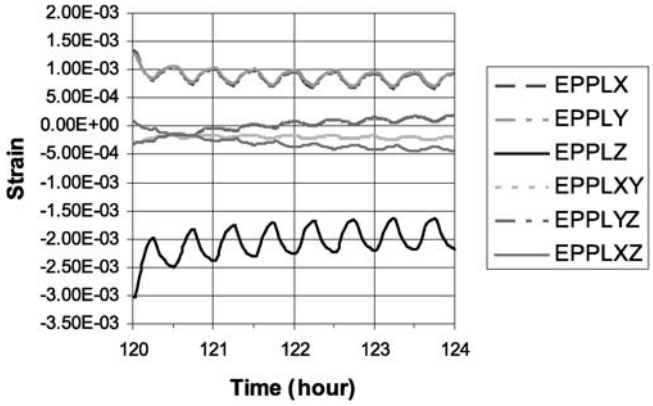


FIGURE 25.22. Time-dependent strains in the solder joint due to cyclic temperatures.

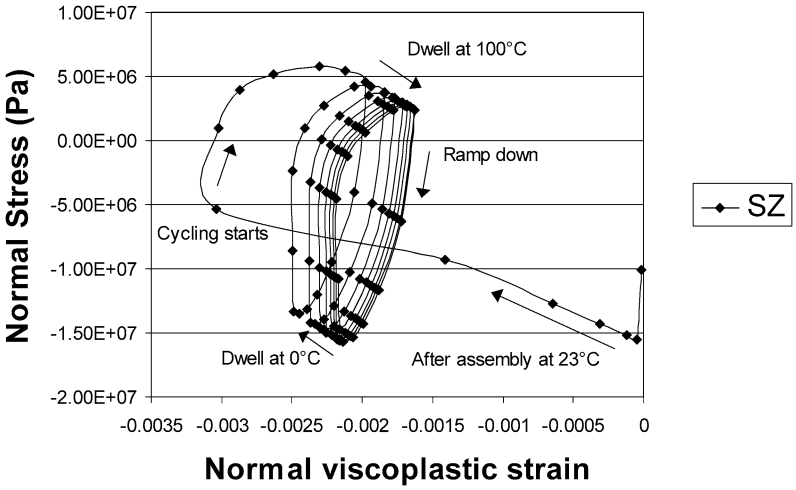


FIGURE 25.23. Hysteresis loops of stress and viscoplastic strain.

relations. The characteristic fatigue life (N_a) for the solder ball with a pad diameter of a can be written as

$$N_a = A(\Delta W_{ave})^B + C(\Delta W_{ave})^D, \tag{25.26}$$

where A, B, C, D are constants that depend on material and solder joint design and assembly process and ΔW_{ave} is the volume averaged viscoplastic strain energy density increment per cycle.

In practice, several complete thermal cycles are simulated in order to establish a stable stress–strain hysteresis loop. The A, B, C, D constants are determined by correlating the simulated results to the temperature cycling test results. Figure 25.24 shows the viscoplastic strain energy density increment calculated from the eighth to the seventh cycle. It is noticed that the solder ball with the highest accumulation in viscoplastic strain energy density is located just under the corner of the die edge. Experimental results show that the

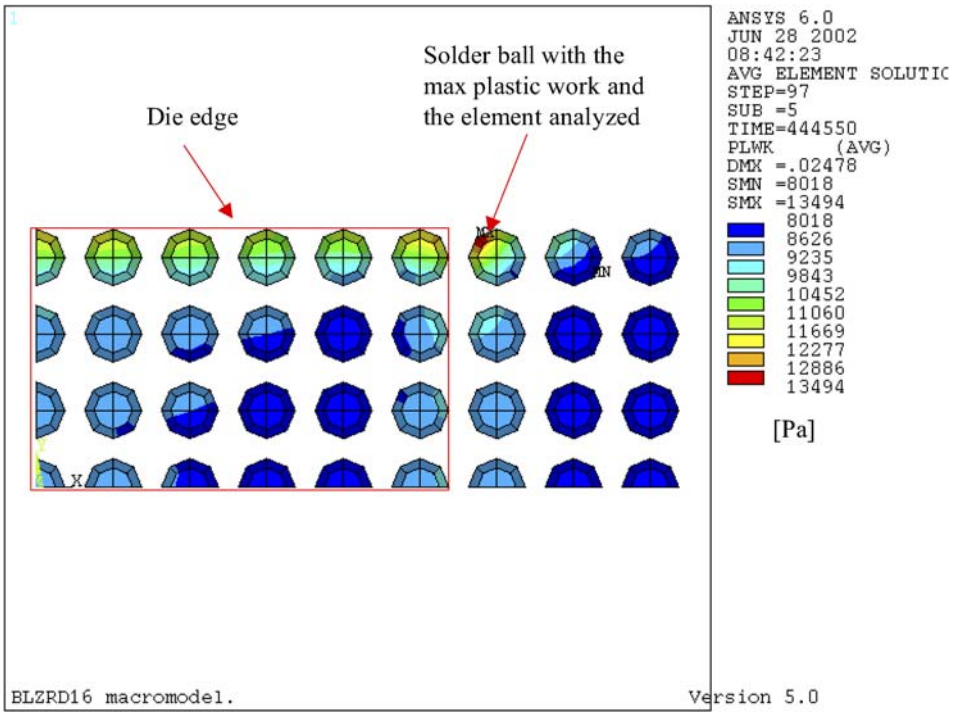


FIGURE 25.24. Calculated plastic work (viscoplastic strain energy density) increment.

balls under the corner of the die edge usually fail first in accelerated temperature cycling test.

25.3.5. Thermal Stress Due to Temperature Cycling: Summary

Dynamic thermal stress as a result of thermal cycling has a great impact on the long-term reliability and integrity of the electronic packages. A non-linear finite element simulation methodology is introduced and applied to simulate the time-dependent thermal stress/strain in the solder joints of a FC-PBGA assembly. The simulated results can be correlated with temperature cycling test results and can be used to predict solder joint reliability under various use and testing conditions.

25.4. RESIDUAL STRESS IN POLYMER-BASED LOW DIELECTRIC CONSTANT (LOW-*k*) MATERIALS

Bulge testing methods have been applied to measure thin film mechanical properties of metals [22–24], semiconductors (Si) [25], and dielectric materials [26–30]. However, it has rarely been applied to polymeric materials [27,29] perhaps the difficulty lies in sample preparation of very thin and pinhole free polymer membranes. The polyimide tested in Refs. [27] and [29] lie in the thickness range of several microns. However, the interlayer dielectric (ILD) materials used in multilevel Very Large Scale Integrated (VLSI) intercon-

nects are thinner films, typically one micron or less. The preparation of such thin polymeric membranes of low-k material without pin-holes is challenging.

In this section, we report a method to prepare thin polymeric membranes for bulge testing. By combining this test with the bending beam measurement of film thermal stresses, we have determined some of the mechanical properties of poly-arylethers (PAE) low-k thin films.

The PAE film, as reported in Ref. [31], was experimentally found to be susceptible to etching by oxidizing chemicals such as acids. It could only survive in H_2O_2 solution for a few minutes and was easily dry etched in O_2 plasma, but it was not attackable by bases. We adopted two approaches, namely, A and B, to be given below for preparing these PAE thin films for bulge tests.

In approach A, a 0.5-μm-thick, low-stress, and low-pressure-chemical-vapor-deposited (LPCVD) nitride film was patterned into square windows by dry etch in a $CF_4 : O_2 = 125 : 25$ mTorr gas mixture at 200 W, followed by Si etching in KOH 30 wt% solution at 90°C. A bilayer of Al (500 Å) top layer/Ti (2000 Å) was e-beam evaporated onto the nitride film. The reason for using the bilayer metal film is that the stress is tensile in the e-beam deposited Ti, but compressive in Al. The low-k material was then spun onto the Al/Ti film, then pre-baked for 1 min at 150°C, and cured for 60 min at 400°C. The thickness of the low-k film was varied by changing the spin rate of the coater and by multiple coatings, and measured by a Tencor Alpha-step Model 200. The nitride layer was then removed in the $CF_4 : O_2 = 125 : 25$ mTorr gas mixture at 100 W, using the Ti layer as the etch stop. Then the Ti film was etched using $H_2O_2 : NH_4OH = 1 : 2$ solution at room temperature and the thin Al was removed by KOH 30 wt% solution, leaving a window-like low-k film for bulge testing [32] see Figure 25.25(a) for a schematic cross-sectional view.

In approach B [32,33], we use a Si membrane as the sacrificial layer. The starting silicon substrate was a wafer of boron diffusion-doped to a depth of 3 μm with a concentration higher than 5×10^{19} B/cm^3. A wet thermal oxide of 5000 Å was grown on the Si substrate at 1050°C, followed by a deposition of 1500 Å LPCVD stoichiometric nitride. The residual stresses in the thermal oxide and nitride were about 330 MPa compressive and 1 GPa tensile, respectively, so the bending moment exerted on the Si substrate by the combined dielectric layers was minimized. A window was opened on the backside by dry etching the nitride as mentioned previously, and followed by wet etching the thermal oxide in buffered oxide etchant (BOE). Then the Si was etched in two steps; first, the etching by 30 wt% KOH at 75°C to a thickness of about 100 μm, and then the etching by ethylene diamine pyrocatechol (EDP) at 90°C until the etch stopped on the B-doped region. The result was a uniform Si membrane of 4.6 μm. The uniformity was better than 0.16 μm over a 1600 μm × 1600 μm area as measured by a Twyman-Green (TG) laser interferometer [34]. Finally the nitride was stripped in $CF_4 : O_2 = (125 : 25$ mTorr) gas mixture at 100 Watts, and the underlying oxide was removed by BOE etch. The process flow is schematically shown in Figure 25.25(b).

The low-k material was spun onto the Si membrane, then pre-baked and cured. To remove the Si membrane, it was first dry etched in the $CF_4 : O_2$ (125 : 25 mTorr) gas mixture at 100 Watts for 1 min in order to remove the native oxide, and then immediately loaded into a home-made pulsed XeF_2 etcher. The chamber pressure was 300 mTorr, and the etching time was 2 min for each cycle, and 2 to 3 cycles was performed. The completion of etching was determined visually through a plexi-glass cover on the etcher, aided by the observation that the low-k film itself is transparent. In comparing approach A and B, we found that the latter is less likely to break the polymer membrane. The key advantage of

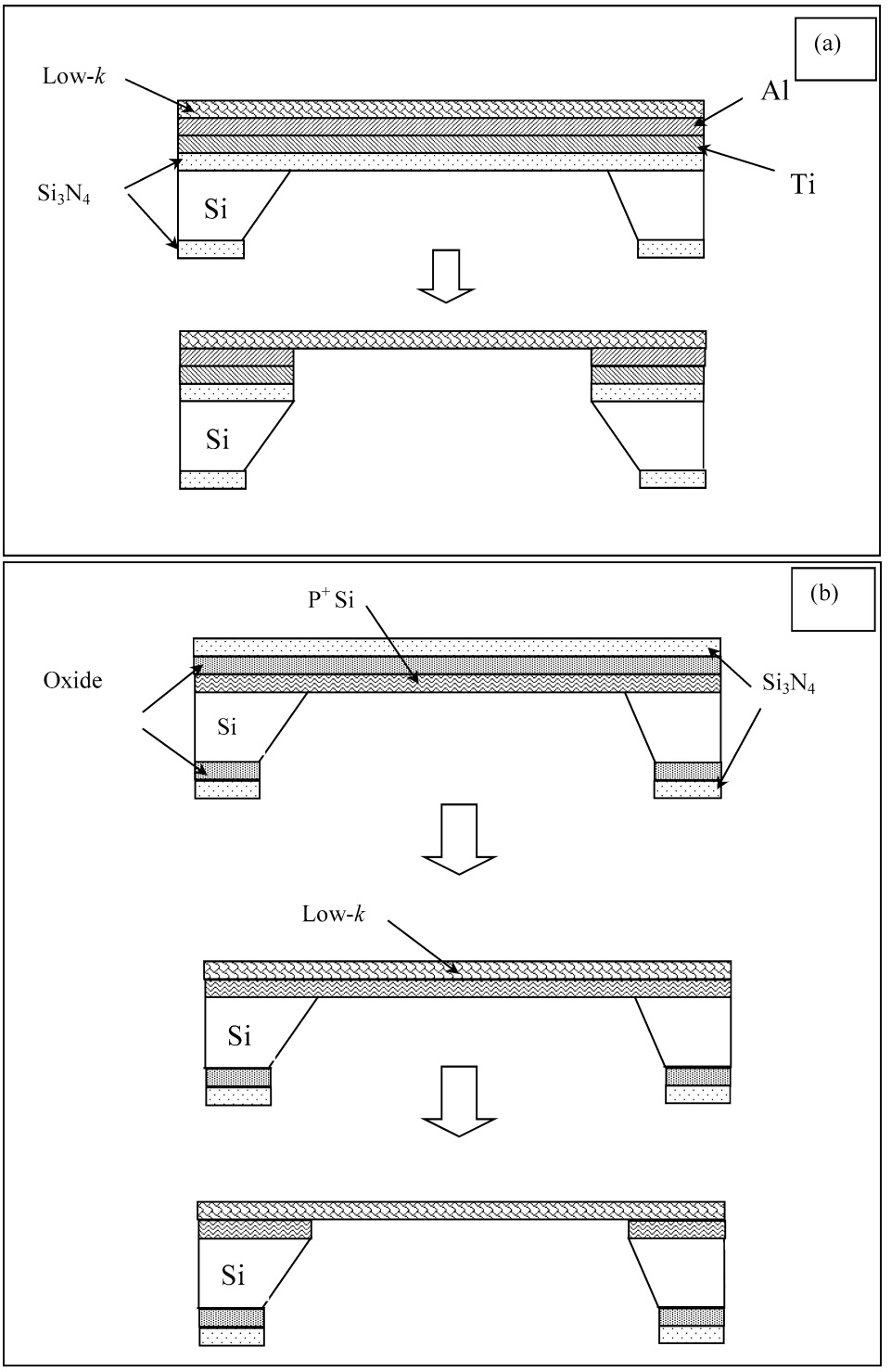


FIGURE 25.25. Cross-sectional view of the steps used to prepare the low-k membrane (a) Al/Ti/Si_3N_4 as the sacrificial layers, (b) P^+ (boron-doped) Si as the sacrificial layer.

the XeF_2 etcher is that it does not affect the thin polymer film and produces no pinholes in the film.

The PAE test film was prepared using approach B. It was coated at 3000 rpm for 30 sec, and the thickness was measured to be 4.6 ± 0.1 kÅ.

The raw data from a bulge test provides the total deflection of the membrane (δ) versus differential pressure (P). For a square membrane test sample, there are two approaches to extract the initial stress and biaxial modulus from the raw data. The first is a direct fit between P and δ, using the analytical formula, e.g., given by Pan et al. [26]:

$$P = \frac{C_1 \cdot t}{R^2} \cdot \sigma_0 \cdot \delta + \frac{C_2 \cdot f(\gamma) \cdot t}{R^4} \cdot \frac{E}{1-\gamma} \cdot \delta^3, \tag{25.27}$$

where $C_1 = 3.41$, $C_2 = 1.37$, $f(\gamma) = 1.446 - 0.427 \cdot \gamma$. R is half of the length of each side of the square, γ is Poisson's ratio, σ_0 is the residual stress, t is the film thickness and $E/1-\gamma$ is the biaxial modulus of the film. The second method of data extraction is to calculate the stress and strain at the center of the membrane [32]:

$$\sigma = \frac{P \cdot R^2}{C_3 \cdot t \cdot \delta}, \tag{25.28}$$

$$\varepsilon = C_4 \frac{\delta^2}{R^2}, \tag{25.29}$$

where $C_3 = 3.04$ and $C_4 = 0.451$ for square-shaped membranes. Poisson's ratio was assumed to be 0.42.

The maximum membrane deflection (δ) versus the differential pressure (P) curve for the dense polymeric film is shown in Figure 25.26 to be nearly linear. Since TG laser interferometer was used to measure the maximum membrane deflection, we were limited by the total number of countable fringes up to ~65. The initial stress in the film was calculated to be 35.2 ± 0.2 MPa after a curve fitting of Figure 25.26 using Equation (25.27). It is obvious that if we calculate the biaxial modulus using Equation (25.27), the result will be very unreliable because of the large initial stress. If we increase pressure for larger deformation, we are limited by the countable fringes, besides we might cause plastic deformation and creep in the polymeric film. Therefore, Equations (25.28) and (25.29) were used to calculate the stress (σ) and strain (ε). In the calculation, the small membrane deflection region was rejected, because in this region the measurement accuracy was relatively low. The biaxial modulus was estimated to be 6.40 ± 0.35 GPa on the basis of the curve fitting of the σ–ε curve, see the inset in the lower right corner of Figure 25.26.

In order to estimate the coefficient of thermal expansion of the low-k films, the thermal stress versus temperature curve was measured using a Flexus 2-300 machine (KLA-Tencor Corporation, San Jose, CA) under a nitrogen flow of 3 liter/min. The substrates used were 320-μm thick double-side-polished Si wafers. The low-k film was coated on the front side of the wafer and the wafer curvature was measured from the backside. The PAE films were coated at 3000 rpm for 30 seconds. Four samples of each kind were tested. We found that the initial stress as well as the slope of the stress-temperature curve varies significantly. The initial stress of the PAE film varied from 32 to 68 MPa. This is tentatively attributed to the difference in the coating solution and the cooling temperature profile during hard-bake. In general, the initial stress measured by bulge testing is lower than that by the Flexus testing; this may be due to relaxation of the film stress after part of

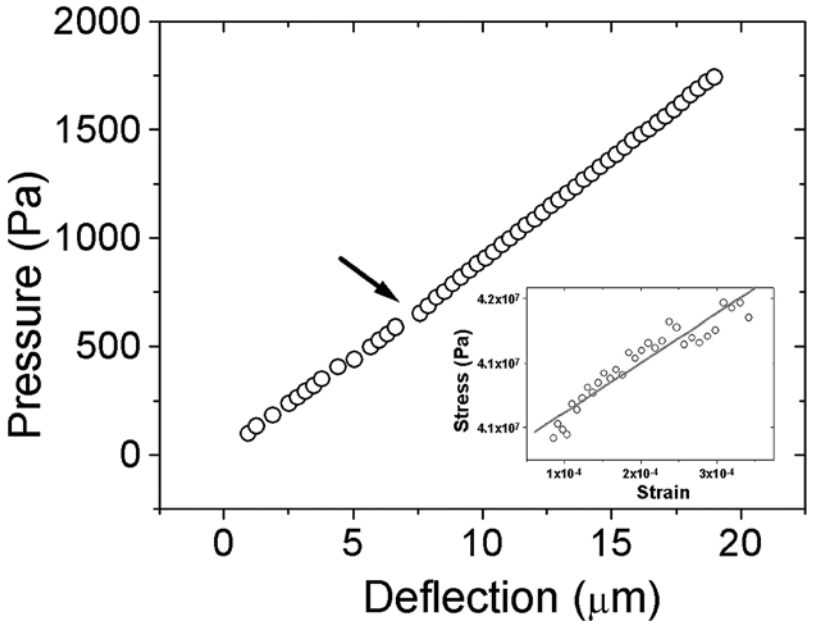


FIGURE 25.26. The total deflection of a 0.46-μm thick dense PAE membrane versus the applied differential pressure. Lower-right inset: the stress versus strain at the center of the membrane.

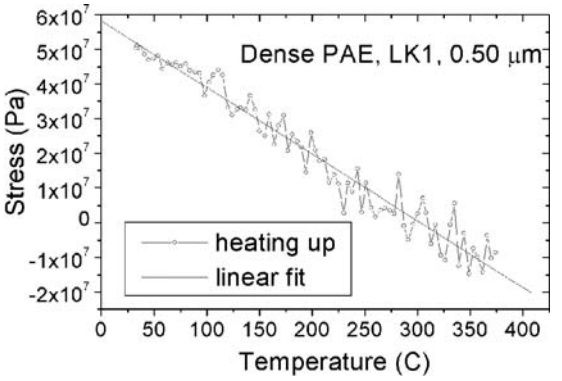


FIGURE 25.27. The stress versus temperature relationship of a blanket PAE film on a 4-inch (100) Si wafer.

the substrate Si was removed in the bulge testing. The slope of the stress-temperature curve is linked to the biaxial modulus and CTE of the film through the following equation:

$$\frac{d\sigma}{dT} = \left(\frac{E}{1-\gamma}\right)_f \cdot (\alpha_s - \alpha_f), \tag{25.30}$$

where $(E/1-\gamma)_f$, α_s, and α_f are the biaxial modulus of the film, the substrate CTE and the film CTE, respectively.

Figure 25.27 shows a typical stress versus temperature curve, where the temperature profile was ramping up from room temperature (24°C) to 375°C in 1.75 hr, held at 375°C

for 0.25 hr, and ramped down to room temperature in 2 hr. The slope of the PAE film ranges from 1.526×10^5 Pa/°C to 1.923×10^5 Pa/°C, and the corresponding CTE calculated range from 26.8–32.6 ppm/°C.

In summary, we have extended the bulge testing method to polymeric thin film of sub-micron thickness. The biaxial modulus and thermal expansion coefficient of the PAE film was obtained.

REFERENCES

1. F.M. D'Heuvle, Metallurgical topics in silicon device interconnections: Thin film stresses, International Materials Reviews, 34(2), p. 53 (1989).
2. J.E. Ashton and J.M. Whitney, Theory of Laminated Plates, Vol. IV, Progress in Materials Science Series, Stamford, CN: Technomic, 153, 1970.
3. E. Suhir, Predicted bow of plastic packages of integrated circuit (IC) device, Thermal Stress and Strain in Microelectronics Packaging, Van Nostrand Reinhold, New York, 1993.
4. R. Jones, Mechanics of Composite Materials, Hemisphere Publishing Corporation, 1975.
5. R.J. Jaccodine and W.A. Schlegel, Measurement of strains at Si-SiO_2 interface, J. Appl. Phys., 37(6), p. 2429 (1966).
6. P.A. Flinn, D.S. Gardner, and W.D. Nix, Measurement and interpretation of stress in aluminum-based metallization as a function of thermal history, IEEE Trans. ED, 34(3), p. 689 (1987).
7. R. Glang, R.A. Holmwood, and R.L. Rosenfeld, Determination of stress in films on single crystalline silicon substrates, Rev. Sci. Instru., 36(1), p. 7 (1965).
8. A. Segmüller, J. Angilelo, and S.J. LaPlaca, Automatic x-ray diffraction measurement of the lattice curvature of substrate wafers for the determination of linear strain patterns, J. Appl. Phys., 51(12), p. 6224 (1980).
9. S.G. Malhotra, Z.U. Rek, S.M. Yalisove, and J.C. Bilello, Depth dependence of residual strains in polycrystalline Mo thin films using high-resolution x-ray diffraction, J. Appl. Phys., 79(9), p. 6872 (1996).
10. K. Ajito, J.P.H. Sukamto, L.A. Nagahara, K.Hashimoto, and A. Fujishima, Strain imaging analysis of Si using Raman microscopy, J. Va. Sci. Technol. A, 13(3), p. 1234 (1995).
11. J. Groenen, A. Mlayah, and R. Carles, Strain in InAs islands grown on InP(001) analyzed by Raman spectroscopy, Applied Physics Letters, 69(7), p. 943 (1996).
12. Y. Guo, D. Mitchell, V. Sarihan, and T. Lee, A testing method and device for intrinsic stress measurement in wafer bumping process, IEEE Transactions On Components And packaging Technologies, 23(2) (2000).
13. G. G. Stoney, Proc. R. Soc. London, Ser. A, 82, p. 172 (1909).
14. R. W. Hoffman, The mechanical properties of thin condensed films, Phys. Thin Film, vol. 3, 1966, pp. 211–273.
15. D. Post, B. Han, and P. Ifju, High Sensitivity Moire: Experimental Analysis for Mechanics and Materials, Chap. 2, Springer-Verlag, New York, 1994.
16. L. Mercado, Effect of die size and gap height on die stress in flip-chip PBGA packages, Motorola Internal Report, 1999.
17. L. Anand, Constitutive equations for hot-working of metals, International Journal of Plasticity, 1, pp. 213–231 (1985).
18. S.B. Brown, K.H. Kim, and L. Anand, An internal variable constitutive model for hot working of metals, International Journal of Plasticity, 5, pp. 95–130 (1989).
19. R. Darveaux, K. Banerji, A. Mawer, and G. Dody, Reliability of plastic ball grid array assembly, in J. Lau, Ed., Ball Grid Array Technology, McGraw-Hill, NY, 1995.
20. R. Darveaux, Solder joint fatigue life model, Proc. TMS Annual Meeting, Orlando, FL, 1997, pp. 213–218.
21. R. Darveaux, Effect of simulation methodology on solder joint crack growth correlation, ECTC, Las Vegas, NV, 2000, pp. 1048–1063.
22. T. Tsakalakos, The bulge test: a comparison of theory and experiment for isotropic and anisotropic films, Thin Solid Films, 75, pp. 293–305 (1981).
23. E.I. Bromley, J.N. Randall, D.C. Flanders, and R.W. Mountain, A technique for the determination of stress in thin films, Journal of Vacuum Science and Technology, B1(4), pp. 1364–1366 (1983).
24. D. Maier-Schneider, A. Ersoy, J. Maibach, D. Schneider, and E. Obermeier, Influence of annealing on elastic properties of LPCVD silicon nitride and LPCVD polysilicon, Sensors and Materials, 7(2), pp. 121–129 (1995).

25. J.J. Vlassak and W.D. Nix, A new bulge test technique for the determination of Young's modulus and Poisson's ratio of thin films, Journal of Material Research, 7(12), pp. 3242–3249 (1992).
26. J.Y. Pan, P. Lin, F. Maseeh, and S.D. Senturia, Verification of FEM analysis of load-deflection methods for measuring mechanical properties of thin films, in Technical Digest IEEE Solid-state Sensors and Actuators Workshop, IEEE, New York, 1990, pp. 70–73.
27. G.F. Cardinale and R.W. Tustison, Fracture strength and biaxial modulus measurement of plasma silicon nitride films, Thin Solid Films, 207, pp. 126–130 (1992).
28. M.G. Allen, M. Mehregany, R.T. Howe, and S.D. Senturia, Microfabricated structures for the in-situ measurement of residual stress, Young's modulus, and ultimate strain of thin films, Applied Physics Letters, 51, pp. 241–243 (1987).
29. O. Tabata, K. Kawahata, S. Sugiyama, and I. Igarashi, Mechanical property measurement of thin films using load-deflection of composite rectangular membranes, Sensors and Actuators, 20, pp. 135–141 (1989).
30. D. Maier-Schneider, J. Maibach, and E. Obermeier, Computer-aided characterization of the elastic properties of thin films, Journal of Micromechanical and Microengineering, 2, pp. 173–175 (1992).
31. Y.H. Xu, Y.-P. Tsai, K.N. Tu, B. Zhao, Q.-Z. Liu, M. Brongo, G.T.T. Sheng, and C.H. Tung, Dielectric property and microstructure of a porous polymer material with ultra low dielectric constant, Applied Physics Letters, 75, pp. 853–855 (1999).
32. D.W. Zheng, Ph.D. dissertation, University of California at Los Angeles, USA, 1999.
33. R.E. Acosta, J.R. Maldonado, L.K. Towart, and J.M. Warlaumont, B-Si masks for storage ring X-ray lithography, Solid State Technology, 27, pp. 205–208 (1984).
34. D. Zheng, R.P. Hwang, X.Y. Dou, C.P. Yeh, M. Prakash, K. Boardman, and G. Ridsdale, Warpage analysis of 144-pin TQFP during reflow using image processing, Proceedings of 47th Electronic Components and Technology Conference, IEEE, New York, NY, 1997, pp. 1176–1181.

Index

Numbers in bold indicate the volume number.

abrasion 602 (**1**)
absorption coefficient 86 (**1**)
absorption rates 65 (**1**)
accelerated life test (ALT) 203 (**2**), 210 (**2**), 212 (**2**), 214 (**2**), 217 (**2**), 218 (**2**), 230 (**2**), 255 (**2**), 256 (**2**)
– model 223 (**2**)
accelerated stress testing (AST) 454 (**1**), 190 (**2**), 194 (**2**), 200 (**2**)
accelerated test conditions 215 (**2**)
accelerated testing 316 (**1**), 348 (**1**), 385 (**1**), 514 (**1**), 4 (**2**), 102 (**2**), 213 (**2**), 215 (**2**), 222 (**2**), 683 (**2**)
accelerated tests 454 (**1**), 301 (**2**)
accelerated thermal cycling (ATC) 485 (**1**), 257 (**2**), 291 (**2**)
acceleration factor 216 (**2**), 218 (**2**), 225 (**2**), 500 (**2**)
acceleration models 388 (**1**)
activation energy 44 (**1**), 215 (**1**), 367 (**1**), 371 (**1**), 576 (**1**), 127 (**2**), 130 (**2**)
adherend failure 6 (**2**)
adhesion 40 (**1**), 137 (**1**), 141 (**1**)–143 (**1**), 146 (**1**), 156 (**1**)–157 (**1**), 165 (**1**)–166 (**1**), 169 (**1**), 170 (**1**), 175 (**1**)–177 (**1**), 263 (**1**), 269 (**1**), 273 (**1**), 538 (**1**), 392 (**2**), 404 (**2**), 405 (**2**), 425 (**2**), 426 (**2**), 431 (**2**)–433 (**2**), 441 (**2**)–445 (**2**), 452 (**2**), 506 (**2**)–508 (**2**), 547 (**2**), 620 (**2**)
– failures 166 (**2**), 318 (**2**)
– promoter 392 (**2**), 506 (**2**)
– strength 404 (**2**), 409 (**2**), 414 (**2**), 419 (**2**), 426 (**2**), 432 (**2**), 619 (**2**)
adhesive 172 (**1**), 302 (**1**), 310 (**1**), 463 (**1**), 479 (**1**), 480 (**1**), 506 (**1**)–508 (**1**), 519 (**1**), 520 (**1**), 523 (**1**), 6 (**2**)–12 (**2**), 116 (**2**), 383 (**2**), 386 (**2**), 387 (**2**), 390 (**2**), 398 (**2**), 399 (**2**), 432 (**2**), 442 (**2**), 449 (**2**), 466 (**2**), 488 (**2**), 496 (**2**), 497 (**2**), 501 (**2**), 503 (**2**), 508 (**2**)–510 (**2**), 513 (**2**), 517 (**2**), 522 (**2**), 529 (**2**), 554 (**2**), 565 (**2**), 571 (**2**), 572 (**2**), 574 (**2**), 597 (**2**), 601 (**2**), 612 (**2**), 613 (**2**), 617 (**2**), 619 (**2**)
– bonding 302 (**1**), 506 (**1**), 7 (**2**), 116 (**2**), 397 (**2**), 460 (**2**), 469 (**2**), 488 (**2**), 503 (**2**), 514 (**2**), 515 (**2**), 572 (**2**), 573 (**2**), 602 (**2**), 606 (**2**)
– failure 310 (**1**), 6 (**2**)
– strength 172 (**1**), 8 (**2**), 260 (**2**), 405 (**2**), 514 (**2**), 532 (**2**), 546 (**2**)
adsorption theory 463 (**2**), 464 (**2**), 508 (**2**)
advanced configuration power interface (ACPI) 293 (**1**)
Al–Ni bond interfaces 124 (**2**)
Al–Ni interface 124 (**2**)
ALD-SAM coating 318 (**2**)
AlGaAs 139 (**1**)
Alloy 42 substrate 373 (**1**)
analysis of variance (ANOVA) 230 (**1**), 232 (**1**), 251 (**1**), 319 (**1**)
analytical, computational, and experimental solutions (ACES) 324 (**2**)
– methodology 324 (**2**)
analytical modeling 270 (**1**), 557 (**1**), 6 (**2**), 16 (**2**), 223 (**2**), 261 (**2**), 330 (**2**), 334 (**2**), 697 (**2**)
analytical stress model 272 (**1**), 13 (**2**), 16 (**2**), 474 (**2**)
analytical thermal model 78 (**1**)
Anand model 430 (**1**), 452 (**1**)
angular acceleration 328 (**2**)

anisotropic conductive adhesive (ACA) 527 (**2**), 529 (**2**), 533 (**2**), 536 (**2**), 572 (**2**), 619 (**2**)
anisotropic conductive film (ACF) 527 (**2**), 532 (**2**), 612 (**2**), 619 (**2**), 625 (**2**)
anisotropic conductive paste (ACP) 527 (**2**), 532 (**2**)
anisotropic elastic constants 156 (**1**)
anisotropic thermal conductivity 186 (**1**)
anisotropic thermal expansion 186 (**1**)
any layer inner via hole (ALIVH) 271 (**2**)
apparent Bragg wavelength shift 98 (**1**)
apparent elastic modulus 435 (**1**), 437 (**1**), 439 (**1**)
approximate fully cure dependent model 34 (**1**)
approximate method 11 (**1**)
approximation method 238 (**1**)
arc-discharge 186 (**1**)
area array packages (AAPs) 284 (**1**)
area array technology 283 (**1**), 398 (**1**)
Arrhenius equation 61 (**1**), 365 (**1**), 370 (**1**), 130 (**2**), 186 (**2**)
atomic force microscopy (AFM) 148 (**1**), 192 (**1**), 604 (**1**), 515 (**2**)
atomic layer deposition (ALD) 316 (**2**)
Au–Al interfaces 122 (**2**)
Au–Al wire bonds 122 (**2**)
Au–Au interfaces 123 (**2**)
axial elastic compression 560 (**1**)
axial free vibrations 674 (**1**)
axial power variation 75 (**1**)
axial spring constant 560 (**1**)
axial strength testing 616 (**1**)
axisymmetrical bending 687 (**2**)

ball bond shear test 89 (**2**), 102 (**2**), 117 (**2**)
ball grid array (BGA) 6 (**1**), 284 (**1**), 291 (**1**), 313 (**1**), 318 (**1**), 411 (**1**), 481 (**1**), 557 (**1**), 8 (**2**), 109 (**2**), 138 (**2**), 257 (**2**), 270 (**2**), 281 (**2**), 370 (**2**), 383 (**2**), 389 (**2**), 573 (**2**), 611 (**2**)
– reliability 401 (**1**)
– solder balls 398 (**1**)
– solder joint 400 (**1**)
ballistic electron transport 182 (**1**)
bare fiber 269 (**1**), 271 (**1**)
$BaTiO_3$ 112 (**1**)
beam bending 687 (**2**)
beam propagation method (BPM) 25 (**2**)
belt slide apparatus 605 (**1**)
benzocyclobutene (BCB) 142 (**1**), 79 (**2**), 108 (**2**), 137 (**2**), 313 (**2**)
BeO 484 (**1**)
BGO 112 (**1**)
bi-material assemblies 6 (**2**)
bi-material model 651 (**2**)
bias on the substrate 197 (**1**)
biaxial elastic modulus 147 (**1**)
bimaterial plate (BMP) 690 (**2**), 697 (**2**)
bimodality 577 (**1**)
binary phase diagram 354 (**1**)
BiTe 484 (**1**)
bithermal loading 481 (**1**)
board level model 322 (**1**)
body of knowledge (BOK) 283 (**1**)
Boltzmann-Arrhenius equation 225 (**2**)
bond failure 87 (**2**)
bond pad metallurgy 100 (**2**)
bond strength 103 (**2**), 108 (**2**), 396 (**2**), 503 (**2**), 508 (**2**)
bondability 79 (**2**), 101 (**2**), 110 (**2**), 185 (**2**)
bonding layer 14 (**2**), 475 (**2**), 604 (**2**)
bonding machine 77 (**2**), 80 (**2**), 90 (**2**), 110 (**2**)
bonding materials 6 (**2**), 15 (**2**), 184 (**2**), 218 (**2**), 384 (**2**), 473 (**2**)
bonding matrix 479 (**2**)
bonding mechanism 419 (**2**)
bonding pad 104 (**2**), 306 (**2**), 606 (**2**)
bonding strength 352 (**1**), 404 (**2**), 419 (**2**)
bonding wire 80 (**2**)
bondline 385 (**2**), 388 (**2**), 393 (**2**), 394 (**2**), 398 (**2**), 511 (**2**), 514 (**2**)
borosilicate glass 142 (**1**)
boundary, initial, and loading (BIL) 324 (**2**)
– conditions 324 (**2**)
boundary element method (BEM) 551 (**1**), 324 (**2**)
Bragg grating (BG) 65 (**1**), 4 (**2**), 16 (**2**)
Bragg wavelength 65 (**1**)
breaking strength 599 (**1**), 98 (**2**)
brittle fracture 418 (**1**), 454 (**1**), 204 (**2**)
brittle intermetallic compounds 313 (**1**), 122 (**2**), 225 (**2**)
brittle intermetallic layers 318 (**1**)
brittle materials 221 (**1**)
Bueche-Zhurkov equation 227 (**2**), 229 (**2**)
buffered oxide etchant (BOE) 704 (**2**)
bulk composite 185 (**1**)
bulk index grating 67 (**1**)
bulk relaxation modulus 21 (**1**)

buoyancy 361 (**1**), 362 (**1**), 152 (**2**), 172 (**2**)
– force 361 (**1**)
– time 362 (**1**)
Burgers model 6 (**1**)
burn-in 310 (**1**), 145 (**2**), 192 (**2**), 205 (**2**), 220 (**2**), 301 (**2**), 625 (**2**)

camera stage 463 (**1**)
cantilever 192 (**1**), 305 (**2**), 307 (**2**), 318 (**2**), 483 (**2**), 484 (**2**)
carbon nanotube (CNT) 181 (**1**)
catalyst deposition 189 (**1**)
CCGA assembly 306 (**1**)
CdTe 112 (**1**)
central composite design (CCD) 233 (**1**)
ceramic ball grid array (CBGA) 283 (**1**), 294 (**1**)–297 (**1**), 299 (**1**), 303 (**1**), 305 (**1**), 309 (**1**), 386 (**1**), 489 (**1**)
ceramic column grid array (CCGA) 283 (**1**), 286 (**1**)
ceramic packages 293 (**1**)
channel dispersion 127 (**1**)
chaotic response 646 (**1**)
characteristic life times 342 (**1**)
chemical mechanical planarization (CMP) 141 (**1**), 155 (**1**), 192 (**1**)
chemical mechanical polishing (CMP) 193 (**1**), 240 (**2**)
chemical potential 324 (**1**)
chemical reaction 6 (**1**), 144 (**1**), 164 (**1**), 330 (**1**)
chemical vapor deposition (CVD) 144 (**1**), 186 (**1**), 187 (**1**)
chip carrier 310 (**1**), 483 (**1**), 501 (**1**), 263 (**2**), 563 (**2**), 623 (**2**)
chip density 71 (**2**)
chip on flex (COF) 612 (**2**)
chip on glass (COG) 612 (**2**)
chip scale package (CSP) 284 (**1**), 286 (**1**), 313 (**1**), 335 (**1**), 411 (**1**), 557 (**1**), 109 (**2**), 135 (**2**), 149 (**2**), 257 (**2**), 268 (**2**), 272 (**2**), 276 (**2**), 281 (**2**), 406 (**2**), 413 (**2**), 528 (**2**), 611 (**2**)
chip-on-board (COB) 94 (**2**), 108 (**2**)
chips per wafer (CPW) 136 (**2**)
chromatic mirror 115 (**1**)
cladding radius 120 (**1**)
cladding refractive index 69 (**1**), 84 (**1**), 120 (**1**)
Class I 206 (**2**)
Class II 207 (**2**)
Class III 207 (**2**)
classic wetting theory 358 (**1**)
coated fibers 272 (**1**)
coefficient of hygro expansion (CHE) 531 (**1**)
coefficient of hygroscopic swelling (CHS) 494 (**1**), 496 (**1**)
– mismatch 499 (**1**)
coefficient of thermal expansion (CTE) 5 (**1**), 186 (**1**), 199 (**1**), 212 (**1**), 215 (**1**), 253 (**1**), 271 (**1**), 294 (**1**), 380 (**1**), 430 (**1**), 459 (**1**), 483 (**1**), 524 (**1**), 530 (**1**), 532 (**1**), 635 (**1**), 3 (**2**), 11 (**2**), 14 (**2**), 32 (**2**), 38 (**2**), 39 (**2**), 75 (**2**), 87 (**2**), 97 (**2**), 122 (**2**), 129 (**2**), 136 (**2**), 160 (**2**), 234 (**2**), 241 (**2**), 273 (**2**), 305 (**2**), 350 (**2**), 373 (**2**), 456 (**2**), 478 (**2**), 481 (**2**), 488 (**2**), 619 (**2**), 653 (**2**), 670 (**2**), 680 (**2**), 685 (**2**), 686 (**2**), 706 (**2**)
– mismatch 219 (**1**), 289 (**1**), 294 (**1**), 395 (**1**), 405 (**1**), 499 (**1**), 9 (**2**), 127 (**2**), 130 (**2**), 139 (**2**), 149 (**2**), 185 (**2**), 273 (**2**), 275 (**2**), 307 (**2**), 373 (**2**), 496 (**2**), 497 (**2**), 502 (**2**), 522 (**2**), 689 (**2**), 691 (**2**)
Coffin-Manson equation 225 (**2**), 559 (**2**)
Coffin-Manson fatigue model 432 (**1**)
Coffin-Manson model 224 (**1**), 446 (**1**)
Coffin-Manson relationship 224 (**1**), 293 (**1**), 300 (**1**)
coherent beams 476 (**1**)
cohesive failure 416 (**1**), 6 (**2**), 386 (**2**), 506 (**2**)
collected volatile condensable materials (CVCM) 511 (**2**)
commercial-off-the-shelf (COTS) 283 (**1**)
compact tension (CT) 244 (**2**)
compatibility 129 (**1**), 377 (**1**), 131 (**2**), 302 (**2**)
complete failure 243 (**2**)
complete optical failure 4 (**2**)
compliance 10 (**1**), 15 (**1**), 16 (**1**), 22 (**1**), 24 (**1**)–26 (**1**), 42 (**1**), 147 (**1**), 162 (**1**), 294 (**1**), 408 (**1**), 688 (**1**), 5 (**2**), 59 (**2**), 77 (**2**), 136 (**2**), 149 (**2**), 293 (**2**), 473 (**2**), 500 (**2**)
– functions 10 (**1**)
compliant adhesive 473 (**2**), 522 (**2**)
component forward/backward compatibility 386 (**1**)
component level model 322 (**1**)
component thermal stability 386 (**1**)
compound 5 (**1**), 26 (**1**), 47 (**1**), 144 (**1**), 293 (**1**)
compressive strength 392 (**2**)
compressive stress 267 (**2**), 355 (**2**), 357 (**2**), 697 (**2**)
computer-aided design (CAD) 136 (**1**)

conducting failure analysis 177 (**2**), 180 (**2**), 269 (**2**)
conductive anodic filament (CAF) 264 (**2**), 267 (**2**)
consistent composite plate model 685 (**2**)
constant load effect 646 (**1**)
constitutive models 4 (**1**), 212 (**1**), 262 (**1**), 435 (**1**)
contact force 147 (**2**), 219 (**2**), 576 (**2**), 581 (**2**)
contact resistance 181 (**1**), 197 (**1**), 76 (**2**), 139 (**2**), 542 (**2**), 572 (**2**), 575 (**2**), 587 (**2**), 595 (**2**), 599 (**2**), 600 (**2**), 613 (**2**), 615 (**2**), 623 (**2**)
contemporary systems 463 (**1**)
continuous stiffness method (CSM) 163 (**1**)
conventional passive alignment methods 158 (**2**)
converse piezoelectric effect 138 (**1**)
convolution integral 11 (**1**)
cooling chamber 469 (**1**)
coplanar waveguide (CPW) 313 (**2**)
copper heat sink 349 (**2**), 350 (**2**), 354 (**2**)
copper pad 491 (**1**), 97 (**2**), 374 (**2**), 596 (**2**), 601 (**2**)
copper wire 97 (**2**)
core radius 84 (**1**), 120 (**1**)
core refractive index 84 (**1**), 120 (**1**)
corner adhesive bonds 302 (**1**)
corrosion 6 (**1**), 144 (**1**), 155 (**1**), 262 (**1**), 269 (**1**), 273 (**1**), 274 (**1**), 388 (**1**), 389 (**1**), 395 (**1**), 100 (**2**), 119 (**2**), 130 (**2**), 182 (**2**), 184 (**2**), 219 (**2**), 255 (**2**), 256 (**2**), 265 (**2**), 286 (**2**), 287 (**2**), 396 (**2**), 404 (**2**), 433 (**2**), 476 (**2**), 477 (**2**), 542 (**2**), 544 (**2**), 557 (**2**), 589 (**2**), 616 (**2**), 620 (**2**)
corrosion inhibitors 557 (**2**), 617 (**2**)
cost effectiveness 136 (**1**), 252 (**1**), 454 (**1**), 203 (**2**), 400 (**2**), 621 (**2**)
coupled axial-radial axisymmetric vibrations 677 (**1**)
coupling coefficient 69 (**1**), 83 (**1**), 122 (**1**)
coupling efficiency 23 (**2**), 28 (**2**), 30 (**2**), 42 (**2**), 48 (**2**), 50 (**2**), 64 (**2**), 152 (**2**), 168 (**2**), 204 (**2**), 373 (**2**), 379 (**2**), 383 (**2**), 489 (**2**), 492 (**2**), 499 (**2**), 510 (**2**), 518 (**2**), 520 (**2**), 523 (**2**)
coupling loss 121 (**1**), 122 (**1**), 27 (**2**), 380 (**2**), 492 (**2**)
crack 165 (**1**), 169 (**1**), 170 (**1**), 220 (**1**), 222 (**1**)–225 (**1**), 418 (**1**)–422 (**1**), 431 (**1**), 527 (**1**)–529 (**1**), 531 (**1**)–537 (**1**), 544 (**1**)–550 (**1**), 594 (**1**)–598 (**1**), 600 (**1**), 601 (**1**), 608 (**1**)–610 (**1**), 612 (**1**)–616 (**1**), 620 (**1**)–625 (**1**), 348 (**2**), 387 (**2**), 413 (**2**), 443 (**2**), 451 (**2**), 457 (**2**), 466 (**2**), 477 (**2**), 494 (**2**), 514 (**2**), 691 (**2**), 701 (**2**)
– detection 243 (**2**), 246 (**2**)
– geometry 610 (**1**)
– opening displacement (COD) 248 (**2**)
– propagation 165 (**1**), 169 (**1**), 223 (**1**), 419 (**1**), 529 (**1**)269 (**2**)
– surface displacement extrapolation method (CSDEM) 531 (**1**), 550 (**1**)
crack growth 222 (**1**), 224 (**1**), 405 (**1**), 433 (**1**), 596 (**1**), 600 (**1**), 614 (**1**), 623 (**1**)
– model 596 (**1**)
– rate 596 (**1**), 623 (**1**)
creep 4 (**1**), 16 (**1**), 22 (**1**), 24 (**1**)–26 (**1**), 42 (**1**), 156 (**1**), 213 (**1**), 252 (**1**), 272 (**1**), 300 (**1**), 301 (**1**), 389 (**1**), 390 (**1**), 393 (**1**)–395 (**1**), 405 (**1**), 454 (**1**), 4 (**2**), 51 (**2**), 219 (**2**), 274 (**2**), 701 (**2**)
– deformation 215 (**1**), 393 (**1**), 395 (**1**), 274 (**2**)
– description 10 (**1**)
– model 212 (**1**)–215 (**1**), 389 (**1**), 440 (**1**), 57 (**2**)
– rate 10 (**1**), 214 (**1**), 301 (**1**), 316 (**1**), 334 (**1**), 392 (**1**), 430 (**1**), 442 (**1**), 491 (**1**), 228 (**2**)
– strain 10 (**1**), 14 (**1**), 42 (**1**), 212 (**1**)–215 (**1**), 334 (**1**), 389 (**1**), 402 (**1**), 430 (**1**), 443 (**1**), 491 (**1**), 138 (**2**), 497 (**2**), 499 (**2**), 511 (**2**)
– strength 390 (**1**)
– test 440 (**1**), 53 (**2**)
– – program 434 (**1**)
creep-fatigue interaction model 432 (**1**)
critical stress 421 (**1**)
cross correlation algorithm 233 (**2**)–235 (**2**)
cross correlation coefficient 236 (**2**)
crystal elastic moduli 147 (**1**)
CT specimen 246 (**2**), 249 (**2**)
CU layers 464 (**1**)
cumulative distribution function (CDF) 388 (**1**), 583 (**1**), 47 (**2**), 180 (**2**), 182 (**2**), 184 (**2**)
cumulative structure function 632 (**2**), 639 (**2**), 641 (**2**)
cure dependent modeling 53 (**1**)
cure-temperature modeling 4 (**1**)
curing 4 (**1**), 5 (**1**), 124 (**1**), 281 (**2**), 384 (**2**), 398 (**2**), 597 (**2**), 606 (**2**)
– polymers 34 (**1**)
– profile 6 (**1**)

– temperature 6 (**1**), 530 (**1**), 7 (**2**), 14 (**2**), 159 (**2**), 387 (**2**), 536 (**2**), 689 (**2**), 691 (**2**)
current density 193 (**1**), 389 (**1**), 220 (**2**)
curvature 146 (**1**), 157 (**1**), 164 (**1**), 182 (**1**), 233 (**1**), 238 (**1**), 240 (**1**), 385 (**1**), 435 (**1**), 483 (**1**), 59 (**2**), 411 (**2**), 412 (**2**), 685 (**2**), 688 (**2**), 690 (**2**), 697 (**2**)
– method 147 (**1**)
cycles-to-failure (CTF) 290 (**1**)–292 (**1**), 296 (**1**), 125 (**2**)
cyclic strain 405 (**1**), 8 (**2**)
cylindrical bending 687 (**2**)

D-optimal design 234 (**1**)
damage accumulation 572 (**1**), 580 (**1**), 593 (**1**)
damage mechanisms 274 (**2**)
dashpots 6 (**1**)
data acquisition 465 (**1**), 605 (**1**)
daughter distribution dynamics 583 (**1**)
debonding 179 (**1**), 263 (**1**), 403 (**2**), 432 (**2**), 461 (**2**), 622 (**2**)
deformed and resonant cantilever 165 (**1**)
degradation rate 223 (**2**)
delamination 6 (**1**), 140 (**1**), 146 (**1**), 164 (**1**)–170 (**1**), 179 (**1**), 263 (**1**), 272 (**1**), 381 (**1**), 525 (**1**), 540 (**1**)–544 (**1**), 547 (**1**), 12 (**2**), 130 (**2**), 185 (**2**), 219 (**2**), 242 (**2**), 260 (**2**), 283 (**2**), 350 (**2**), 353 (**2**), 394 (**2**), 399 (**2**), 403 (**2**), 404 (**2**), 406 (**2**), 414 (**2**), 419 (**2**), 423 (**2**), 427 (**2**), 432 (**2**), 447 (**2**), 507 (**2**), 619 (**2**), 630 (**2**), 633 (**2**), 643 (**2**), 646 (**2**), 685 (**2**)
delamination temperature 539 (**1**), 542 (**1**)
deleterious effect 143 (**1**), 148 (**1**), 281 (**2**), 442 (**2**)
dense wavelength division multiplexing (DWDM) 362 (**2**)
density measurement 48 (**1**)
dependability 270 (**1**), 206 (**2**), 211 (**2**), 213 (**2**)
deposition method 157 (**1**)
design for reliability (DFR) 155 (**1**), 270 (**1**), 429 (**1**), 432 (**1**), 433 (**1**)
design of experiments (DoE) 242 (**1**)
– method 225 (**1**), 226 (**1**)
– procedure 206 (**1**)
design-qualification AST 190 (**2**), 193 (**2**)
deviatoric creep compliance 15 (**1**)
device under test (DUT) 314 (**2**)
die fracture 276 (**2**)
die-substrate assemblies 6 (**2**)
dielectric deposition 189 (**1**)
dielectric dry-etch 189 (**1**)
dielectric film 121 (**1**), 137 (**1**), 140 (**1**)
differential scanning calorimetry (DSC) 42 (**1**), 385 (**2**), 388 (**2**), 444 (**2**), 535 (**2**)
differential structure function 633 (**2**), 634 (**2**), 637 (**2**)
differential thermal expansion 155 (**1**)
diffraction efficiency 120 (**1**)
diffusion 112 (**1**), 113 (**1**), 143 (**1**), 152 (**1**), 155 (**1**), 177 (**1**), 262 (**1**), 323 (**1**), 329 (**1**), 373 (**1**), 385 (**1**), 526 (**1**), 587 (**1**), 102 (**2**), 103 (**2**), 130 (**2**), 184 (**2**), 192 (**2**), 219 (**2**), 225 (**2**), 274 (**2**), 347 (**2**), 354 (**2**), 357 (**2**), 396 (**2**), 433 (**2**), 437 (**2**), 462 (**2**), 508 (**2**), 527 (**2**), 536 (**2**), 542 (**2**), 704 (**2**)
– barrier 143 (**1**), 177 (**1**), 124 (**2**), 284 (**2**)
– coefficient 526 (**1**), 587 (**1**), 434 (**2**)
– creep 274 (**2**)
– kinetics 324 (**1**)
– models 373 (**1**)
– reactions 177 (**1**), 323 (**1**), 324 (**1**), 330 (**1**), 385 (**1**), 130 (**2**)
– theory 434 (**2**), 508 (**2**)
diffusional creep 334 (**1**)
digital fringe multiplication (DFM) 478 (**1**)
digital image correlation (DIC) 233 (**2**), 234 (**2**), 238 (**2**), 241 (**2**), 250 (**2**)
digital light processor (TM) systems 136 (**1**)
digital mirror device (DMD) 301 (**2**)
dipping technique 517 (**2**)
direct chip attachment (DCA) 8 (**2**)
direct numerical method 82 (**1**)
direct piezoelectric effect 668 (**1**)
directional coupler (DC) 330 (**2**)
disclinations 182 (**1**)
discrete element modeling (DEM) 576 (**2**)
dislocation 8 (**2**), 13 (**2**), 185 (**2**), 344 (**2**), 348 (**2**)
– creep 274 (**2**)
– glide 274 (**2**)
– separation (DS) 342 (**2**), 344 (**2**)
dislocations 152 (**1**), 182 (**1**)
dispensing technologies 515 (**2**)
dispersion management filters 129 (**1**)
dispersion slope compensation 128 (**1**), 129 (**1**)
dissolution 145 (**1**), 330 (**1**)
distance from the neutral point (DNP) 219 (**1**), 300 (**1**), 491 (**1**)
distributed Bragg reflector (DBR) 67 (**1**), 113 (**1**), 140 (**1**), 363 (**2**)
– laser diodes (DBR LD) 67 (**1**)
distributed feedback (DFB) 362 (**2**)

divergence angle 26 (**2**), 27 (**2**), 37 (**2**)–40 (**2**)
double simple shear tool (DST) 39 (**1**)
double-sided mirror image assembly 291 (**1**)
downtime 180 (**2**)
drive gear 334 (**2**), 338 (**2**)
drop impact loading 454 (**1**)
drop testing 334 (**1**), 341 (**1**)
drop tests 348 (**1**)
drops-to-failure 319 (**1**), *see also* times-to-failure (TTF)
dual-coated fiber 269 (**1**), 3 (**2**), 15 (**2**)
ductile fracture 220 (**1**), 418 (**1**)
ductile materials 222 (**1**)
Duffing equation 630 (**1**), 633 (**1**), 635 (**1**)
dummy component 416 (**1**)
durability 263 (**1**), 270 (**1**), 302 (**1**), 387 (**1**), 206 (**2**), 253 (**2**), 261 (**2**), 341 (**2**), 392 (**2**), 432 (**2**), 478 (**2**)
DWDM systems 117 (**1**)
dwell time 316 (**1**)
dynamic fatigue 601 (**1**), 609 (**1**)
– test 604 (**1**), 621 (**1**), 624 (**1**)
dynamic loading 270 (**1**), 555 (**1**)
dynamic mechanical analysis (DMA) 19 (**1**), 27 (**1**), 54 (**2**), 385 (**2**)
dynamic mechanical elongation tests 56 (**1**)
dynamic model 163 (**1**)
dynamic physical damage 572 (**1**)
dynamic response 272 (**1**), 555 (**1**), 557 (**1**)
dynamic strength 556 (**1**), 261 (**2**)
dynamic stress 272 (**1**), 556 (**1**), 698 (**2**), 701 (**2**)
dynamic testing 19 (**1**), 27 (**1**)

E-modulus 5 (**1**)
edge dispensing method 163 (**2**)
edge-indentation method 405 (**2**), 428 (**2**)
EDX elemental map 288 (**2**)
effect of contact metalization dissolution 325 (**1**)
effect of damping 678 (**1**), 679 (**1**)
effective contact area 547 (**2**), 591 (**2**)
effective contact length 590 (**2**)
effective refractive index 68 (**1**), 270 (**1**), 273 (**1**)
elastic approximations 17 (**1**)
elastic constant 8 (**2**)
elastic constants 138 (**1**)
elastic energy distribution 572 (**1**)
elastic model 50 (**1**), 212 (**1**), 213 (**1**)
elastic modulus 17 (**1**), 400 (**1**), 442 (**2**), 448 (**2**)
elastic potential energy 145 (**1**)
elastic stability 270 (**1**), 272 (**1**), 16 (**2**)
elastic strain 222 (**1**), 317 (**1**), 430 (**1**), 410 (**2**)
elastic stress 492 (**1**), 7 (**2**)
elastic-plastic-creep analysis 430 (**1**)
elasto-plastic analyses 223 (**1**), 414 (**1**), 416 (**1**)
elastomer 460 (**1**)
electric enthalpy density 671 (**1**)
electrical failure 285 (**2**), 557 (**2**)
electrically conductive adhesive (ECA) 503 (**2**), 527 (**2**), 534 (**2**), 571 (**2**), 612 (**2**)
electro-optic (Pockels) effect 112 (**1**)
electrochemical migration (ECM) 264 (**2**), 265 (**2**), 286 (**2**), 288 (**2**), 290 (**2**)
electrodeposition 145 (**1**)
electroless gold plating 103 (**2**)
electroless nickel and immersion gold (ENIG) 303 (**1**), 310 (**1**), 381 (**1**), 270 (**2**), 272 (**2**), 283 (**2**)
electromagnetic interference (EMI) 193 (**2**)
electromigration 144 (**1**), 153 (**1**), 191 (**1**), 380 (**1**), 79 (**2**), 100 (**2**), 124 (**2**), 127 (**2**), 182 (**2**), 184 (**2**), 204 (**2**), 219 (**2**), 267 (**2**), 536 (**2**), 618 (**2**)
electron back scatter diffraction (EBSD) 149 (**1**)
electron transport characteristics 192 (**1**)
electronic manufacturing service (EMS) 371 (**2**)
electronic speckle pattern interferometry (ESPI) 158 (**1**), 652 (**2**), 655 (**2**), 674 (**2**)
electronics 102 (**1**), 103 (**1**), 283 (**1**)–285 (**1**), 315 (**1**), 351 (**1**), 555 (**1**), 556 (**1**), 629 (**1**), 136 (**2**), 177 (**2**), 234 (**2**), 243 (**2**), 255 (**2**), 270 (**2**), 294 (**2**), 304 (**2**), 311 (**2**), 323 (**2**), 364 (**2**), 366 (**2**), 389 (**2**), 392 (**2**), 397 (**2**), 403 (**2**), 404 (**2**), 432 (**2**), 442 (**2**), 462 (**2**), 487 (**2**), 493 (**2**), 527 (**2**), 529 (**2**), 571 (**2**), 595 (**2**), 606 (**2**), 607 (**2**), 611 (**2**), 613 (**2**), 618 (**2**), 630 (**2**), 656 (**2**), 703 (**2**)
– packaging 4 (**1**), 273 (**1**), 283 (**1**), 284 (**1**), 121 (**2**), 132 (**2**), 149 (**2**), 177 (**2**), 302 (**2**), 489 (**2**), 611 (**2**), 687 (**2**)
electrostatic discharge (ESD) 184 (**2**)
electrostatically driven microengines 324 (**2**)
elongation to failure 156 (**1**)
encapsulated package 139 (**2**)
end-contacted nanotubes 197 (**1**)
energy dispersive spectroscopy (EDS) 350 (**2**)
energy-based fatigue models 405 (**1**), 433 (**1**)

engineering reliability 204 (**2**), *see also* reliability engineering
environmental chamber 463 (**1**)
environmental stress screening (ESS) 190 (**2**), 194 (**2**)
epitaxial film 137 (**1**)
epoxy adhesive 508 (**1**), 443 (**2**), 446 (**2**), 582 (**2**)
epoxy molding compound (EMC) 138 (**2**)
epoxy resin 6 (**1**), 34 (**1**), 44 (**1**), 390 (**2**), 433 (**2**), 600 (**2**), 601 (**2**), 614 (**2**), 616 (**2**), 619 (**2**)
epoxy viscosity 164 (**2**), 169 (**2**)
equation of motion 561 (**1**), 565 (**1**)
equilibrium moduli 17 (**1**)
equilibrium phase diagram 328 (**1**)
Er-doped fiber amplifiers (EDFA) 114 (**1**)
erbium-doped fiber amplifier (EDFA) 489 (**2**)
eutectic solder 294 (**1**), 298 (**1**), 351 (**1**), 354 (**1**), 360 (**1**), 382 (**1**), 397 (**1**), 401 (**1**), 411 (**1**), 412 (**1**), 418 (**1**), 453 (**1**), 491 (**1**), 572 (**2**)
evolution of microstructure 347 (**1**)
excluded area 579 (**2**), 607 (**2**)
excluded volume 579 (**2**)
exposure time 127 (**1**)
Eyring equation 227 (**2**)

fab integrated packaging (FIP) 138 (**2**)
Fabry-Perot filter with DBRs 116 (**1**)
Fabry-Pérot interferences 512 (**2**)
face-centered-cubic (FCC) structure 146 (**1**)
failure 3 (**1**), 136 (**1**), 144 (**1**), 156 (**1**), 157 (**1**), 185 (**1**), 205 (**1**), 215 (**1**), 223 (**1**), 253 (**1**), 262 (**1**), 270 (**1**), 290 (**1**), 293 (**1**), 309 (**1**), 314 (**1**), 317 (**1**), 318 (**1**), 380 (**1**), 386 (**1**), 411 (**1**)–414 (**1**), 431 (**1**)–433 (**1**), 524 (**1**), 573 (**1**), 595 (**1**), 631 (**1**), 3 (**2**), 6 (**2**), 12 (**2**), 47 (**2**), 87 (**2**), 104 (**2**), 123 (**2**), 124 (**2**), 130 (**2**), 177 (**2**), 178 (**2**), 191 (**2**), 203 (**2**), 205 (**2**), 241 (**2**), 255 (**2**), 256 (**2**), 258 (**2**), 262 (**2**), 274 (**2**), 279 (**2**), 282 (**2**), 294 (**2**), 301 (**2**), 315 (**2**), 348 (**2**), 357 (**2**), 394 (**2**), 401 (**2**), 403 (**2**), 546 (**2**), 630 (**2**), 683 (**2**)
– analysis 217 (**1**), 223 (**1**), 321 (**1**), 654 (**1**), 661 (**1**), 89 (**2**), 177 (**2**), 179 (**2**), 250 (**2**), 255 (**2**), 269 (**2**), 294 (**2**), 635 (**2**), 647 (**2**)
– criterion 221 (**1**), 405 (**1**), 573 (**1**), 268 (**2**)
– density function 217 (**1**)
– detection 302 (**1**), 269 (**2**), 635 (**2**)
– diagram 658 (**1**)
– distribution 217 (**1**), 218 (**1**), 220 (**1**), 299 (**1**), 318 (**1**), 319 (**1**), 389 (**1**), 524 (**1**), 572 (**1**), 595 (**1**), 605 (**1**), 655 (**1**), 47 (**2**), 177 (**2**), 268 (**2**), 350 (**2**), 604 (**2**)
– free life 319 (**1**), 342 (**1**)
– in time (FIT) 181 (**2**)
– intensity 220 (**1**), 388 (**1**), 524 (**1**), 551 (**1**), 623 (**1**), 661 (**1**), 180 (**2**), 204 (**2**), 226 (**2**)
– link 454 (**1**), 206 (**2**), 262 (**2**)
– mechanism 136 (**1**), 185 (**1**), 298 (**1**), 314 (**1**), 177 (**2**), 183 (**2**), 184 (**2**), 204 (**2**), 221 (**2**), 242 (**2**), 255 (**2**), 267 (**2**), 269 (**2**), 273 (**2**), 274 (**2**), 313 (**2**), 349 (**2**), 384 (**2**), 401 (**2**), 433 (**2**), 552 (**2**)
– mode 144 (**1**), 152 (**1**), 315 (**1**), 127 (**2**), 178 (**2**), 186 (**2**), 204 (**2**), 214 (**2**), 242 (**2**), 254 (**2**), 262 (**2**), 272 (**2**), 348 (**2**), 357 (**2**), 689 (**2**), 691 (**2**)
– probability 261 (**1**), 388 (**1**), 597 (**1**)
– rate 215 (**1**), 321 (**1**), 574 (**1**), 3 (**2**), 94 (**2**), 177 (**2**), 217 (**2**), 254 (**2**), 547 (**2**), 601 (**2**), 624 (**2**)
– site 153 (**1**), 299 (**1**), 204 (**2**), 262 (**2**)
– software 550 (**1**), 187 (**2**), 213 (**2**), 649 (**2**)
failure-time distribution 255 (**2**)
fall time 124 (**1**)
far infrared Fizeau interferometry (FIFI) 476 (**1**), 514 (**1**)
fast Fourier transform (FFT) 92 (**2**)
fatigue 156 (**1**), 220 (**1**), 222 (**1**), 294 (**1**), 337 (**1**), 390 (**1**), 400 (**1**), 412 (**1**), 431 (**1**)–435 (**1**), 443 (**1**)–448 (**1**), 453 (**1**), 454 (**1**), 595 (**1**), 599 (**1**), 614 (**1**), 620 (**1**)–625 (**1**), 3 (**2**), 9 (**2**), 75 (**2**), 125 (**2**), 132 (**2**), 204 (**2**), 219 (**2**), 221 (**2**), 274 (**2**), 286 (**2**), 303 (**2**), 562 (**2**)
– analysis 222 (**1**)
– – inelastic energy 222 (**1**)
– – steady-state crack growth 222 (**1**)
– – strain based 222 (**1**)
– – stress based 222 (**1**)
– ductility coefficient 432 (**1**), 446 (**1**), 226 (**2**)
– effect 595 (**1**)
– failure 215 (**1**)–217 (**1**), 220 (**1**)–223 (**1**), 300 (**1**), 388 (**1**), 395 (**1**), 411 (**1**)–414 (**1**), 422 (**1**), 454 (**1**), 595 (**1**), 608 (**1**)–613 (**1**), 616 (**1**), 623 (**1**)–625 (**1**), 104 (**2**), 192 (**2**), 204 (**2**), 268 (**2**), 315 (**2**), 562 (**2**)
– – mode 417 (**1**)
– fracture 342 (**1**), 396 (**1**), 412 (**1**)
– – mode 419 (**1**)

– life 213 (**1**), 223 (**1**), 240 (**1**), 294 (**1**), 395 (**1**), 413 (**1**), 415 (**1**), 424 (**1**), 431 (**1**), 451 (**1**), 292 (**2**)
– – model 390 (**1**), 398 (**1**), 402 (**1**), 405 (**1**)
– – – thermal 402 (**1**)
– reliability 398 (**1**), 401 (**1**), 413 (**1**), 424 (**1**)
– resistance 156 (**1**), 75 (**2**)
– strength 363 (**1**), 414 (**1**), 426 (**1**), 9 (**2**)
– test program 435 (**1**)
– tests 242 (**1**), 388 (**1**)–390 (**1**), 396 (**1**), 402 (**1**), 411 (**1**), 416 (**1**), 431 (**1**), 445 (**1**), 454 (**1**), 610 (**1**)–612 (**1**), 215 (**2**), 262 (**2**)
ferrules 272 (**1**)
fiber 11 (**1**), 65 (**1**)–71 (**1**), 81 (**1**), 102 (**1**), 112 (**1**), 119 (**1**), 133 (**1**), 150 (**1**), 170 (**1**), 182 (**1**), 194 (**1**), 269 (**1**), 272 (**1**), 274 (**1**), 277 (**1**), 380 (**1**), 389 (**1**), 459 (**1**), 468 (**1**), 483 (**1**), 502 (**1**), 557 (**1**), 571 (**1**), 577 (**1**), 580 (**1**), 585 (**1**), 592 (**1**), 595 (**1**), 605 (**1**), 23 (**2**), 109 (**2**), 152 (**2**), 158 (**2**), 164 (**2**), 166 (**2**), 185 (**2**), 186 (**2**), 332 (**2**), 367 (**2**), 368 (**2**), 395 (**2**), 477 (**2**), 487 (**2**), 488 (**2**), 497 (**2**), 520 (**2**), 523 (**2**)
– Bragg grating (FBG) 67 (**1**), 70 (**1**), 113 (**1**), 126 (**1**)
– lifetime 597 (**1**), 613 (**1**), 614 (**1**)
– – model 595 (**1**)
– optics 111 (**1**)–114 (**1**), 272 (**1**), 557 (**1**), 332 (**2**), 394 (**2**)
– – directional coupler 330 (**2**)
– strain 11 (**1**), 97 (**1**), 269 (**1**)–271 (**1**), 389 (**1**), 483 (**1**), 605 (**1**), 667 (**1**), 688 (**1**), 692 (**1**)–695 (**1**), 395 (**2**), 511 (**2**)
– strength 277 (**1**), 571 (**1**), 578 (**1**)
fiber-on-board (FOB) 368 (**2**)
fiber-optic system 23 (**2**), 27 (**2**), 37 (**2**), 67 (**2**)
fiber-optics structural mechanics (FOSM) 270 (**1**), 277 (**1**), 65 (**2**)
field failures 385 (**1**), 205 (**2**), 207 (**2**), 254 (**2**), 256 (**2**), 267 (**2**)
filler 25 (**1**), 44 (**1**), 48 (**1**), 347 (**1**), 349 (**1**), 385 (**2**), 392 (**2**), 435 (**2**), 478 (**2**), 527 (**2**), 529 (**2**)–531 (**2**), 536 (**2**), 549 (**2**), 557 (**2**), 565 (**2**), 572 (**2**), 574 (**2**), 578 (**2**), 579 (**2**), 582 (**2**), 583 (**2**), 591 (**2**), 612 (**2**)–614 (**2**), 619 (**2**), 625 (**2**)
filter 114 (**1**), 117 (**1**), 130 (**1**), 134 (**1**), 343 (**2**), 423 (**2**), 427 (**2**), 487 (**2**), 655 (**2**)
fine pitch bonding 105 (**2**)
fine pitch surface mount components 285 (**1**)
fine pitch wirebonding 107 (**2**)
fine tuning 275 (**1**)
finite difference method (FDM) 208 (**1**), 324 (**2**)
finite element analysis (FEA) 87 (**1**), 295 (**1**), 400 (**1**), 432 (**1**), 433 (**1**), 448 (**1**), 630 (**1**), 4 (**2**), 5 (**2**), 62 (**2**), 416 (**2**), 438 (**2**)
– modeling 448 (**1**)
– thermal model 80 (**1**)
finite element method (FEM) 10 (**1**), 35 (**1**), 163 (**1**), 206 (**1**), 208 (**1**), 209 (**1**), 294 (**1**), 321 (**1**), 322 (**1**), 508 (**1**), 524 (**1**), 25 (**2**), 324 (**2**), 681 (**2**), 683 (**2**), 691 (**2**)
– method 209 (**1**)
– model 243 (**1**), 263 (**1**), 508 (**1**), 681 (**2**)
– packages 10 (**1**), 50 (**1**)
first and second order second moment methods 44 (**2**)
first-time-to-failure (FTTF) 456 (**1**)
fixturing 397 (**2**)
flat-topped tunable filter 114 (**1**)
flex circuit 8 (**2**)
flexibilizers 392 (**2**)
flip chip (FC) 6 (**1**), 62 (**1**), 141 (**1**), 283 (**1**), 286 (**1**), 291 (**1**), 310 (**1**), 400 (**1**), 411 (**1**), 450 (**1**), 501 (**1**)–504 (**1**), 517 (**1**), 8 (**2**), 74 (**2**), 79 (**2**), 80 (**2**), 161 (**2**), 261 (**2**), 304 (**2**), 492 (**2**), 502 (**2**), 527 (**2**), 546 (**2**), 611 (**2**)
– assembly 74 (**2**), 304 (**2**), 320 (**2**), 652 (**2**), 697 (**2**)
– attachment 286 (**1**), 288 (**1**), 72 (**2**)
– BGA (FCBGA) 283 (**1**), 286 (**1**), 652 (**2**)
– components 313 (**1**)
– failures 6 (**1**)
– gold bump 235 (**2**)
– interconnection 283 (**1**), 310 (**1**), 313 (**1**), 294 (**2**)
– module 552 (**2**), 689 (**2**)
– on board (FCOB) 136 (**2**)
– package 284 (**1**), 286 (**1**)–288 (**1**), 291 (**1**), 310 (**1**), 501 (**1**), 519 (**1**), 145 (**2**), 651 (**2**), 652 (**2**), 656 (**2**), 659 (**2**), 660 (**2**), 662 (**2**), 668 (**2**), 670 (**2**), 672 (**2**), 674 (**2**), 685 (**2**), 688 (**2**)
– plastic ball grid array (FC-PBGA) 481 (**1**), 493 (**1**), 685 (**2**)
– – package assembly 516 (**1**)
– reflow soldering 77 (**2**)
– underfill inspection 282 (**2**)
flow soldering 412 (**1**)
flow stress 165 (**1**), 344 (**1**), 418 (**1**)
flux chemical reactions 360 (**1**)
focused ion beam imaging (FIB) 148 (**1**)

four-beam moiré interferometry 476 (**1**)
four-wave mixing (FWM) 112 (**1**)
fractional factorial experiments (FFEs) 232 (**1**)
fracture 5 (**1**), 156 (**1**), 165 (**1**), 220 (**1**), 325 (**1**), 342 (**1**), 347 (**1**), 412 (**1**), 419 (**1**), 523 (**1**)–525 (**1**), 529 (**1**), 3 (**2**), 262 (**2**), 284 (**2**), 303 (**2**), 395 (**2**), 506 (**2**), 600 (**2**)
– analysis 221 (**1**), 222 (**1**), 347 (**1**), 523 (**1**), 529 (**1**), 5 (**2**), 184 (**2**), 247 (**2**), 452 (**2**)
– mechanics approach 524 (**1**)
– toughness testing 169 (**1**)
free space optical interconnect (FSOI) 366 (**2**)
free vibrations 690 (**1**)
frequency response 646 (**1**), 648 (**1**)
frequency scanning 30 (**1**)
frequency-modified Coffin-Manson model 447 (**1**)
frequency-modified Morrow model 447 (**1**)
Fresnel reflection 373 (**2**), 512 (**2**)
fretting corrosion 287 (**2**)
fringe patterns 469 (**1**), 485 (**1**), 488 (**1**), 498 (**1**), 334 (**2**), 337 (**2**), 654 (**2**)
fringe-locus function 331 (**2**)
full array configuration 289 (**1**)
full factorial design 229 (**1**)
full width at half maximum (FWHM) 71 (**1**), 87 (**1**), 350 (**2**), 357 (**2**)
fully cure dependent modeling 50 (**1**), 51 (**1**)
fully cured polymer 18 (**1**)
fully state dependent modeling 10 (**1**), 34 (**1**), 49 (**1**)
function variable incidence matrix 30 (**2**)
function-rich 253 (**2**)
functionality 3 (**1**), 53 (**1**)–57 (**1**), 132 (**1**), 244 (**1**), 252 (**1**), 262 (**1**), 288 (**1**), 629 (**1**), 253 (**2**), 336 (**2**), 348 (**2**), 391 (**2**), 424 (**2**), 487 (**2**), 496 (**2**), 506 (**2**), 555 (**2**)
fundamental moisture absorption 435 (**2**)
fused biconical taper (FBT) 272 (**1**), 16 (**2**)
– couplers 686 (**1**), 16 (**2**)

G-Helix 139 (**2**)
GaAs 112 (**1**), 139 (**1**), 9 (**2**), 109 (**2**), 131 (**2**), 350 (**2**), 357 (**2**), 363 (**2**)
– laser 493 (**2**)
galvanic corrosion 264 (**2**), 287 (**2**), 557 (**2**), 575 (**2**), 595 (**2**), 607 (**2**), 615 (**2**), 616 (**2**), 618 (**2**)
garbage in garbage out (GIGO) 208 (**1**)
Gatan image filter (GIF) 343 (**2**)
Gaussian beam 26 (**2**), 491 (**2**)
generalized Galerkin procedure 630 (**1**), 673 (**1**), 677 (**1**)
geometrical nonlinearity 668 (**1**)
glass FBG 82 (**1**)
glass fiber 65 (**1**), 380 (**1**), 10 (**2**), 392 (**2**), 477 (**2**), 491 (**2**), 559 (**2**), 562 (**2**)
glass optical fiber 68 (**1**), 81 (**1**), 10 (**2**)
glass solder 496 (**2**)
glass transition temperature 4 (**1**), 460 (**1**), 468 (**1**), 530 (**1**), 7 (**2**), 85 (**2**), 109 (**2**), 128 (**2**), 143 (**2**), 385 (**2**), 488 (**2**), 510 (**2**), 547 (**2**), 559 (**2**), 571 (**2**), 620 (**2**), 662 (**2**), 664 (**2**)
glassy modulus 17 (**1**)
glob-top manner 152 (**2**)
global mismatch 10 (**2**)
global stress 299 (**1**), 10 (**2**)
global-local modeling techniques 450 (**1**)
gold electroplating 103 (**2**)
gold thermosonic ball bond 113 (**2**)
gold-tin eutectics 9 (**2**)
Graham-Walles equation 228 (**2**)
grain 142 (**1**), 147 (**1**)–149 (**1**), 157 (**1**), 175 (**1**), 191 (**1**), 212 (**1**), 261 (**1**), 315 (**1**), 325 (**1**), 335 (**1**), 87 (**2**), 103 (**2**), 241 (**2**), 274 (**2**), 352 (**2**), 500 (**2**)
– boundary (GB) 275 (**2**), 352 (**2**)
– boundary sliding 334 (**1**), 274 (**2**)
grain growth 307 (**1**), 335 (**1**)
graph conversion 31 (**2**)
graph partitioning 34 (**2**)
graph partitioning algorithms 34 (**2**)
graph partitioning analysis 37 (**2**)
graphene 182 (**1**), *see also* graphitic carbon
graphitic carbon 181 (**1**)
graphoepitaxy 137 (**1**)
grating 65 (**1**), 80 (**1**), 115 (**1**), 461 (**1**), 476 (**1**), 510 (**1**), 384 (**2**), 654 (**2**)
– length 70 (**1**), 120 (**1**)
– period effect 81 (**1**)
– pitch 462 (**1**)
gravimetric method 146 (**1**)
group delay (GD) 127 (**1**)
growth 182 (**1**)–184 (**1**), 186 (**1**)–189 (**1**), 197 (**1**), 222 (**1**)–224 (**1**), 335 (**1**), 363 (**1**)–376 (**1**), 385 (**1**), 413 (**1**), 595 (**1**)–601 (**1**), 621 (**1**)–625 (**1**), 8 (**2**), 13 (**2**), 83 (**2**), 186 (**2**), 226 (**2**), 293 (**2**), 317 (**2**), 348 (**2**), 354 (**2**), 460 (**2**), 618 (**2**)
– kinetics 365 (**1**)
– of the intermetallic compound 413 (**1**)

– temperature 197 (**1**)
– time 187 (**1**)

H-PDLC 123 (**1**)
Hall-Petch rule 172 (**1**)
Hamilton principle 671 (**1**)
hard solder 9 (**2**)
hard-core model 578 (**2**)
hardness 159 (**1**), 163 (**1**), 172 (**1**), 175 (**1**), 335 (**1**)
harmonic forcing 651 (**1**)
heat generation 77 (**1**)
heat loss 109 (**1**)
heat transfer 79 (**1**), 87 (**1**), 209 (**1**), 525 (**1**)
heavily alloyed gold 102 (**2**)
heteroepitaxy 137 (**1**)
high cycle fatigue 222 (**1**)
high density interconnect (HDI) 73 (**2**), 146 (**2**), 271 (**2**), 291 (**2**)
high performance dispersion compensators 126 (**1**)
high-cycle fatigue 454 (**1**)
high-speed testing 606 (**1**)
high-temperature-storage (HTS) 485 (**1**)
highly accelerated life test (HALT) 454 (**1**), 4 (**2**), 194 (**2**), 217 (**2**), 218 (**2**)
highly accelerated stress screening (HASS) 194 (**2**)
hinged micromirror 325 (**2**), 327 (**2**)
holographic grating 118 (**1**)
holographic memory assemblies 11 (**2**)
holographic polymer dispersed liquid crystal (H-PDLC) 113 (**1**)
homoepitaxy 139 (**1**)
hot air solder leveling (HASL) 297 (**1**), 302 (**1**), 310 (**1**), 381 (**1**), 272 (**2**)
Hsai-Wu failure criteria 654 (**1**)
humidity 155 (**1**), 262 (**1**), 275 (**1**), 388 (**1**), 500 (**1**), 523 (**1**), 572 (**1**), 576 (**1**), 587 (**1**), 595 (**1**), 606 (**1**), 178 (**2**), 212 (**2**), 220 (**2**), 254 (**2**), 255 (**2**), 303 (**2**), 318 (**2**), 386 (**2**), 393 (**2**), 396 (**2**), 398 (**2**), 435 (**2**), 444 (**2**), 453 (**2**), 556 (**2**), 574 (**2**), 595 (**2**), 601 (**2**), 615 (**2**), 620 (**2**), 623 (**2**)
– chamber 435 (**2**)
hybrid integration 304 (**2**)
hydrophilic 586 (**1**), 587 (**1**), 458 (**2**)
hydrophilic nanoparticle 586 (**1**)
hydrophobic 274 (**1**), 587 (**1**), 318 (**2**), 458 (**2**), 477 (**2**), 618 (**2**)
hydrophobic nanoparticle 586 (**1**)
hydrophobicity 273 (**1**), 458 (**2**), 477 (**2**)
hygroscopic deformation 494 (**1**)
hygroscopic strain 495 (**1**), 499 (**1**)
hygroscopic stress 496 (**1**)
hygroscopic swelling 500 (**1**)
– properties 494 (**1**)
hygrostrain 531 (**1**)
hygrostress 525 (**1**)
hyperbolic-sine model 442 (**1**)
hypercorrosion 286 (**2**)
hypergraph 33 (**2**)

imaging lens 516 (**1**)
IMC growth 364 (**1**)
IMC layer growth 366 (**1**)
IMC phase 274 (**2**), 275 (**2**)
immersion interfeometer 478 (**1**)
In composition fluctuation 343 (**2**), 344 (**2**), 347 (**2**)
in-circuit test (ICT) 380 (**1**)
in-plane thermal field 640 (**1**)
indentation 146 (**1**), 162 (**1**), 385 (**1**), 603 (**1**)
indentation-induced delamination test 404 (**2**)
independent plastic strain 430 (**1**)
indium tin oxide (ITO) 403 (**2**), 419 (**2**)
inelastic energy approach 224 (**1**)
inelastic strain 222 (**1**), 491 (**1**), 9 (**2**)
InGaAs 140 (**1**)
initial curvature 558 (**1**), 15 (**2**)
initial spring constant 561 (**1**)
initial stage of intermetallic growth 372 (**1**)
initial strength 575 (**1**), 597 (**1**), 599 (**1**)
initial stress 7 (**1**), 8 (**1**), 392 (**1**), 706 (**2**)
initial wetting time 361 (**1**), 363 (**1**)
ink-jet dispensing 517 (**2**)
inner lead bonding (ILB) 75 (**2**)
InP 112 (**1**)
insertion loss 120 (**1**), 372 (**2**)
integrated circuit (IC) 155 (**1**), 252 (**1**), 263 (**1**), 264 (**1**), 429 (**1**), 450 (**1**), 486 (**1**), 13 (**2**), 71 (**2**), 79 (**2**), 109 (**2**), 113 (**2**), 135 (**2**), 182 (**2**), 186 (**2**), 190 (**2**), 318 (**2**), 320 (**2**), 366 (**2**), 380 (**2**), 621 (**2**)
inter-particle conduction 540 (**2**)
interaction integral method 536 (**1**)
interconnect failure 273 (**2**)
interconnection 112 (**1**), 210 (**1**), 215 (**1**), 283 (**1**), 313 (**1**), 315 (**1**), 317 (**1**), 323 (**1**), 327 (**1**), 329 (**1**), 337 (**1**), 339 (**1**), 342 (**1**), 351 (**1**), 389 (**1**), 390 (**1**), 405 (**1**), 459 (**1**), 491 (**1**), 273 (**2**)
– electronics industries (IPC) 283 (**1**)
– level model 322 (**1**)

interdiffusion 81 (**2**), 103 (**2**), 114 (**2**), 219 (**2**), 508 (**2**)
interface failure mode 413 (**1**)
interface fatigue crack 426 (**1**)
interface fatigue mode 412 (**1**)
interface fracture mode 417 (**1**)
interface fracture toughness 529 (**1**)
interface metallurgy 76 (**2**), 349 (**2**), 357 (**2**)
interface strengths 262 (**1**)
interfacial adhesion 263 (**1**), 318 (**2**), 403 (**2**), 433 (**2**), 441 (**2**), 444 (**2**), 449 (**2**)
interfacial bond 271 (**2**)
interfacial compliance 7 (**2**), 12 (**2**), 59 (**2**), 473 (**2**)–475 (**2**), 479 (**2**), 481 (**2**), 483 (**2**)
interfacial conductivity 593 (**2**)
interfacial contact materials 196 (**1**)
interfacial failure 262 (**1**), 276 (**2**), 432 (**2**), 469 (**2**)
interfacial fracture 527 (**1**), 285 (**2**), 404 (**2**), 437 (**2**)–439 (**2**), 449 (**2**), 451 (**2**), 452 (**2**), 456 (**2**), 461 (**2**), 468 (**2**)
– toughness 461 (**2**), 462 (**2**), 469 (**2**)
interfacial fracture mechanics 527 (**1**)
interfacial hydrophobicity 457 (**2**), 458 (**2**)
interfacial metallurgical reactions 357 (**1**)
interfacial reactions 347 (**1**), 360 (**1**)
interfacial shearing stress 474 (**2**), 482 (**2**), 485 (**2**), 495 (**2**)
interfacial strength 263 (**1**), 260 (**2**), 263 (**2**), 403 (**2**), 411 (**2**), 414 (**2**), 423 (**2**), 425 (**2**), 426 (**2**), 469 (**2**)
interfacial stress 7 (**2**), 401 (**2**), 473 (**2**), 481 (**2**), 483 (**2**), 652 (**2**)
interfacial tension coefficient 485 (**2**)
interferometer (IT) 331 (**2**), 692 (**2**)
intermetallic bond 383 (**1**), 271 (**2**)
intermetallic compound 267 (**2**)
intermetallic compound (IMC) 330 (**1**), 412 (**1**)
– layer 417 (**1**)
intermetallic growth 262 (**1**), 363 (**1**), 83 (**2**), 93 (**2**)
– kinetics 363 (**1**), 376 (**1**)
intermetallic layer 330 (**1**), 339 (**1**), 360 (**1**), 365 (**1**), 375 (**1**), 415 (**1**)
internal compressive forces 476 (**2**)
interposer 294 (**1**)–296 (**1**), 305 (**1**), 309 (**1**), 519 (**1**), 81 (**2**), 278 (**2**), 371 (**2**)
intra-particle resistance 542 (**2**)
intrinsic stress 146 (**1**), 243 (**2**)
inverse piezoelectric effect 668 (**1**)
Ishikawa diagram 219 (**1**)
isothermal mechanical fatigue test 413 (**1**)
isotropic conductive adhesive (ICA) 527 (**2**), 529 (**2**), 538 (**2**), 572 (**2**), 575 (**2**), 595 (**2**), 604 (**2**), 613 (**2**)
isotropic laminate 631 (**1**)
isotropic materials 14 (**1**)
isotropic PFBG 80 (**1**)

J-integral 542 (**1**), 416 (**2**), 419 (**2**)

Kelvin element 6 (**1**), 52 (**2**)
kernel function 11 (**1**)
kickstand 310 (**2**), 313 (**2**)
$KNbO_3$ 112 (**1**)
Knoop hardness scale 97 (**2**)
– micro-hardness 354 (**2**)
known good die (KGD) 80 (**2**), 136 (**2**), 145 (**2**)
Kriging error estimation 245 (**1**)

laminated microstructure 660 (**1**)
land grid array (LGA) 138 (**2**), 260 (**2**)
large displacement 564 (**1**)
laser 67 (**1**), 71 (**1**), 85 (**1**), 94 (**1**)–101 (**1**), 118 (**1**), 128 (**1**), 139 (**1**), 147 (**1**), 186 (**1**), 273 (**1**), 298 (**1**), 483 (**1**), 514 (**1**), 23 (**2**), 73 (**2**), 144 (**2**), 216 (**2**), 221 (**2**), 229 (**2**), 234 (**2**), 271 (**2**), 272 (**2**), 314 (**2**), 330 (**2**), 332 (**2**), 477 (**2**), 487 (**2**), 489 (**2**), 518 (**2**), 523 (**2**), 573 (**2**), 655 (**2**), 679 (**2**)
– ablation 186 (**1**)
– bars 357 (**2**)
– beam 70 (**1**), 119 (**1**), 147 (**1**), 476 (**1**), 485 (**1**), 514 (**1**)–516 (**1**), 23 (**2**), 32 (**2**), 40 (**2**), 91 (**2**), 489 (**2**), 491 (**2**), 499 (**2**), 520 (**2**), 655 (**2**)
– DFB 67 (**1**), 363 (**2**)
– diode 71 (**1**), 85 (**1**), 95 (**1**)–101 (**1**), 25 (**2**), 222 (**2**), 330 (**2**), 343 (**2**), 348 (**2**), 354 (**2**), 362 (**2**), 489 (**2**)
– – high power 348 (**2**), 357 (**2**)
– emission 351 (**2**)
– Excimer 313 (**2**)
– Fabry-Perot (FP) 363 (**2**)
– interferometer 91 (**2**), 110 (**2**), 234 (**2**), 330 (**2**), 704 (**2**)
– interferometry 476 (**1**), 514 (**1**), 683 (**2**)
– power 70 (**1**), 86 (**1**), 93 (**1**), 97 (**1**)–99 (**1**), 101 (**1**), 119 (**1**), 485 (**1**), 28 (**2**), 514 (**2**)
– pump 487 (**2**)
– ridge 357 (**2**)
– welding 483 (**1**), 488 (**2**)
latin hypercube design 235 (**1**)

lattice-mismatch strain 8 (**2**)
lead-free flux 380 (**1**)
lead-free solder 283 (**1**), 324 (**1**), 353 (**1**), 411 (**1**), 435 (**1**), 557 (**1**), 261 (**2**), 273 (**2**), 603 (**2**), 612 (**2**)
– alloy compositions 378 (**1**)
– joint 424 (**1**)
– – interconnections 351 (**1**)
lead-free soldering processes 330 (**1**), 377 (**1**), 381 (**1**)
life distribution model 455 (**1**)
lifetime 6 (**1**), 165 (**1**), 218 (**1**), 269 (**1**), 273 (**1**), 319 (**1**), 586 (**1**), 587 (**1**), 595 (**1**), 8 (**2**), 75 (**2**), 94 (**2**), 122 (**2**), 195 (**2**), 208 (**2**), 214 (**2**), 216 (**2**), 221 (**2**), 225 (**2**), 226 (**2**), 229 (**2**), 315 (**2**), 477 (**2**), 574 (**2**)
– failure fraction 194 (**2**)
light emitting diode (LED) 6 (**1**), 71 (**1**), 216 (**2**), 343 (**2**), 489 (**2**)
lightly alloyed gold 102 (**2**)
linear acceleration 328 (**2**)
linear elastic fracture mechanics (LEFM) 220 (**1**), 248 (**2**)
linear state dependent viscoelasticity 13 (**1**)
linear thermally-dependent contributions 673 (**1**)
linear visco-elastic modeling 4 (**1**), 18 (**1**)
liquid crystal display (LCD) 225 (**2**), 258 (**2**), 527 (**2**), 572 (**2**)
liquid crystal (LC) 113 (**1**), 118 (**1**), 186 (**2**)
– polymer (LCP) 302 (**2**)
liquid encapsulant 6 (**1**)
liquid Pb-free solders 376 (**1**)
liquid solder 330 (**1**), 332 (**1**)
load 7 (**1**), 17 (**1**), 42 (**1**), 158 (**1**), 163 (**1**), 170 (**1**), 208 (**1**), 222 (**1**), 270 (**1**), 302 (**1**), 388 (**1**), 400 (**1**), 414 (**1**), 419 (**1**), 435 (**1**), 454 (**1**), 460 (**1**), 475 (**1**), 476 (**1**), 508 (**1**), 532 (**1**), 556 (**1**), 571 (**1**), 575 (**1**), 584 (**1**), 600 (**1**), 604 (**1**), 629 (**1**), 640 (**1**), 661 (**1**), 684 (**1**), 696 (**1**), 32 (**2**), 48 (**2**), 124 (**2**), 204 (**2**), 212 (**2**), 235 (**2**), 259 (**2**), 261 (**2**), 331 (**2**), 334 (**2**), 349 (**2**), 354 (**2**), 400 (**2**), 404 (**2**), 405 (**2**), 407 (**2**), 410 (**2**), 422 (**2**), 469 (**2**), 506 (**2**), 522 (**2**), 550 (**2**), 562 (**2**), 572 (**2**), 586 (**2**), 686 (**2**)
– application time 7 (**1**)
– sensor 616 (**1**)
loading conditions 17 (**1**), 206 (**1**), 215 (**1**), 270 (**1**), 314 (**1**), 398 (**1**), 451 (**2**)
local area networks (LAN) 363 (**2**)
local mismatch 10 (**2**)
local stress 10 (**2**), 241 (**2**), 685 (**2**)
localization length 184 (**1**)
localized galvanic corrosion 287 (**2**)
lognormal failure rate distribution 184 (**2**)
long-range thickness variation (LR-TV) 347 (**2**)
long-term reliability 272 (**1**), 379 (**1**), 483 (**1**), 6 (**2**), 172 (**2**), 182 (**2**), 186 (**2**), 209 (**2**), 215 (**2**), 223 (**2**), 432 (**2**), 450 (**2**), 595 (**2**), 604 (**2**), 607 (**2**), 698 (**2**), 703 (**2**)
longitudinal relaxation modulus 27 (**1**)
longitudinal strain 26 (**1**), 323 (**1**)
low cycle fatigue models 443 (**1**)
low-cycle fatigue 222 (**1**), 454 (**1**)
low-pressure chemical vapor deposition (LPCVD) 325 (**2**), 704 (**2**)

macrocrack 275 (**2**)
maintainability 206 (**2**)
maintenance 211 (**2**), 400 (**2**)
– preventive 211 (**2**)
– reactive 211 (**2**)
mandrel device 616 (**1**)
Manson-Coffin's law 412 (**1**)
manufacturability 144 (**1**), 285 (**1**)
manufacturing-qualification AST 193 (**2**)
mask 67 (**1**), 68 (**1**), 85 (**1**), 128 (**1**)–130 (**1**), 306 (**1**), 311 (**1**), 74 (**2**), 152 (**2**), 155 (**2**), 658 (**2**), 659 (**2**), 678 (**2**)
master curve 23 (**1**), 26 (**1**), 48 (**1**), 57 (**1**)
mathematical modeling 400 (**1**)
matrix creep fatigue model 432 (**1**)
maximum acceleration 273 (**1**), 562 (**1**)
maximum acceleration (deceleration) 566 (**1**)
maximum compressive force 561 (**1**), 565 (**1**)
maximum displacement 494 (**1**), 561 (**1**)
maximum stress criteria 221 (**1**), 524 (**1**), 654 (**1**), *see also* normal stress
maximum stress failure criteria 657 (**1**)
maximum wetting force 361 (**1**)
Maxwell element 6 (**1**)
mean downtime (MDT) 181 (**2**)
mean operating time (MOT) 181 (**2**)
mean-time-between failures (MTBF) 181 (**2**), 562 (**2**)
mean-time-to-failure (MTTF) 456 (**1**), 181 (**2**), 212 (**2**), 226 (**2**)
mechanical failure 144 (**1**), 215 (**1**), 314 (**1**), 524 (**1**), 551 (**1**), 105 (**2**), 206 (**2**), 315 (**2**)
mechanical fatigue test 414 (**1**), 416 (**1**)
mechanical interlocking 452 (**2**), 462 (**2**), 488 (**2**), 508 (**2**)

mechanical load 388 (**1**), 400 (**1**), 454 (**1**), 508 (**1**), 627 (**1**), 631 (**1**), 633 (**1**)–635 (**1**), 640 (**1**), 657 (**1**), 254 (**2**), 456 (**2**), 501 (**2**)
mechanical loading 215 (**1**), 269 (**1**), 406 (**1**), 233 (**2**), 272 (**2**), 473 (**2**)
mechanical reliability 274 (**1**), 411 (**1**), 523 (**1**), 571 (**1**), 595 (**1**), 613 (**1**), 141 (**2**), 270 (**2**), 293 (**2**), 373 (**2**), 478 (**2**), 542 (**2**), *see also* structural reliability
mechanical shearing 390 (**1**)
mechanical shock 283 (**1**), 305 (**1**), 315 (**1**), 317 (**1**), 341 (**1**), 215 (**2**), 280 (**2**), 497 (**2**)
mechanical shock test 315 (**1**), 317 (**1**)
mechanical simulation 49 (**1**), 322 (**1**)
mechanical strength 141 (**1**), 198 (**1**), 275 (**1**), 357 (**1**), 388 (**1**), 191 (**2**), 443 (**2**), 554 (**2**), 578 (**2**)
median strength 607 (**1**), 608 (**1**)
melting point 5 (**1**), 145 (**1**), 342 (**1**), 352 (**1**), 367 (**1**), 411 (**1**), 75 (**2**), 97 (**2**), 123 (**2**), 127 (**2**), 311 (**2**), 615 (**2**)
melting temperature 5 (**1**), 149 (**1**), 212 (**1**), 298 (**1**), 351 (**1**), 383 (**1**), 430 (**1**), 75 (**2**), 498 (**2**), 612 (**2**)
membrane deflection 159 (**1**), 160 (**1**), 225 (**1**), 706 (**2**)
membrane deformation 678 (**2**), 681 (**2**), 683 (**2**)
mesh density 211 (**1**)
metal film 137 (**1**), 141 (**1**), 479 (**1**)
metal-oxide-silicon (MOS) transistor 140 (**1**)
metalization systems 331 (**1**)
metalized fiber 269 (**1**), 16 (**2**)
metallurgical interface 122 (**2**), 129 (**2**)
metallurgical structures 406 (**1**)
metastable solubility 331 (**1**)
metropolitan access networks (MAN) 362 (**2**)
micro-optics 310 (**2**)
micro-optoelectromechanical systems (MOEMS) 507 (**1**), 323 (**2**), 338 (**2**), 380 (**2**)
micro-Raman spectroscopy 350 (**2**)
micro-via technology 272 (**2**)
microbending 272 (**1**), 684 (**1**), 3 (**2**), 15 (**2**)
microcrack 307 (**1**), 579 (**1**), 580 (**1**), 592 (**1**), 110 (**2**), 185 (**2**), 275 (**2**), 507 (**2**)
– growth 580 (**1**), 186 (**2**)
microDAC 234 (**2**)
microelectromechanical systems (MEMS) 135 (**1**), 142 (**1**), 262 (**1**), 273 (**1**), 627 (**1**), 4 (**2**), 10 (**2**), 105 (**2**), 241 (**2**), 299 (**2**), 310 (**2**), 323 (**2**), 477 (**2**)
– device 301 (**2**), 311 (**2**)
– failures 303 (**2**)
– packaging 299 (**2**), 300 (**2**), 302 (**2**)
microelectronics 3 (**1**), 135 (**1**), 136 (**1**), 155 (**1**), 193 (**1**), 200 (**1**), 205 (**1**), 217 (**1**), 253 (**1**), 259 (**1**), 272 (**1**), 273 (**1**), 460 (**1**), 475 (**1**), 556 (**1**), 629 (**1**), 3 (**2**), 8 (**2**), 14 (**2**), 16 (**2**), 81 (**2**), 87 (**2**), 101 (**2**), 111 (**2**), 203 (**2**), 204 (**2**), 230 (**2**), 244 (**2**), 299 (**2**), 310 (**2**), 383 (**2**), 403 (**2**), 432 (**2**), 442 (**2**), 473 (**2**), 480 (**2**), 488 (**2**), 563 (**2**), 573 (**2**), 698 (**2**)
– bonding wire 95 (**2**)
microengine 328 (**2**), 334 (**2**), 335 (**2**)
micromachining 157 (**1**), 324 (**2**)
micromechanical model 572 (**1**)
micromirror 142 (**1**), 507 (**1**), 299 (**2**), 324 (**2**), 337 (**2**), 338 (**2**)
microscopic moiré interferometry 475 (**1**), 477 (**1**), 502 (**1**)
microspheres 479 (**2**)
microspring contact on silicon technology (MOST) 139 (**2**)
microstrain 461 (**1**), 670 (**1**)
microstructural characterization 321 (**1**)
microstructural effect 138 (**1**)
microstructure 137 (**1**), 173 (**1**), 212 (**1**), 261 (**1**), 313 (**1**), 334 (**1**), 351 (**1**), 273 (**2**), 275 (**2**)
microsystem 205 (**1**), 212 (**1**), 262 (**1**), 324 (**2**)
microtensile testing (MT) 157 (**1**)
mirror 65 (**1**), 291 (**1**), 462 (**1**), 483 (**1**), 506 (**1**), 302 (**2**), 363 (**2**), 370 (**2**), 452 (**2**)
– holders 463 (**1**)
– integrated 371 (**2**)
Mises equivalent stresses 419 (**1**)
mode field diameter 26 (**2**), 518 (**2**)
mode mixity phase angle 528 (**1**), 529 (**1**), 549 (**1**)
model molding compounds 44 (**1**)
modified crack surface displacement extrapolation method (MCSDEM) 531 (**1**)
modified J-integral (MJI) method 532 (**1**), 549 (**1**)
modified passive alignment method 162 (**2**)
modified virtual crack closure method (MVCCM) 533 (**1**), 550 (**1**)
modulator 330 (**2**), 333 (**2**), 487 (**2**)
moiré field 461 (**1**)
moiré grating 461 (**1**)

moiré interferometry 460 (**1**), 470 (**1**), 475 (**1**), 476 (**1**), 484 (**1**), 486 (**1**), 263 (**2**), 652 (**2**)
moiré pattern 461 (**1**), 463 (**1**), 469 (**1**), 477 (**1**)
moisture 5 (**1**), 169 (**1**), 262 (**1**), 269 (**1**), 508 (**1**), 525 (**1**), 571 (**1**), 595 (**1**), 13 (**2**), 16 (**2**), 82 (**2**), 186 (**2**), 267 (**2**), 294 (**2**), 303 (**2**), 386 (**2**), 392 (**2**), 395 (**2**), 396 (**2**), 401 (**2**), 432 (**2**), 433 (**2**), 450 (**2**), 464 (**2**), 503 (**2**), 515 (**2**), 555 (**2**), 642 (**2**)
– absorption 65 (**1**), 525 (**1**), 185 (**2**), 396 (**2**), 432 (**2**), 442 (**2**), 514 (**2**)
– concentration distribution 440 (**2**)
– diffusion 262 (**1**), 525 (**1**), 544 (**1**), 436 (**2**), 536 (**2**)
– diffusion coefficient 433 (**2**)
– ingress 6 (**1**)
– preconditioning 62 (**1**), 523 (**1**)–526 (**1**), 542 (**1**), 444 (**2**), 452 (**2**), 460 (**2**)
– sensitivity 386 (**1**), 523 (**1**)
– – level (MSL) 5 (**1**), 386 (**1**)
– swelling coefficient 456 (**2**)
mold compound 494 (**1**)
molding compound 5 (**1**), 252 (**1**), 524 (**1**), 8 (**2**), 13 (**2**), 14 (**2**), 432 (**2**)
molding plastic 389 (**1**)
molecular dynamic 172 (**1**)
Morrow's energy-based model 446 (**1**)
multi-beam optical stress sensor (MOSS) 385 (**1**)
multi-channel dispersion-slope compensator 126 (**1**)
multi-mode fiber (MMF) 27 (**2**), 167 (**2**), 365 (**2**)
multi-objective optimization 263 (**1**), 42 (**2**)
multi-points approach 11 (**1**)
multi-stage peel test (MPT) 405 (**2**), 407 (**2**), 410 (**2**), 427 (**2**)
multichip module (MCM) 87 (**2**), 571 (**2**), 606 (**2**), 659 (**2**)
multicore fiber cable 274 (**1**)
multiquantum well (MQW) 342 (**2**)
multiwalled carbon nanofibers (CNFs) 194 (**1**)
multiwalled CNT (MWCNT) 182 (**1**), 184 (**1**), 186 (**1**), 194 (**1**), 196 (**1**)
mutually-pumped phase conjugators (MPPC) 112 (**1**)

nano wafer level ACF (WLACF) 625 (**2**)
nano-electromechanical systems (NEMS) 136 (**1**), 247 (**2**), 320 (**2**)
nano-indentation 141 (**1**), 146 (**1**), 153 (**1**), 157 (**1**), 173 (**1**), 175 (**1**), 385 (**1**)
nano-machines 404 (**2**)
nano-particle material (NPM) 269 (**1**), 273 (**1**), 585 (**1**), 16 (**2**), 474 (**2**), 476 (**2**), 478 (**2**), 479 (**2**)
– cladding 274 (**1**)
NPM-based fibers 274 (**1**), 275 (**1**), 585 (**1**), 587 (**1**), 590 (**1**)
nanoDAC 239 (**2**), 242 (**2**), 246 (**2**)
nanofiber 182 (**1**), 188 (**1**), 192 (**1**)
nanostructures 181 (**1**)
National Electronics Manufacturing Initiative (NEMI) 429 (**1**)
Nickel substrate 373 (**1**)
non-conductive adhesive (NCA) 529 (**2**), 554 (**2**)
non-destructive evaluations (NDE) 222 (**2**), 477 (**2**)
non-linear deformation 351 (**1**), 489 (**1**)
non-linear plastic 430 (**1**)
non-solder mask defined (NSMD) 260 (**2**)
non-volatile 273 (**1**)
nondestructive pull test (NDPT) 89 (**2**)
nonhomogeneous Mathieu-Hill equation 675 (**1**)
nonlinear dynamics 630 (**1**)
– response 653 (**1**)
nonlinear thermo-elasticity theories 630 (**1**)
normal stress 221 (**1**), 440 (**1**), 442 (**1**), 527 (**1**), 14 (**2**), 61 (**2**), 63 (**2**), 395 (**2**), 494 (**2**), 684 (**2**)
Norton creep law 228 (**2**)
null field 465 (**1**)
– patterns 498 (**1**)
numerical aperture (NA) 23 (**2**), 26 (**2**)
numerical modeling 208 (**1**), 271 (**1**), 475 (**1**), 4 (**2**), 358 (**2**)

Oliver and Pharr method 162 (**1**)
on-wafer floating pad technology 139 (**2**)
ongoing AST 194 (**2**)
ongoing ESS 194 (**2**)
Optica Cross-connect (OXC) 371 (**2**)
optical beam 24 (**2**), 300 (**2**), 323 (**2**)
optical code division multiplexing 132 (**1**)
optical efficiency 342 (**2**), 373 (**2**)
optical fiber 67 (**1**), 68 (**1**), 79 (**1**), 86 (**1**), 112 (**1**), 114 (**1**), 142 (**1**), 269 (**1**)–272 (**1**), 274 (**1**), 275 (**1**), 483 (**1**), 572 (**1**), 576 (**1**), 609 (**1**), 683 (**1**), 4 (**2**), 23 (**2**), 24 (**2**), 37 (**2**), 63 (**2**), 151 (**2**), 157 (**2**), 160 (**2**), 166 (**2**), 226 (**2**), 365 (**2**), 367 (**2**), 478 (**2**),

488 (**2**), 490 (**2**), 498 (**2**), 502 (**2**), 512 (**2**), 518 (**2**)
– curvature 271 (**1**)
– reliability 575 (**1**)
– strength 582 (**1**)
– system 270 (**1**)
optical networking forum (OIF) 365 (**2**)
optical power loss 72 (**1**)
optimal fitting of a regression model objective 227 (**1**)
opto-electrical circuit board (OECB) 369 (**2**), 380 (**2**)
optoelectronic 139 (**1**), 273 (**1**), 557 (**1**), 3 (**2**), 14 (**2**), 151 (**2**), 165 (**2**), 203 (**2**), 204 (**2**), 225 (**2**), 230 (**2**), 299 (**2**), 311 (**2**), 330 (**2**), 338 (**2**), 383 (**2**), 403 (**2**), 473 (**2**), 480 (**2**), 487 (**2**), 498 (**2**), 503 (**2**), 522 (**2**)
optoelectronic holography (OEH) 330 (**2**)
optoelectronic laser interferometric microscopy (OELIM) 324 (**2**), 332 (**2**), 338 (**2**)
optoelectronics package 483 (**1**)
organic polymers 142 (**1**)
organic solderability preservative (OSP) 302 (**1**), 313 (**1**), 325 (**1**), 381 (**1**), 272 (**2**)
organic surface protection (OSP) 597 (**2**)
outer lead bonding (OLB) 76 (**2**)

package to board interconnection shear strength (PBISS) 263 (**2**)
packaging 3 (**1**), 4 (**1**), 35 (**1**), 261 (**1**), 262 (**1**), 135 (**2**), 177 (**2**), 255 (**2**), 294 (**2**), 302 (**2**), 311 (**2**), 320 (**2**), 348 (**2**), 358 (**2**), 383 (**2**), 390 (**2**), 432 (**2**), 433 (**2**), 442 (**2**), 450 (**2**), 473 (**2**), 477 (**2**), 487 (**2**), 489 (**2**), 495 (**2**), 522 (**2**), 527 (**2**), 529 (**2**), 546 (**2**), 572 (**2**), 607 (**2**), 611 (**2**), 613 (**2**), 625 (**2**), 634 (**2**), 636 (**2**), 652 (**2**), 657 (**2**), 660 (**2**), 697 (**2**)
pad design 302 (**1**)
pad metallurgy 74 (**2**), 273 (**2**)
parameterized polymer modeling (PPM) 53 (**1**), 60 (**1**), 62 (**1**)
Paris-Erdogan equation 226 (**2**)
partial differential equation (PDE) 207 (**1**), 324 (**2**)
partly state dependent modeling 10 (**1**), 34 (**1**), 35 (**1**)
passive alignment method 163 (**2**), 166 (**2**)
Pb-free HASL 381 (**1**)
Pb-free solder 351 (**1**), 267 (**2**), 274 (**2**)
– joint failure 388 (**1**)
– reliability 401 (**1**)
Pb-free solder alloys 353 (**1**), 380 (**1**), 384 (**1**), 401 (**1**), 404 (**1**), 429 (**1**), 273 (**2**)
Pb-Sn alloys 273 (**2**)
Peck and Black equations 227 (**2**)
peel test 405 (**2**), 426 (**2**)
peeling failure 412 (**1**)
peeling stress 7 (**2**), 61 (**2**)
percolation probability 581 (**2**), 585 (**2**), 586 (**2**), 594 (**2**)
percolation theory 538 (**2**), 574 (**2**), 578 (**2**), 581 (**2**), 612 (**2**)
peripheral array configuration 289 (**1**)
personal digital assistant (PDA) 253 (**2**), 259 (**2**), 572 (**2**), 621 (**2**)
phase fraction diagram 328 (**1**)
phase grating 461 (**1**), 479 (**1**)
phase separation 342 (**2**)
phase shifter 115 (**1**)
phase shifting technique 512 (**1**)
photo detector (PD) 366 (**2**), 489 (**2**), 513 (**2**)
photo-sensitive glasses 112 (**1**)
photoluminescence (PL) 346 (**2**)
photonic 65 (**1**), 114 (**1**), 136 (**1**), 270 (**1**), 272 (**1**), 475 (**1**), 555 (**1**), 4 (**2**), 8 (**2**), 16 (**2**), 23 (**2**), 32 (**2**), 151 (**2**), 204 (**2**), 221 (**2**)
– element 137 (**1**)
– reliability engineering 270 (**1**), 223 (**2**)
– system 111 (**1**), 23 (**2**), 44 (**2**), 67 (**2**)
photopolymers 112 (**1**)
photorefractive effect 112 (**1**)
photorefractive materials 112 (**1**)
photoresist 145 (**1**), 154 (**2**)
photosensitive optical fiber 67 (**1**)
photovoltaic effect 112 (**1**)
physical nonlinearity 668 (**1**), 678 (**1**)
physical vapor deposition (PVD) 144 (**1**)
physically nonlinear axial vibrations 683 (**1**)
piezoelectric coefficients 669 (**1**)
piezoelectric rod 671 (**1**), 672 (**1**), 683 (**1**)
– subject 678 (**1**)
piezoelectricity 668 (**1**)
pigtails 272 (**1**)
pin through hole (PTH) 611 (**2**)
pin-transfer dispensing 517 (**2**)
pitch 285 (**1**), 286 (**1**), 170 (**2**), 267 (**2**), 284 (**2**), 654 (**2**)
pitfall 204 (**2**), 219 (**2**), 255 (**2**), 268 (**2**)
pitting corrosion 287 (**2**)
placement machine 143 (**2**)
planar lightwave circuit (PLC) 225 (**2**), 520 (**2**)
– pigtailing 520 (**2**)

plasma enhanced CVD (PECVD) 187 (**1**)
plasma power 197 (**1**)
plastic ball grid array (PBGA) 263 (**2**)
plastic deformation 162 (**1**), 169 (**1**), 214 (**1**), 222 (**1**), 334 (**1**), 491 (**1**)
plastic encapsulated microcircuit (PEM) 494 (**1**)
plastic model 212 (**1**), 214 (**1**)
plastic package assembly reliability 289 (**1**)
plastic packages 288 (**1**)
plastic quad flat package (PQFP) 496 (**1**), 542 (**1**)
plastic strain 145 (**1**), 153 (**1**), 214 (**1**), 223 (**1**), 300 (**1**), 317 (**1**), 401 (**1**), 430 (**1**)–432 (**1**), 443 (**1**)–447 (**1**), 552 (**2**), 562 (**2**), 701 (**2**)
plastic wire-bond ball grid array 283 (**1**), 288 (**1**)
plated-through-hole (PTH) 289 (**1**), 302 (**1**), 502 (**1**)
plating material 424 (**1**)
pneumatic piston 605 (**1**)
Poincare map 637 (**1**), 643 (**1**), 644 (**1**)
point of presence (POP) 362 (**2**)
Poisson's ratio 16 (**1**), 23 (**1**), 26 (**1**), 147 (**1**), 159 (**1**), 167 (**1**), 212 (**1**), 398 (**1**), 524 (**1**), 572 (**1**), 635 (**1**), 234 (**2**), 347 (**2**), 409 (**2**), 413 (**2**), 422 (**2**), 451 (**2**), 481 (**2**), 549 (**2**), 653 (**2**), 687 (**2**)
poly-arylethers (PAE) 704 (**2**)
polybutadiene rubber 5 (**1**)
polycarbonate 4 (**1**)
polyethylene 5 (**1**)
poly(ethylene terephthalate) (PET) 403 (**2**), 419 (**2**)
polyimide (PI) 142 (**1**), 145 (**1**), 79 (**2**), 138 (**2**), 313 (**2**), 364 (**2**), 413 (**2**)
polyisoprene rubber 5 (**1**)
polymer 3 (**1**), 65 (**1**), 113 (**1**), 142 (**1**), 157 (**1**), 185 (**1**), 212 (**1**), 263 (**1**), 269 (**1**), 277 (**1**), 585 (**1**), 13 (**2**), 108 (**2**), 131 (**2**), 138 (**2**), 244 (**2**), 370 (**2**), 379 (**2**), 403 (**2**), 529 (**2**), 534 (**2**), 540 (**2**), 574 (**2**), 589 (**2**), 612 (**2**)–614 (**2**)
– fiber Bragg grating (PFBG) 66 (**1**), 92 (**1**), 99 (**1**)
– film 137 (**1**)
– materials 3 (**1**), 65 (**1**)
PFBG–LED illumination 87 (**1**)
PFBG–SM LD illumination 92 (**1**)
polymer adhesion 535 (**2**)
polymer materials 313 (**2**)
polymer matrix 536 (**2**), 614 (**2**), 619 (**2**)
polymer-coated fiber 269 (**1**), 272 (**1**), 11 (**2**), 16 (**2**)
polymethylmethacrylate (PMMA) 84 (**1**)
– FBG 102 (**1**)
– PFBG 84 (**1**)
PolyMUMPS 304 (**2**)
polynomial model 238 (**1**)
polypropylene 5 (**1**)
polysilicon 324 (**2**)
polystyrene 4 (**1**)
popcorn 511 (**1**), 512 (**1**), 524 (**1**), 525 (**1**), 185 (**2**)
– cracking 524 (**1**), 13 (**2**)
– effect 511 (**1**)
positive displacement piston 517 (**2**)
post-screening experiment 256 (**1**)
potato chip effect 145 (**1**)
power cycling 315 (**1**), 389 (**1**), 400 (**1**), 413 (**1**), 520 (**1**), 8 (**2**), 87 (**2**), 96 (**2**), 125 (**2**), 126 (**2**), 192 (**2**), 215 (**2**), 293 (**2**), 698 (**2**)
power spectral density (PSD) 91 (**2**)
predictive modeling 262 (**1**), 557 (**1**), 222 (**2**), 447 (**2**)
pressurized blister test 404 (**2**)
pressurized bulge testing 165 (**1**)
printed circuit board (PCB) 264 (**1**), 352 (**1**), 377 (**1**), 459 (**1**), 482 (**1**), 556 (**1**), 8 (**2**), 244 (**2**), 313 (**2**), 366 (**2**), 380 (**2**), 573 (**2**), 599 (**2**), 604 (**2**), 692 (**2**)
printed wiring board (PWB) 283 (**1**), 295 (**1**), 313 (**1**), 629 (**1**), 631 (**1**), 108 (**2**), 135 (**2**), 257 (**2**), 270 (**2**), 271 (**2**), 276 (**2**), 294 (**2**), 652 (**2**)
pristine fiber 608 (**1**), 612 (**1**)
probabilistic design 46 (**2**), 49 (**2**), 66 (**2**), 230 (**2**)
probabilistic models 15 (**2**)
probability 183 (**1**), 184 (**1**), 217 (**1**), 235 (**1**), 261 (**1**), 319 (**1**), 388 (**1**), 455 (**1**), 571 (**1**), 574 (**1**), 581 (**1**), 589 (**1**)–591 (**1**), 15 (**2**), 35 (**2**), 44 (**2**), 67 (**2**), 87 (**2**), 104 (**2**), 177 (**2**), 184 (**2**), 201 (**2**), 203 (**2**), 209 (**2**), 229 (**2**), 230 (**2**), 253 (**2**), 268 (**2**), 389 (**2**), 443 (**2**), 533 (**2**), 574 (**2**), 578 (**2**), 589 (**2**), 607 (**2**)
– density distribution 217 (**1**), 197 (**2**)
– density function (PDF) 319 (**1**), 455 (**1**), 574 (**1**), 590 (**1**), 180 (**2**), 195 (**2**), 229 (**2**)
probability-based reliability engineering 205 (**2**)

product development/verification tests (PDTs) 215 (**2**), 217 (**2**)
product level drop tests 258 (**2**)
product reliability 136 (**1**), 219 (**1**), 313 (**1**)
production-sampling AST 193 (**2**)
Prony series 11 (**1**), 52 (**1**)
pull strength 104 (**2**), 185 (**2**)
pull test 258 (**2**)
pure gold 102 (**2**)
pyramidal structure 639 (**2**)

quad flat pack no-lead (QFN) 270 (**2**), 272 (**2**), 642 (**2**)
quad flat package (QFP) 284 (**1**), 544 (**1**), 573 (**2**)
qualification tests (QTs) 260 (**1**), 483 (**1**), 210 (**2**), 212 (**2**), 214 (**2**), 217 (**2**)
quality assurance (QA) 283 (**1**)
quantum confined (QC) 342 (**2**)
quantum efficiency 348 (**2**)
quantum resistance 184 (**1**)

radial basis function (RBF) model 241 (**1**)
Raman microscopy 677 (**2**)
Raman spectroscopy 346 (**2**)
random vibration tests 215 (**2**)
ray tracing method 25 (**2**)
Rayleigh criterion 514 (**1**)
real-time observation 470 (**1**), 485 (**1**)
receiver 364 (**2**), 368 (**2**), 487 (**2**)
recrystallization 151 (**1**), 335 (**1**)
recrystallized grain boundaries (RGB) 275 (**2**)
reference beam 332 (**2**)
reference grating 510 (**1**)
reference mirror 506 (**1**)
reflected power spectrum 90 (**1**), 97 (**1**)
reflection peak 120 (**1**)
reflectivity spectrum 89 (**1**), 96 (**1**)
reflow process 285 (**1**), 298 (**1**), 329 (**1**), 387 (**1**), 413 (**1**), 460 (**1**), 491 (**1**)
reflow soldering 5 (**1**), 316 (**1**), 330 (**1**), 381 (**1**), 412 (**1**), 8 (**2**), 74 (**2**), 496 (**2**)
refractive index 68 (**1**), 69 (**1**), 84 (**1**), 113 (**1**), 115 (**1**), 118 (**1**), 120 (**1**)–123 (**1**), 128 (**1**), 273 (**1**)–275 (**1**), 27 (**2**), 387 (**2**), 477 (**2**), 491 (**2**), 503 (**2**), 520 (**2**)
refractive losses 27 (**2**)
regular Fabry-Perot etalon 114 (**1**)
regular Fabry-Perot filter 116 (**1**)
regular fiber 587 (**1**)
relative dispersion slope (RDS) 127 (**1**)
relative humidity (RH) 500 (**1**), 576 (**1**), 587 (**1**), 225 (**2**), 254 (**2**), 316 (**2**), 438 (**2**), 615 (**2**), 618 (**2**)
relaxation description 13 (**1**), 18 (**1**)
relaxation modulus functions 8 (**1**)
relaxation shear modulus 27 (**1**)
relaxation time 7 (**1**)
reliability 3 (**1**), 53 (**1**), 144 (**1**), 217 (**1**), 270 (**1**), 283 (**1**), 288 (**1**), 292 (**1**), 315 (**1**), 337 (**1**), 395 (**1**), 454 (**1**), 475 (**1**), 524 (**1**), 557 (**1**), 571 (**1**), 585 (**1**), 591 (**1**), 684 (**1**), 4 (**2**), 15 (**2**), 46 (**2**), 47 (**2**), 83 (**2**), 87 (**2**), 97 (**2**), 104 (**2**), 110 (**2**), 124 (**2**), 129 (**2**), 136 (**2**), 152 (**2**), 177 (**2**), 178 (**2**), 186 (**2**), 187 (**2**), 203 (**2**), 204 (**2**), 208 (**2**), 209 (**2**), 211 (**2**)–213 (**2**), 230 (**2**), 253 (**2**), 267 (**2**), 268 (**2**), 281 (**2**), 293 (**2**), 294 (**2**), 300 (**2**), 301 (**2**), 303 (**2**), 316 (**2**), 348 (**2**), 357 (**2**), 362 (**2**), 363 (**2**), 372 (**2**), 383 (**2**), 384 (**2**), 388 (**2**), 395 (**2**), 397 (**2**), 399 (**2**), 403 (**2**), 433 (**2**), 442 (**2**), 473 (**2**), 488 (**2**), 513 (**2**), 522 (**2**), 554 (**2**), 559 (**2**), 572 (**2**), 602 (**2**), 612 (**2**), 623 (**2**), 630 (**2**), 644 (**2**), 647 (**2**), 652 (**2**), 677 (**2**), 691 (**2**)
– costs money 208 (**2**)
– distribution function (RDF) 571 (**1**), 574 (**1**)
– engineering 270 (**1**), 319 (**1**), 204 (**2**), 205 (**2**), 214 (**2**)
– failures 6 (**1**), 143 (**1**), 205 (**1**), 217 (**1**), 262 (**1**), 270 (**1**), 289 (**1**), 303 (**1**), 314 (**1**), 395 (**1**), 455 (**1**), 4 (**2**), 13 (**2**), 182 (**2**), 210 (**2**), 258 (**2**), 317 (**2**)
– hardware 270 (**1**), 191 (**2**), 294 (**2**)
– lead-free CSPs in thermal cycling 337 (**1**)
– lead-free lead-free CSPs in drop testing 341 (**1**)
– models 389 (**1**)
– software 550 (**1**), 110 (**2**), 188 (**2**)
– standards 135 (**1**), 483 (**1**), 207 (**2**), 208 (**2**)
– test 199 (**1**), 260 (**1**), 454 (**1**)
– testing 305 (**1**), 315 (**1**), 316 (**1**), 335 (**1**), 483 (**1**), 4 (**2**), 255 (**2**), 267 (**2**), 280 (**2**), 294 (**2**), 314 (**2**), 341 (**2**), 546 (**2**), 685 (**2**)
reparability 206 (**2**)
repeatability 126 (**1**), 603 (**1**), 110 (**2**), 206 (**2**), 259 (**2**), 370 (**2**), 399 (**2**)
repeated stress concentration 413 (**1**)
residual resistivity ratio (RRR) 151 (**1**)
residual strain 96 (**2**)
residual stress 4 (**1**), 145 (**1**), 150 (**1**), 156 (**1**), 163 (**1**), 166 (**1**), 170 (**1**), 385 (**1**), 460 (**1**),

267 (**2**), 405 (**2**), 409 (**2**), 413 (**2**), 423 (**2**), 432 (**2**), 700 (**2**)
resin 6 (**1**), 35 (**1**), 39 (**1**)–41 (**1**), 55 (**1**), 336 (**1**), 459 (**1**), 503 (**1**), 109 (**2**), 244 (**2**), 272 (**2**), 441 (**2**), 509 (**2**), 613 (**2**)
resistance heater 465 (**1**)
resistance welding 483 (**1**), 488 (**2**)
response surface modeling (RSM) 206 (**1**), 236 (**1**), 242 (**1**)
– method 226 (**1**)
response time 129 (**1**)
response transition 651 (**1**)
Restriction of Hazardous Substances (ROHS) 429 (**1**)
rework methods 384 (**1**)
rheologically simple 20 (**1**)
rise time 124 (**1**)
risk 283 (**1**), 285 (**1**), 313 (**1**), 319 (**1**), 337 (**1**), 386 (**1**), 81 (**2**), 180 (**2**), 263 (**2**), 267 (**2**), 268 (**2**), 291 (**2**), 318 (**2**), 372 (**2**), 514 (**2**), 547 (**2**)
robust design 206 (**1**), 226 (**1**), 228 (**1**), 251 (**1**), 40 (**2**), 41 (**2**), 190 (**2**)
robust design objective 228 (**1**)
robust scheme 470 (**1**)
robustness 140 (**1**), 205 (**1**), 260 (**1**), 263 (**1**), 23 (**2**), 187 (**2**), 192 (**2**), 199 (**2**), 397 (**2**)
root cause 205 (**1**), 215 (**1**), 337 (**1**), 384 (**1**), 204 (**2**), 255 (**2**), 348 (**2**)
– analysis 215 (**1**), 177 (**2**), 180 (**2**), 294 (**2**)
root of the sum of the squares (RSS) 331 (**2**)
Rouard method 82 (**1**)
rubber 4 (**1**), 5 (**1**), 479 (**1**), 605 (**1**), 392 (**2**)
rupture strain 419 (**2**), 421 (**2**)

safe stress model 598 (**1**)
sample preparation 468 (**1**), 602 (**1**)
sandwich beam shear tool (SBT) 39 (**1**)
scanning acoustic microscopy (SAM) 321 (**1**), 523 (**1**), 539 (**1**), 541 (**1**), 282 (**2**), 353 (**2**), 644 (**2**)
scanning electron microscopy (SEM) 148 (**1**), 602 (**1**), 72 (**2**), 127 (**2**), 234 (**2**), 350 (**2**), 512 (**2**)
scanning force microscopy (SFM) 233 (**2**), 238 (**2**), 239 (**2**), 242 (**2**), 250 (**2**)
scanning probe microscopy 185 (**1**)
scanning transmission electron microscopy-atomic number (STEM-Z) 343 (**2**)
scratch test 404 (**2**)
screening experiment 255 (**1**)
screening objective 227 (**1**)
sea of leads (SoL) 139 (**2**), 140 (**2**)
second-order elastic constants 137 (**1**)
seeding by particle 602 (**1**)
self-assembled monolayer (SAM) 301 (**2**), 316 (**2**), 318 (**2**), 621 (**2**)
self-pumped phase conjugators (SPPC) 112 (**1**)
semi-crystalline polymers 5 (**1**)
semiconductor optical amplifier (SOA) 489 (**2**)
sequential RSM 244 (**1**)
shadow moiré 459 (**1**), 476 (**1**), 509 (**1**), 652 (**2**), 656 (**2**), 661 (**2**)
Shapiro-Wilk test 319 (**1**)
shear creep compliance 24 (**1**)
shear modulus 27 (**1**), 212 (**1**), 60 (**2**), 97 (**2**), 347 (**2**)
shear relaxation modulus 15 (**1**), 19 (**1**)
– curve 33 (**1**)
shear strain 219 (**1**), 481 (**1**)
shear strength 93 (**2**), 102 (**2**), 114 (**2**), 263 (**2**), 451 (**2**), 547 (**2**), 554 (**2**), 555 (**2**), 599 (**2**), 601 (**2**)
shear stress 19 (**1**), 28 (**1**), 41 (**1**), 218 (**1**), 334 (**1**), 527 (**1**), 61 (**2**), 67 (**2**), 136 (**2**), 275 (**2**), 352 (**2**), 394 (**2**), 493 (**2**), 510 (**2**), 511 (**2**), 537 (**2**), 684 (**2**)
shearing strength 390 (**1**)
shearing stress 474 (**2**)
shock load 302 (**1**), 556 (**1**), 261 (**2**)
shock response spectrum (SRS) 259 (**2**)
shock test 556 (**1**), 262 (**2**), 642 (**2**)
short bare fibers 685 (**1**)
short-range thickness variation SR-TV 343 (**2**)
short-term reliability 272 (**1**), 9 (**2**), 187 (**2**), 215 (**2**)
shutter 129 (**1**)
side-contacted nanotubes 197 (**1**)
silane 393 (**2**), 506 (**2**), 508 (**2**)
silica glass 65 (**1**)
– fiber 571 (**1**)
silicon nitride film 142 (**1**)
silicon optical bench (SiOB) 152 (**2**), 162 (**2**)
silicon-on-insulator (SOI) 135 (**1**)
silicon-on-silicon flip-chip assemblies 12 (**2**)
silver-tin 9 (**2**)
simulated annealing 34 (**2**), 35 (**2**)
single mode laser diode (SM LD) 71 (**1**), 95 (**1**)
single walled CNT (SWCNT) 182 (**1**), 183 (**1**)
single-coated fiber 269 (**1**)

single-mode fiber (SMF) 27 (**2**), 151 (**2**), 167 (**2**), 330 (**2**), 333 (**2**), 362 (**2**), 365 (**2**), 489 (**2**), 492 (**2**), 513 (**2**)
single-region power-law model 596 (**1**)
sinusoidal vibration tests 215 (**2**)
SiO_2 185 (**1**)
small displacement 558 (**1**)
small form factor pluggable (SFP) 364 (**2**)
Sn-Ag-Bi 353 (**1**), 412 (**1**)
Sn-Ag-Cu 351 (**1**), 353 (**1**), 359 (**1**), 375 (**1**), 378 (**1**), 389 (**1**), 412 (**1**)
Sn-Ag-Cu alloys 378 (**1**)
Sn-Cu 331 (**1**), 353 (**1**), 355 (**1**), 378 (**1**), 412 (**1**)
– alloy 378 (**1**)
Sn-Pb eutectic solder 351 (**1**), 273 (**2**)
Sn-Pb solder 352 (**1**), 384 (**1**), 386 (**1**)–388 (**1**), 394 (**1**), 401 (**1**), 405 (**1**), 411 (**1**), 413 (**1**), 423 (**1**), 424 (**1**), 433 (**1**), 446 (**1**), 273 (**2**)
Sn-Zn-Bi 353 (**1**), 412 (**1**)
SnPb-based solders 313 (**1**)
soft solder 9 (**2**), 129 (**2**), 348 (**2**)
soft-core model 578 (**2**)
solder 294 (**1**)–299 (**1**), 302 (**1**)–306 (**1**), 310 (**1**), 313 (**1**)–318 (**1**), 323 (**1**)–325 (**1**), 330 (**1**)–334 (**1**), 347 (**1**), 361 (**1**), 375 (**1**)–391 (**1**), 393 (**1**)–405 (**1**), 411 (**1**)–425 (**1**), 439 (**1**)–455 (**1**), 491 (**1**)–494 (**1**), 8 (**2**), 77 (**2**), 145 (**2**), 185 (**2**), 258 (**2**), 262 (**2**), 274 (**2**), 284 (**2**), 306 (**2**), 309 (**2**)–311 (**2**), 348 (**2**), 349 (**2**), 377 (**2**), 384 (**2**), 389 (**2**), 450 (**2**), 479 (**2**), 488 (**2**), 497 (**2**)–500 (**2**), 502 (**2**), 522 (**2**), 527 (**2**), 542 (**2**), 546 (**2**), 557 (**2**), 565 (**2**), 601 (**2**), 604 (**2**), 606 (**2**), 611 (**2**), 619 (**2**), 620 (**2**), 625 (**2**), 646 (**2**), 672 (**2**), 685 (**2**), 698 (**2**), 699 (**2**)
– ball 483 (**1**), 489 (**1**), 517 (**1**), 74 (**2**), 136 (**2**), 263 (**2**), 270 (**2**), 274 (**2**), 282 (**2**), 476 (**2**), 700 (**2**)
– bump 143 (**1**), 313 (**1**), 337 (**1**)–339 (**1**), 401 (**1**), 502 (**1**), 74 (**2**), 306 (**2**), 652 (**2**), 653 (**2**)
– column interposer 294 (**1**)
– fatigue failure mode 412 (**1**), 415 (**1**)
– glass 143 (**1**), 358 (**1**), 389 (**1**), 502 (**1**), 12 (**2**), 74 (**2**), 143 (**2**)
– interconnection 283 (**1**), 306 (**1**), 313 (**1**), 328 (**1**), 351 (**1**), 256 (**2**)
– interface 272 (**1**), 298 (**1**), 310 (**1**), 324 (**1**), 330 (**1**)–333 (**1**), 339 (**1**), 347 (**1**), 357 (**1**), 411 (**1**)–413 (**1**), 415 (**1**)–423 (**1**), 525 (**1**), 270 (**2**), 356 (**2**), 493 (**2**), 495 (**2**)
– mask 306 (**1**), 311 (**1**), 74 (**2**), 260 (**2**), 273 (**2**), 450 (**2**), 657 (**2**), 658 (**2**), 674 (**2**)
– – defined (SMD) 260 (**2**)
– materials 272 (**1**)
– pad 297 (**1**), 302 (**1**)–304 (**1**), 325 (**1**), 338 (**1**), 347 (**1**), 398 (**1**), 411 (**1**), 417 (**1**), 491 (**1**), 525 (**1**), 76 (**2**), 110 (**2**), 140 (**2**), 185 (**2**), 312 (**2**)
– paste 296 (**1**), 298 (**1**), 304 (**1**), 316 (**1**), 337 (**1**), 341 (**1**), 354 (**1**), 382 (**1**), 387 (**1**), 77 (**2**), 542 (**2**), 573 (**2**)
– reflow 285 (**1**), 298 (**1**), 316 (**1**), 337 (**1**), 380 (**1**), 386 (**1**), 388 (**1**), 413 (**1**), 429 (**1**), 513 (**1**), 525 (**1**), 75 (**2**), 102 (**2**), 110 (**2**), 142 (**2**), 185 (**2**), 557 (**2**), 621 (**2**)
– reflow process 429 (**1**), 523 (**1**)
– strength 141 (**1**), 273 (**1**), 318 (**1**), 342 (**1**), 391 (**1**), 414 (**1**), 424 (**1**), 262 (**2**), 263 (**2**)
– volume 303 (**1**)
– wetting 330 (**1**), 351 (**1**), 357 (**1**), 358 (**1**), 383 (**1**), 387 (**1**), 143 (**2**)
solder joint 220 (**1**), 225 (**1**), 272 (**1**), 289 (**1**), 297 (**1**), 351 (**1**), 356 (**1**), 380 (**1**), 413 (**1**), 8 (**2**), 75 (**2**), 80 (**2**), 136 (**2**), 142 (**2**), 163 (**2**), 192 (**2**), 225 (**2**), 260 (**2**), 261 (**2**), 268 (**2**), 275 (**2**)–277 (**2**), 279 (**2**), 281 (**2**), 293 (**2**), 309 (**2**), 313 (**2**), 432 (**2**), 498 (**2**), 499 (**2**), 502 (**2**), 552 (**2**), 604 (**2**), 651 (**2**), 685 (**2**), 698 (**2**), 700 (**2**)
– failure 220 (**1**), 225 (**1**), 290 (**1**), 295 (**1**), 306 (**1**), 388 (**1**)–390 (**1**), 411 (**1**)–414 (**1**), 418 (**1**)–425 (**1**), 434 (**1**), 451 (**1**), 281 (**2**)
– fracture 220 (**1**), 225 (**1**), 390 (**1**), 396 (**1**), 413 (**1**), 417 (**1**)–419 (**1**), 424 (**1**), 454 (**1**), 280 (**2**)
– reliability 285 (**1**), 292 (**1**), 297 (**1**), 306 (**1**), 389 (**1**)
solderability 302 (**1**), 325 (**1**), 336 (**1**), 384 (**1**), 185 (**2**), 272 (**2**), 284 (**2**)
soldered assembly 315 (**1**), 309 (**2**)
soldering 253 (**1**), 254 (**1**), 314 (**1**), 316 (**1**), 323 (**1**), 330 (**1**), 337 (**1**), 348 (**1**), 351 (**1**), 74 (**2**), 185 (**2**), 302 (**2**), 488 (**2**), 493 (**2**), 494 (**2**), 511 (**2**), 522 (**2**), 552 (**2**), 563 (**2**), 602 (**2**), 604 (**2**), 606 (**2**), 633 (**2**), 646 (**2**)
solid surface 183 (**1**), 192 (**1**), 360 (**1**)
solidification 324 (**1**), 328 (**1**), 330 (**1**), 347 (**1**), 352 (**1**)
– structure 325 (**1**)
Solomon's low cycle fatigue model 432 (**1**)

specimen gratings 461 (**1**), 479 (**1**)
spectral dispersion 89 (**1**)
spectral heat generation density 87 (**1**), 93 (**1**)
spectral power density 71 (**1**), 86 (**1**), *see also* power spectral density (PSD)
splitter 506 (**1**), 186 (**2**), 332 (**2**), 384 (**2**), 487 (**2**), 656 (**2**)
spreading 145 (**1**), 357 (**1**)
– rate 360 (**1**)
spring 6 (**1**), 209 (**1**), 558 (**1**), 15 (**2**), 31 (**2**), 52 (**2**), 73 (**2**), 261 (**2**), 326 (**2**)
– constant 209 (**1**), 558 (**1**), 564 (**1**), 15 (**2**)
– element 6 (**1**), 209 (**1**), 561 (**1**), 52 (**2**)
staggered array configuration 289 (**1**)
stain rate 700 (**2**)
static fatigue 610 (**1**), 611 (**1**), 4 (**2**)
– test 600 (**1**), 605 (**1**), 620 (**1**), 623 (**1**)
static stress fatigue 186 (**2**)
static testing 19 (**1**)
– of creep compliances 24 (**1**)
statistical modeling 584 (**1**)
statistical testing 319 (**1**)
steady-state crack growth 223 (**1**)
steady-state creep strain rate 440 (**1**)
steady-state in-plane thermal field 638 (**1**)
steady-state peeling 409 (**2**)
steady-state temperature effect 633 (**1**)
steady-state thermal field 633 (**1**), 634 (**1**), 660 (**1**)
stencil design 302 (**1**)
stencil parameter 304 (**1**)
stiff adherends 473 (**2**)
stiffness coefficients 29 (**1**)
stiffness matrix 210 (**1**)
stochastic modeling 594 (**1**), 24 (**2**)
stochastical model 240 (**1**)
storage bulk modulus 39 (**1**)
storage longitudinal modulus 39 (**1**)
storage modulus curve 33 (**1**)
storage shear modulus 39 (**1**)
strain 7 (**1**), 97 (**1**), 139 (**1**), 143 (**1**), 173 (**1**), 185 (**1**), 210 (**1**), 211 (**1**), 213 (**1**), 262 (**1**), 270 (**1**), 294 (**1**), 313 (**1**), 316 (**1**), 318 (**1**), 333 (**1**), 385 (**1**), 389 (**1**), 419 (**1**), 430 (**1**), 448 (**1**), 460 (**1**), 475 (**1**), 568 (**1**), 572 (**1**), 668 (**1**), 685 (**1**), 687 (**1**), 3 (**2**), 52 (**2**), 54 (**2**), 61 (**2**), 97 (**2**), 138 (**2**), 141 (**2**), 219 (**2**), 233 (**2**), 234 (**2**), 241 (**2**), 257 (**2**)–259 (**2**), 347 (**2**), 373 (**2**), 403 (**2**), 425 (**2**), 426 (**2**), 442 (**2**), 451 (**2**), 456 (**2**), 497 (**2**), 500 (**2**), 552 (**2**), 686 (**2**)
– analysis 430 (**1**), 462 (**1**), 480 (**1**)
– based approach 223 (**1**)
– constant 7 (**1**)
– distribution 218 (**1**), 385 (**1**), 421 (**1**), 574 (**1**), 578 (**1**), 258 (**2**), 418 (**2**), 677 (**2**)
– effective 7 (**1**)
– energy 166 (**1**), 210 (**1**), 213 (**1**)–215 (**1**), 225 (**1**), 404 (**1**), 433 (**1**), 533 (**1**), 673 (**1**), 692 (**1**), 347 (**2**), 410 (**2**), 476 (**2**), 547 (**2**), 686 (**2**)
– energy release rate (SERR) 166 (**1**), 169 (**1**), 529 (**1**), 549 (**1**)
– field 181 (**1**), 301 (**1**), 385 (**1**), 460 (**1**), 475 (**1**), 499 (**1**), 537 (**1**), 631 (**1**), 673 (**1**), 684 (**1**), 242 (**2**)
– gage techniques 460 (**1**)
– hardening coefficient 412 (**2**), 422 (**2**)
– inelastic 215 (**1**), 222 (**1**), 225 (**1**), 393 (**1**), 403 (**1**), 433 (**1**), 448 (**1**), 243 (**2**), 292 (**2**)
– initial 10 (**1**), 692 (**1**)
– plastic 145 (**1**), 153 (**1**), 223 (**1**), 300 (**1**), 317 (**1**)
– rate 166 (**1**), 214 (**1**), 224 (**1**), 334 (**1**), 342 (**1**), 344 (**1**), 393 (**1**), 430 (**1**), 442 (**1**), 491 (**1**), 537 (**1**), 605 (**1**), 52 (**2**), 125 (**2**), 257 (**2**), 262 (**2**), 279 (**2**), 497 (**2**), 537 (**2**)
– tensile 185 (**1**)
– tensor 532 (**1**)
– volumetric 14 (**1**)
strength 141 (**1**), 157 (**1**), 172 (**1**), 198 (**1**), 263 (**1**), 270 (**1**), 556 (**1**), 571 (**1**), 591 (**1**), 613 (**1**), 625 (**1**), 5 (**2**), 8 (**2**), 11 (**2**), 26 (**2**), 46 (**2**), 75 (**2**), 82 (**2**), 114 (**2**), 191 (**2**), 198 (**2**), 227 (**2**), 263 (**2**), 401 (**2**), 405 (**2**), 473 (**2**), 511 (**2**)
– distribution 357 (**1**), 426 (**1**), 576 (**1**), 582 (**1**), 584 (**1**), 585 (**1**), 591 (**1**), 595 (**1**), 606 (**1**), 655 (**1**), 196 (**2**), 229 (**2**), 258 (**2**)
– distribution function 229 (**2**)
– tests 590 (**1**)
stress 6 (**1**), 136 (**1**), 144 (**1**), 145 (**1**), 185 (**1**), 211 (**1**), 222 (**1**), 248 (**1**), 262 (**1**), 270 (**1**), 272 (**1**), 299 (**1**), 313 (**1**), 316 (**1**), 333 (**1**), 347 (**1**), 380 (**1**), 385 (**1**), 389 (**1**), 419 (**1**), 430 (**1**), 448 (**1**), 460 (**1**), 475 (**1**), 524 (**1**), 556 (**1**), 572 (**1**), 575 (**1**), 595 (**1**), 597 (**1**), 657 (**1**), 668 (**1**), 684 (**1**), 687 (**1**), 3 (**2**), 10 (**2**), 15 (**2**), 23 (**2**), 52 (**2**), 54 (**2**), 77 (**2**), 110 (**2**), 125 (**2**), 136 (**2**), 146 (**2**), 152 (**2**), 189 (**2**), 198 (**2**), 204 (**2**), 216 (**2**), 221 (**2**), 241 (**2**), 275 (**2**), 348 (**2**), 350 (**2**), 355 (**2**), 357 (**2**), 358 (**2**), 373 (**2**), 374 (**2**), 376 (**2**), 378 (**2**), 385 (**2**)–387 (**2**), 393 (**2**), 395 (**2**),

397 (**2**), 404 (**2**), 409 (**2**), 411 (**2**), 442 (**2**), 473 (**2**), 494 (**2**), 497 (**2**), 502 (**2**), 523 (**2**), 574 (**2**), 681 (**2**), 686 (**2**), 688 (**2**), 690 (**2**), 694 (**2**), 697 (**2**)
– compensation layer (SCL) 138 (**2**)
– compressive 145 (**1**), 166 (**1**), 271 (**1**), 384 (**1**)
– cracking 220 (**1**), 225 (**1**), 253 (**1**), 298 (**1**), 325 (**1**), 603 (**1**), 87 (**2**), 185 (**2**)
– distribution 41 (**1**), 218 (**1**), 295 (**1**), 272 (**1**), 299 (**1**), 389 (**1**), 524 (**1**), 527 (**1**), 654 (**1**), 655 (**1**), 58 (**2**), 162 (**2**), 229 (**2**), 419 (**2**), 677 (**2**), 682 (**2**)
– – function 217 (**1**), 388 (**1**), 571 (**1**), 574 (**1**), 229 (**2**)
– field 157 (**1**), 220 (**1**), 536 (**1**), 654 (**1**), 661 (**1**), 242 (**2**)
– film 145 (**1**)
– generation 145 (**1**), 550 (**2**)
– gradient 385 (**1**), 350 (**2**)
– initial 10 (**1**), 392 (**1**)
– intensity factor (SIF) 524 (**1**), 525 (**1**), 528 (**1**), 226 (**2**), 248 (**2**), 451 (**2**), 466 (**2**)
– intrinsic 146 (**1**), 677 (**2**), 678 (**2**), 681 (**2**)
– rate 166 (**1**), 170 (**1**), 214 (**1**), 224 (**1**), 334 (**1**), 344 (**1**), 393 (**1**), 430 (**1**), 442 (**1**), 598 (**1**), 623 (**1**), 52 (**2**), 125 (**2**), 177 (**2**), 256 (**2**), 536 (**2**)
– relaxation 6 (**1**), 8 (**1**), 19 (**1**), 34 (**1**), 52 (**1**), 145 (**1**), 215 (**1**), 389 (**1**), 419 (**1**), 493 (**1**), 4 (**2**), 228 (**2**), 348 (**2**), 423 (**2**), 604 (**2**)
– residual 148 (**1**), 156 (**1**), 609 (**1**)
– screening 254 (**1**)–257 (**1**), 389 (**1**), 191 (**2**)
– service 613 (**1**)
– state 9 (**1**), 334 (**1**)
– tensile hydrostatic 155 (**1**)
– tensor 672 (**1**)
– testing 19 (**1**)–21 (**1**), 28 (**1**), 157 (**1**), 174 (**1**), 262 (**1**), 272 (**1**), 347 (**1**), 390 (**1**), 454 (**1**), 595 (**1**)–598 (**1**), 613 (**1**), 623 (**1**), 177 (**2**), 191 (**2**), 198 (**2**), 542 (**2**)
– thermal 145 (**1**), 156 (**1**), 222 (**1**), 388 (**1**), 398 (**1**), 401 (**1**)–405 (**1**), 500 (**1**), 532 (**1**), 627 (**1**), 654 (**1**), 662 (**1**), 185 (**2**), 201 (**2**), 217 (**2**), 404 (**2**), 703 (**2**)
– thin film 145 (**1**)
– volumetric 14 (**1**)
stress–strain relationship 10 (**1**), 220 (**1**), 269 (**1**), 421 (**1**), 683 (**1**), 688 (**1**), 694 (**1**), 421 (**2**), 422 (**2**), 424 (**2**), 427 (**2**), 497 (**2**), 500 (**2**)
stress-free nano-wires 563 (**1**)
– wire 559 (**1**), 560 (**1**)
stress/strain analysis 270 (**1**), 476 (**1**)
striction 262 (**1**), 303 (**2**)
structural analysis 209 (**1**), 270 (**1**), 277 (**1**), 3 (**2**), 480 (**2**)
structural nonlinearity 271 (**1**)
structural reliability 271 (**1**), 3 (**2**), 4 (**2**), 210 (**2**), 474 (**2**)
structure function 633 (**2**), 643 (**2**), 645 (**2**), 649 (**2**)
structure function evaluation methodology 630 (**2**), 636 (**2**)
structure zone model 148 (**1**)
stud bump 115 (**2**), 377 (**2**), 528 (**2**)
subpixel analysis 236 (**2**)
substrate 6 (**1**), 65 (**1**), 137 (**1**), 187 (**1**), 218 (**1**), 292 (**1**), 296 (**1**), 323 (**1**), 330 (**1**), 333 (**1**), 481 (**1**), 501 (**1**), 517 (**1**), 557 (**1**), 6 (**2**), 13 (**2**), 14 (**2**), 47 (**2**), 53 (**2**), 74 (**2**), 81 (**2**), 109 (**2**), 135 (**2**), 142 (**2**), 149 (**2**), 256 (**2**), 267 (**2**), 307 (**2**), 311 (**2**), 357 (**2**), 364 (**2**), 387 (**2**), 393 (**2**), 395 (**2**), 400 (**2**), 401 (**2**), 403 (**2**), 405 (**2**), 459 (**2**), 466 (**2**), 494 (**2**), 504 (**2**), 552 (**2**), 564 (**2**), 573 (**2**), 605 (**2**), 648 (**2**), 651 (**2**), 657 (**2**), 661 (**2**), 664 (**2**), 669 (**2**), 672 (**2**), 683 (**2**), 688 (**2**), 693 (**2**)
– dissolution 363 (**1**)
– temperature 149 (**1**)
surface insulation resistance (SIR) 384 (**1**), 264 (**2**), 266 (**2**)
surface laminar circuitry (SLC) 271 (**2**)
surface mount technology (SMT) 285 (**1**), 411 (**1**), 429 (**1**), 135 (**2**), 141 (**2**), 163 (**2**), 368 (**2**), 380 (**2**), 527 (**2**), 595 (**2**), 611 (**2**), 613 (**2**)
surface relief grating 67 (**1**)
surface tension 360 (**1**), 389 (**1**), 401 (**1**), 480 (**1**), 74 (**2**), 83 (**2**), 143 (**2**), 163 (**2**), 169 (**2**), 311 (**2**), 476 (**2**), 486 (**2**), 504 (**2**), 619 (**2**)
surface-mounted device (SMD) 595 (**2**), 602 (**2**), 618 (**2**)
switchable optical add/drop multiplexer (SOADM) 121 (**1**)
system camera (CCD) 331 (**2**)
system decomposition 31 (**2**), 34 (**2**)

Taguchi approach 251 (**1**)
Taguchi method 250 (**1**)
tape automated bonding (TAB) 506 (**1**), 72 (**2**), 75 (**2**), 80 (**2**)

– technology 506 (**1**)
target transmission 130 (**1**)
temperature coefficient of resistance (TCR) 541 (**2**), 554 (**2**)
temperature cycling 6 (**1**), 289 (**1**), 388 (**1**), 389 (**1**), 430 (**1**), 454 (**1**), 460 (**1**), 8 (**2**), 124 (**2**), 129 (**2**), 132 (**2**), 185 (**2**), 192 (**2**), 348 (**2**), 394 (**2**), 397 (**2**), 560 (**2**), 563 (**2**), 604 (**2**), 620 (**2**), 698 (**2**), 703 (**2**)
tensile creep compliance 25 (**1**)
tensile failure 386 (**2**)
tensile properties 435 (**1**)
tensile strain 426 (**2**), 502 (**2**)
tensile strength 156 (**1**), 174 (**1**), 175 (**1**), 185 (**1**), 419 (**1**), 96 (**2**), 386 (**2**), 403 (**2**), 419 (**2**)
tensile stress 271 (**1**), 355 (**2**), 357 (**2**), 425 (**2**), 506 (**2**), 691 (**2**)
tensile test program 434 (**1**)
tensile testing 21 (**1**), 157 (**1**)
termination metallurgy 384 (**1**)
ternary eutectic reaction 355 (**1**)
test assemblies 316 (**1**)
test results 169 (**1**), 306 (**1**)
test vehicle (TV) 298 (**1**)
testability 206 (**2**)
testing time 315 (**1**)
tetraethoxysilane (TEOS) 192 (**1**)
theoretical maximum stress 20 (**1**)
theory-of-elasticity method 5 (**2**)
theory-of-elasticity treatment 5 (**2**)
thermal aging 93 (**2**), 114 (**2**), 262 (**2**), 271 (**2**), 293 (**2**), 445 (**2**), 453 (**2**), 461 (**2**)
thermal capacitance 630 (**2**), 632 (**2**), 639 (**2**), 643 (**2**), 645 (**2**)
thermal coefficient of expansion (TCE) 384 (**2**), 386 (**2**), 388 (**2**), *see also* coefficient of thermal expansion (CTE)
thermal conductivity 84 (**1**), 186 (**1**), 212 (**1**), 519 (**1**), 97 (**2**), 389 (**2**), 392 (**2**), 496 (**2**), 603 (**2**), 623 (**2**), 639 (**2**)
thermal CVD 187 (**1**)
thermal cycle (TC) 293 (**1**), 306 (**1**), 449 (**1**), 454 (**1**), 489 (**1**), 357 (**2**), 691 (**2**), 701 (**2**)
– profiles 292 (**1**)
– range 297 (**1**)
– test 297 (**1**), 305 (**1**)
thermal cycling 4 (**1**), 5 (**1**), 186 (**1**), 286 (**1**), 291 (**1**), 299 (**1**)–302 (**1**), 348 (**1**), 555 (**1**), 87 (**2**), 139 (**2**), 192 (**2**), 215 (**2**), 220 (**2**), 242 (**2**), 255 (**2**), 273 (**2**), 274 (**2**), 386 (**2**), 488 (**2**), 510 (**2**), 519 (**2**), 539 (**2**), 542 (**2**), 557 (**2**), 604 (**2**), 618 (**2**), 698 (**2**), *see also* temperature cycling
– loading 450 (**1**)
– reliability 430 (**1**), 255 (**2**), 261 (**2**), 269 (**2**)
– – test 430 (**1**)
– test 316 (**1**), 431 (**1**), 268 (**2**)
thermal deformation 272 (**1**), 400 (**1**), 406 (**1**), 459 (**1**), 470 (**1**), 486 (**1**), 494 (**1**)
thermal effect 636 (**1**)
thermal expansion 4 (**1**), 34 (**1**), 42 (**1**), 65 (**1**), 90 (**1**), 155 (**1**), 164 (**1**), 186 (**1**), 198 (**1**), 222 (**1**), 270 (**1**), 300 (**1**), 306 (**1**), 481 (**1**), 7 (**2**), 14 (**2**), 59 (**2**), 63 (**2**), 87 (**2**), 473 (**2**), 572 (**2**), 604 (**2**)
– coefficient 82 (**1**), 137 (**1**), 145 (**1**), 155 (**1**), *see also* coefficient of thermal expansion (CTE)
thermal fatigue strength 411 (**1**)
thermal fatigue test 413 (**1**)
thermal field 206 (**1**), 657 (**1**), 661 (**1**)
– assumption 633 (**1**), 638 (**1**)
thermal fluctuations 577 (**1**)
thermal gradient 475 (**1**), 215 (**2**), 393 (**2**), 396 (**2**), 398 (**2**)
thermal load 544 (**1**), 642 (**1**), 658 (**1**), 661 (**1**), 3 (**2**), 242 (**2**), 695 (**2**)
thermal loading 270 (**1**), 389 (**1**), 516 (**1**), 233 (**2**), 651 (**2**), 668 (**2**)
thermal mechanical fatigue life assessment 405 (**1**)
thermal mismatch 6 (**1**), 215 (**1**), 389 (**1**), 398 (**1**), 403 (**1**), 405 (**1**), 686 (**1**), 244 (**2**), 306 (**2**), 410 (**2**)
thermal modeling 321 (**1**)
thermal properties 186 (**1**), 321 (**1**), 398 (**1**), 400 (**1**), 387 (**2**), 503 (**2**), 553 (**2**)
thermal resistance 291 (**1**), 602 (**2**), 604 (**2**), 630 (**2**), 632 (**2**), 637 (**2**), 639 (**2**), 643 (**2**), 644 (**2**), 649 (**2**)
thermal resonance 645 (**1**)
thermal response 643 (**1**), 644 (**1**)
thermal shock (TS) 213 (**1**), 291 (**1**), 299 (**1**), 337 (**1**), 383 (**1**), 449 (**1**), 454 (**1**), 215 (**2**), 396 (**2**), 698 (**2**)
– testing 337 (**1**)
thermal simulation 242 (**1**), 321 (**1**), 637 (**2**)
thermal strain 389 (**1**), 401 (**1**), 405 (**1**), 406 (**1**), 499 (**1**), 281 (**2**), 303 (**2**)
thermal stress 155 (**1**), 198 (**1**), 3 (**2**), 12 (**2**), 14 (**2**), 262 (**2**), 273 (**2**), 303 (**2**), 473 (**2**), 493 (**2**), 680 (**2**), 685 (**2**), 698 (**2**), 701 (**2**), 706 (**2**)

– modeling 7 (**2**)
thermal stress-induced voiding 155 (**1**)
thermal-mismatch strain 8 (**2**)
thermally matched assemblies 11 (**2**)
thermo-compression bonding 506 (**1**), 313 (**2**)
thermo-elastic 215 (**1**)
thermo-mechanical reliability 205 (**1**)
thermo-optic coefficient 65 (**1**)
thermo-optic effect 81 (**1**)
thermo-optical analysis 89 (**1**)
thermo-optical model 66 (**1**), 70 (**1**), 80 (**1**)
thermocompression bonding 75 (**2**), 81 (**2**)
thermodynamic 324 (**1**)
thermoelectric cooler (TEC) 469 (**1**), 483 (**1**)
thermomechanical properties 388 (**2**)
thermomechanical reliability 268 (**2**)
thermomechanical shock test 315 (**1**)
thermomechanical testing 317 (**1**)
thermoplastic polymers 4 (**1**)
thermosetting polymers 4 (**1**), 5 (**1**)
thermosonic bonding 81 (**2**), 85 (**2**)
thick film 74 (**2**), 185 (**2**)
thickness variation (TV) 347 (**2**)
thin film 135 (**1**), 479 (**1**), 7 (**2**), 12 (**2**), 73 (**2**), 101 (**2**), 127 (**2**), 137 (**2**), 141 (**2**), 149 (**2**), 309 (**2**), 316 (**2**), 347 (**2**), 404 (**2**), 405 (**2**), 426 (**2**), 496 (**2**), 677 (**2**), 683 (**2**), 708 (**2**)
– structure 142 (**1**), 156 (**1**), 175 (**1**), 7 (**2**)
thin isotropic laminate 630 (**1**), 632 (**1**)
thin small outline package (TSOP) 285 (**1**), 14 (**2**), 272 (**2**)
thin-profile fine-pitch BGA (TFBGA) 270 (**2**)
thixotropic index (TI) 386 (**2**), 389 (**2**)
three-part-structure (TPS) 147 (**2**)
– probe 147 (**2**)
time dependent creep strain 430 (**1**)
time pressure dispensing 516 (**2**)
time–temperature superposition principle 23 (**1**), 32 (**1**)
time-dependent modeling 4 (**1**)
time-dependent thermal field 638 (**1**)
time-to-failure (TTF) 319 (**1**), 587 (**1**), 588 (**1**), 595 (**1**), 597 (**1**), 600 (**1**), 611 (**1**), 618 (**1**), 623 (**1**), 224 (**2**), 527 (**2**), *see also* cycles-to-failure (CTF)
time-to-market 3 (**1**), 205 (**1**), 315 (**1**), 454 (**1**), 203 (**2**), 208 (**2**), 253 (**2**), 254 (**2**), 399 (**2**)
tin whiskering 384 (**1**)
tin whiskers 385 (**1**), 386 (**1**), 267 (**2**)
tin-lead solder 352 (**1**), 440 (**1**), 263 (**2**), 527 (**2**)
tiny thermo-mechanical displacements 6 (**2**)
tire pressure monitoring system (TPMS) 300 (**2**)
total deflection function 559 (**1**)
total internal reflection (TIR) 367 (**2**)
total life approach 431 (**1**)
total mass loss (TML) 511 (**2**)
total shear strain range fatigue model 432 (**1**)
total stress intensity factor 530 (**1**)
toughenners 392 (**2**)
trace fracture 285 (**2**)
transfer matrix method (TMM) 70 (**1**), 82 (**1**)
transient in-plane thermal field 661 (**1**)
transient moisture distribution 438 (**2**)
transient thermal field 633 (**1**), 660 (**1**)
– effect 638 (**1**)
transient transverse thermal field 642 (**1**)
transistor outline (TO) 489 (**2**)
transition probability distribution 581 (**1**)
transmission dispersion 130 (**1**), 131 (**1**)
transmission electron microscopy (TEM) 343 (**2**)
transmission isolation 130 (**1**)
transmission loss 271 (**1**), 272 (**1**), 684 (**1**), 4 (**2**), 15 (**2**), 546 (**2**)
transmission speed 126 (**1**)
transmitter 362 (**2**), 364 (**2**), 368 (**2**), 487 (**2**)
transverse optical (TO) 350 (**2**)
tri-material assembly 58 (**2**), 493 (**2**)
tri-material model 651 (**2**)
true Bragg wavelength shift 98 (**1**)
true interface fatigue mode 415 (**1**)
tunable laser 119 (**1**), 4 (**2**)
twist testing 281 (**2**)
twist-off technique 8 (**2**)
two-region power-law model 598 (**1**)
two-wave mixing (TWM) 112 (**1**)
Twyman-Green interferometry 475 (**1**), 505 (**1**), 679 (**2**), 694 (**2**)

U-groove 515 (**2**)
U displacement field 466 (**1**), 477 (**1**), 494 (**1**)
ultimate tensile stress (UTS) 156 (**1**), 436 (**1**), 437 (**1**), 439 (**1**)
ultra large scale integrated (ULSI) 137 (**1**), 141 (**1**), 143 (**1**), 191 (**1**), 196 (**1**)
ultra short reach (USR) 364 (**2**)
– interconnects 364 (**2**)
– optical interconnects 366 (**2**)
ultra violet (UV) 371 (**2**), 501 (**2**), 521 (**2**), 565 (**2**)
ultra-low expansion (ULE) 497 (**1**), 501 (**1**)
ultrasonic bonding 83 (**2**)

ultrasonic energy 81 (**2**), 125 (**2**)
uncoupled axial vibrations 673 (**1**)
under bump metalization (UBM) 143 (**1**), 683 (**2**)
underfill 6 (**1**), 34 (**1**), 35 (**1**), 37 (**1**), 220 (**1**), 450 (**1**)–453 (**1**), 481 (**1**), 501 (**1**), 502 (**1**), 516 (**1**), 432 (**2**), 532 (**2**), 621 (**2**), 625 (**2**), 651 (**2**), 662 (**2**), 666 (**2**), 668 (**2**), 671 (**2**), 673 (**2**), 674 (**2**), 685 (**2**), 688 (**2**), 698 (**2**)
– flow modeling 532 (**2**)
– polymers 6 (**1**)
uniform curing 113 (**1**), 118 (**1**)
universal testing machine 434 (**1**), 604 (**1**)
US Military Standards (MIL-STDs) 206 (**2**), 207 (**2**)
UV-ozone cleaning 104 (**2**)

V-groove 152 (**2**), 164 (**2**), 384 (**2**), 394 (**2**), 496 (**2**), 506 (**2**)
V-number 27 (**2**)
vapor deposition techniques 385 (**1**)
variable order boundary element method (VOBEM) 536 (**1**), 546 (**1**), 550 (**1**)
variable strain rate shear test 391 (**1**)
VCR during test 465 (**1**)
V displacement field 466 (**1**), 477 (**1**), 494 (**1**)
vector-loop-equations (VLEs) 328 (**2**)
verification tests 538 (**1**)
vertical cavity surface emitting laser (VCSEL) 139 (**1**), 37 (**2**), 63 (**2**), 67 (**2**), 223 (**2**), 363 (**2**), 370 (**2**), 380 (**2**), 489 (**2**)
– transreceiver 23 (**2**)
vertical geometry 191 (**1**)
very large scale integrated (VLSI) 137 (**1**), 190 (**2**), 703 (**2**)
very short reach (VSR) 364 (**2**)
– interconnects 364 (**2**), 380 (**2**)
very-thin-profile fine-pitch BGA (VFBGA) 270 (**2**)
vibration loading 454 (**1**)
virtual master grating 462 (**1**)
virtual prototyping 242 (**1**)
virtual thermo-mechanical prototyping method 205 (**1**)
visco-plastic phenomena 4 (**2**)
viscoelastic models 52 (**2**)
viscoplastic analysis 430 (**1**), 701 (**2**)
viscoplastic solid 698 (**2**)
viscoplastic strain 430 (**1**), 702 (**2**)
– energy density 703 (**2**)
visual inspection 305 (**1**)
volumetric creep compliance 15 (**1**), 25 (**1**)
von-Karman's equations 13 (**2**)

wafer 140 (**1**), 146 (**1**)–148 (**1**), 150 (**1**), 157 (**1**), 164 (**1**)–167 (**1**), 187 (**1**), 188 (**1**), 192 (**1**), 198 (**1**), 261 (**1**), 287 (**1**), 318 (**1**), 347 (**1**), 385 (**1**), 627 (**1**), 111 (**2**), 135 (**2**), 142 (**2**), 153 (**2**), 294 (**2**), 302 (**2**), 343 (**2**), 363 (**2**), 364 (**2**), 395 (**2**), 413 (**2**), 625 (**2**), 652 (**2**), 677 (**2**), 678 (**2**), 706 (**2**)
– curvature technique 164 (**1**)
– level CSP (WLCSP) 136 (**2**), 145 (**2**)
– level packaging (WLP) 135 (**2**), 149 (**2**)
– level test 145 (**2**), 149 (**2**)
– level underfill 141 (**2**)–144 (**2**)
– on wafer (WOW) 139 (**2**)
wafer-level capping 302 (**2**), 320 (**2**)
wafer-level packaging 299 (**2**), 320 (**2**)
warpage 6 (**1**), 398 (**1**), 476 (**1**), 508 (**1**), 509 (**1**), 514 (**1**), 305 (**2**), 309 (**2**), 652 (**2**), 661 (**2**)–663 (**2**), 666 (**2**), 670 (**2**), 672 (**2**)
– analysis 505 (**1**)
– control 5 (**1**)
wave soldering 316 (**1**), 383 (**1**), 185 (**2**)
– process 381 (**1**), 429 (**1**)
waveguide 65 (**1**), 67 (**1**), 74 (**1**), 76 (**1**), 117 (**1**), 121 (**1**), 122 (**1**), 25 (**2**), 313 (**2**), 367 (**2**), 376 (**2**), 380 (**2**), 521 (**2**)
wavelength division multiplexing (WDM) 68 (**1**), 114 (**1**), 126 (**1**)
– filters 132 (**1**)
– systems 114 (**1**)
wavelength selective 2×2 switch 117 (**1**)
WB-PBGA package 511 (**1**)
Weibull distribution 217 (**1**), 218 (**1**), 319 (**1**), 337 (**1**), 342 (**1**), 388 (**1**), 455 (**1**), 571 (**1**), 577 (**1**), 183 (**2**), 269 (**2**)
Weibull failure rate distribution 183 (**2**)
Weibull life distribution model 455 (**1**)
Weibull modulus 604 (**1**)
Weibull parameter 299 (**1**), 319 (**1**), 597 (**1**)
Weibull plot 337 (**1**), 341 (**1**), 577 (**1**), 578 (**1**), 607 (**1**), 608 (**1**), 610 (**1**), 183 (**2**)
Weibull reliability analysis 318 (**1**)
Weibull shape 319 (**1**), 320 (**1**), 183 (**2**)
wettability 360 (**1**), 361 (**1**), 378 (**1**), 382 (**1**), 384 (**1**), 411 (**1**), 458 (**2**)
wetting angle 360 (**1**), 361 (**1**)
wetting balance curve 362 (**1**)
wetting balance test 361 (**1**)
wetting curve 361 (**1**)
whole-field deformation 680 (**2**)

wide area vertical expansion (WAVE) 139 (**2**)
Wilcoxon Rank-Sum test 319 (**1**)
wire 146 (**1**), 180 (**1**), 183 (**1**), 253 (**1**), 283 (**1**), 293 (**1**), 387 (**1**), 486 (**1**), 508 (**1**), 557 (**1**), 564 (**1**), 566 (**1**)
– array (WA) 558 (**1**), 567 (**1**), 474 (**2**), 475 (**2**), 479 (**2**)
– bond 293 (**1**), 511 (**1**), 121 (**2**), 124 (**2**), 185 (**2**), 276 (**2**)
– bonding 486 (**1**), 294 (**2**)
– strength 108 (**2**)
wire-bond plastic ball grid array (WB-PBGA) 486 (**1**)
wirebond failure 94 (**2**), 105 (**2**)
wirebond pull strength 102 (**2**), 104 (**2**), 185 (**2**)
wirebond reliability 87 (**2**)
wirebond testing 89 (**2**)
wirebonding 72 (**2**), 79 (**2**), 111 (**2**), 116 (**2**), 139 (**2**), 606 (**2**)
– machine 110 (**2**)
X-ray diffraction (XRD) 354 (**2**)
X-ray inspection 305 (**1**), 321 (**1**), 282 (**2**)
X-ray photoelectron spectroscopy (XPS) 460 (**2**), 468 (**2**), 541 (**2**)
X-ray reflectometry method 146 (**1**)

yield strain 412 (**2**), 422 (**2**)
yield strength 156 (**1**), 335 (**1**)
yield stress 148 (**1**), 253 (**1**), 333 (**1**), 435 (**1**), 439 (**1**)
Young's modulus 17 (**1**), 23 (**1**), 147 (**1**), 156 (**1**), 159 (**1**), 167 (**1**), 185 (**1**), 198 (**1**), 212 (**1**), 213 (**1**), 253 (**1**), 269 (**1**), 398 (**1**), 501 (**1**), 524 (**1**), 558 (**1**), 562 (**1**), 635 (**1**), 7 (**2**), 11 (**2**), 12 (**2**), 14 (**2**), 32 (**2**), 38 (**2**), 39 (**2**), 241 (**2**), 248 (**2**), 260 (**2**), 374 (**2**), 409 (**2**), 413 (**2**), 421 (**2**), 422 (**2**), 475 (**2**), 476 (**2**), 510 (**2**), 547 (**2**), 549 (**2**), 668 (**2**), 671 (**2**), 687 (**2**)

zero stress condition 394 (**2**)
zero stress temperature 396 (**2**), 402 (**2**)