Benefiting from Thermal and Mechanical Simulation in Micro-Electronics

Edited by
G.Q. Zhang
L.J. Ernst
O. de Saint Leger

Springer-Science+Business Media, B.V.

BENEFITING FROM THERMAL AND MECHANICAL SIMULATION IN MICRO-ELECTRONICS

المنارة للاستشارات

ww

Benefiting from Thermal and Mechanical Simulation in Micro-Electronics

Edited by

G.Q. Zhang

Philips CFT, Eindhoven, The Netherlands

L.J. Ernst

Delft University of Technology, Delft, The Netherlands

and

O. de Saint Leger

COMPETE Network Co-ordinator at MTA, Paris, France

SPRINGER-SCIENCE+BUSINESS MEDIA, B.V.

المنارة للاستشارات

ww

A C.I.P. Catalogue record for this book is available from the Library of Congress.

ISBN 978-1-4419-4873-1 ISBN 978-1-4757-3159-0 (eBook)
DOI 10.1007/978-1-4757-3159-0

EUROSIME 2000 was mainly sponsored by the European Commission

Printed on acid-free paper

All Rights Reserved
© 2000 Springer Science+Business Media Dordrecht
Originally published by Kluwer Academic Publishers in 2000
Softcover reprint of the hardcover 1st edition 2000

No part of the material protected by this copyright notice may be reproduced or
utilized in any form or by any means, electronic or mechanical,
including photocopying, recording or by any information storage and
retrieval system, without written permission from the copyright owner.

المنارة للاستشارات

Organising committee

Chair: Mr. O. de Saint Leger, MTA, France

Secretariat: Dr. G.Q. Zhang, CFT/Philips, The Netherlands

Dr. R. Dudek, IZM, Germany

Dr. L. J. Ernst, Technical University of Delft, The Netherlands

Dr. M. Meuwissen, TNO, The Netherlands

Dr. S.H. Pulko , Hull University, UK

Ms. O.F. Slattery, NMRC, Ireland

Ir. B. Vandevelde, IMEC, Belgium

Dr. I. de Wolf, IMEC, Belgium

المنارة للاستشارات

ww

Contents

Preface ix

Simulation overview in the industry
B. Schwarz 1

Thermal & mechanical problems in microelectronics
O. F. Slattery G. Kelly and J. Greer 17

Solder material characterization and modelling
S. Wiese F. Feustel and E. Meusel 27

Polymer material characterisation and modelling
L. J. Ernst 37

Generic issues in numerical modeling
S. Pulko 59

Modeling of vapor pressure during reflow for electronic packages
X. J. Fan 75

Simulation for fatigue, cracks and delamination
R. Dudek, J. Auersperg and B. Michel 93

Experimental validation of finite element modeling
D. Vogel, C. Jian and I. De Wolf 113

Perspectives of non-linear simulation
M. A. Crisfield and A.J. Burton 135

Simulation-based optimisation in virtual thermo-mechanical prototyping of electronic packages
G.Q. Zhang and H.P. Stehouwer 151

Thermal fatigue reliability optimisation of Flip-Chip assemblies
B. Vandevelde and E. Beyne 165

Product and Process Optimization with simulation
D. den Hertog and P. Stehouwer 179

المنارة للاستشارات

ww

PREFACE

One of the technology trends in the electronics industry is the rapid development and ever-increasing application of advanced thermal & mechanical simulation methodologies and tools. For the leading companies, thermal & mechanical simulation is becoming a daily practice, a standard tool, and an important competitive edge to achieve competitive product and/or process development, with significant impact on the business profitability of electronics industry. On the other hand, the fascinating development of the electronics industry has also formulated many new challenges, impulse and finance possibility for the further development of mechanical and thermal simulation methodologies and tools.

As the first international conference specially dedicated to mechanical and thermal simulations in (micro)-electronics, EuroSimE2000, was held at Eindhoven, The Netherlands, March 23-24. It was initiated and organised by the thematic network COMPETE(www.compete.tm.fr)/EU with major sponsoring from the European Community, aimed to:

- promote further development and application of mechanical and thermal simulation methodologies and tools for the electronics industry
- disseminate competence and results obtained from relevant research projects
- improve communication and exchange information between methodology & tool developers and industry users
- strengthen co-operation between industry, universities, and research institutes.

It brought together about 125 participants from 13 countries, including project-, product- and IT managers, reliability specialists, thermal & mechanical analysts, (micro-) electronic design engineers and other thermal and mechanical professionals. There were 4 sessions with 19 presentations in total. All lecturers were invited and can be considered as leading authorities in their fields of specialization. The first session concerned “The impact of simulation on industry profitability”. The second session handled “Approaches to simulation”. The third session was on “The state-of-the-art methodologies of simulation”. The last session focused on “Design optimization by simulation”.

This conference revealed that

- the state-of-the-art methodologies and tools for thermal and mechanical simulation and optimization are receiving more and more applications in microelectronics and generating substantial business profitability. However, many new challenges and bottlenecks for further application of simulation and optimization methods emerge mainly driven by the rapid development of microelectronics

G. Q. Zhang et al. (eds.), Benefiting from Thermal and Mechanical Simulation in Micro-electronics, ix–x.
© 2000 *Kluwer Academic Publishers.*

ww

- more attention and effort are needed from methodology & tool developers to speed up the development of simulation and optimization methodologies and tools for micro-electronics application, by combining the expertise and strengthens from both electronics industry and simulation & optimization societies.

G.Q. Zhang, L.J. Ernst and O. de Saint Leger

المنارة للاستشارات

ww

SIMULATION IN THE INDUSTRY

by

Dipl. Ing.
Bernd Schwarz

Dr. Hans Jürgen Albrecht
Dipl. Ing. Gunnar Petzold
Dipl. Ing. Jörg Deliga
Dipl. Ing. Christian Wegener

Siemens AG
ZT MF 6
Siemensdamm 50
13 629 Berlin

+49 (0)30 / 386 – 26872
Fax: - 26843
Bernd.Schwarz@blns.siemens.de

Contents

1. Self-Introduction
2. Overview: Simulation in the Industry
3. Examples of Practical Work
4. Accelerated Life Time Tests with Demonstrator Boards versus Simulation
5. Reasons for more investigations in simulation (models)
6. Appendix: Introduction and Operation of TherMoiré-System- a warpage measurement system for surfaces (PCB, Components and other Materials) under Temperature Loading (e.g. Reflow soldering process)

G. Q. Zhang et al. (eds.), Benefiting from Thermal and Mechanical Simulation in Micro-electronics, 1–16.
© 2000 *Kluwer Academic Publishers.*

المنارة للاستشارات

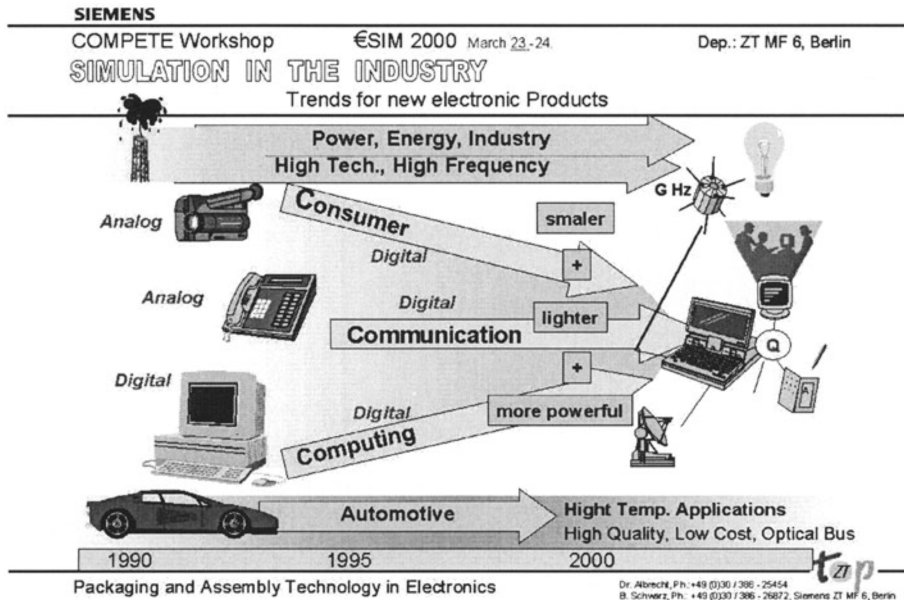
ww

1. Self - Introduction

ZT MF6 Berlin is a Central Department for Corporate Technology and a Center of Assembly Technology for Electronics in the Siemens Company. We are supporting many Siemens divisions, who are manufacturing assembly boards (PCB). Especially for the new technologies we give support to ensure the product qualities and we help to find out critical manufacturing parameters and to optimize and to qualify components, boards, assembly materials and the manufacturing parameters.

2. Overview: Simulation in the Industry

The development of electronic technology is rather fast. The number of assembled components and manufactured boards are double each 2 until 3 years.



In dependence of the products, different assembly technologies are used, because the demands of the product sectors are different.

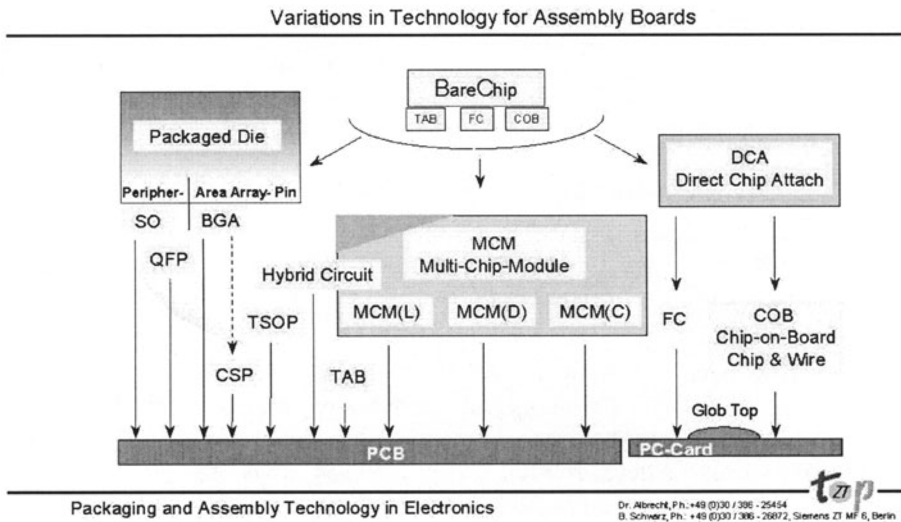
* Some assembly boards need special cooling technologies - e.g. product fields (sectors) where power and high frequency products are manufactured
* others have to work under very hard surrounding conditions - e.g. the automotive products at the engine and under the hood of a car.

The three sectors Consumer, Communication and Computer are growing together more and more. We have a trend to digitally realized product ideas. One common trend is that the structures are getting smaller and smaller and

المنارة للاستشارات
ww

the clock rates (frequencies) of the digitally realized products are increasing very fast. So the size reduction (miniaturisation) is not alone driven by the wishes of the customers, to have small electronic pocket assistants but also to get the performance. High clock rates need short pathways.

Because of this different demands to realize a electronic product, there is a wide range of possibilities to make from the wafer and the single Bare Chip a printed circuit board (PCB).



* On the left side of the picture you see the conventional way to redistribute the pins of the die (Bare Chip) and arrange the I/O's to peripheral leads or area array solder balls. Those packaged components show a trend in direction to chip size (CSP).
* The right side of the picture shoes the direct chip attachment – mostly used for PCMCIA-Cards or Plastic-Cards but also for Multi Chip (Package) Modules or sub-boards.
Those MCM solutions – shown in the middle of the diagram - can be overmolded or encapsulated in an HF-shielding-cup. The substrate materials can be Laminates (e.g. FR4), Deposits or a Ceramic.

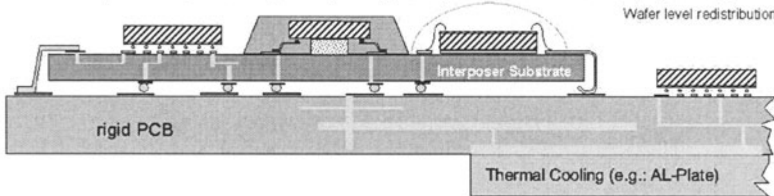
If we take a closer look to the shown variants, we see, that there exist much more differences:
The Die (Chip) can be attached face up - connected by wire bonds - or face down connected by solder balls (Flip Chip FC) or by metal tapes. The

المنارة للاستشارات
ww

spectrum of board assembly technologies and all the different I/O pathways are shown in this schematic picture.

Assembly Board Technologies:

- **Electronic Components:**
  QFP (G-leaded, J-leaded), FCIP, BGA, µBGA, CSP, MCM with COB(IP), COB , FCOB



- **Interposers:**
  rigid, flex, ceramic
- **Boards:**
  FR4, SBU with Liquid Epoxy or RCC (Polyimid-Foil), High Tg Material
- **Assembly Materials:**
  solder (Pb?), Polymer-Package (rigid, Gob Top), Underfiller, ...

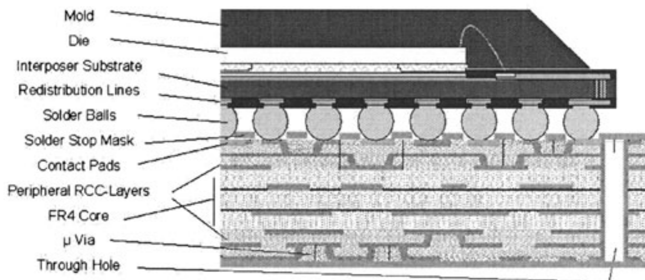
Packaging and Assembly Technology in Electronics

Dr. Albrecht, Ph.:+49 (0)30 / 386 - 25454
B. Schwarz, Ph.: +49 (0)30 / 386 - 26872, Siemens ZT MF 6, Berlin

The Materials of the interposer substrate and of the board, but also the geometry of all shown parts as well as the assembly materials are depending from supplier and manufacturer and most parameters are unknown for those who have to design a new product.
But to predict product quality and lifetime by simulation it is necessary to know this Simulation-Parameters

To get a more detailed view of the (confusing) all in one schema, the following picture shoes a drawing of a cross section of a BGA on a multilayer board.

High Density Interconnection Substrate - with "Chip&Wire"-CSP Schematic View



المنارة للاستشارات
ww

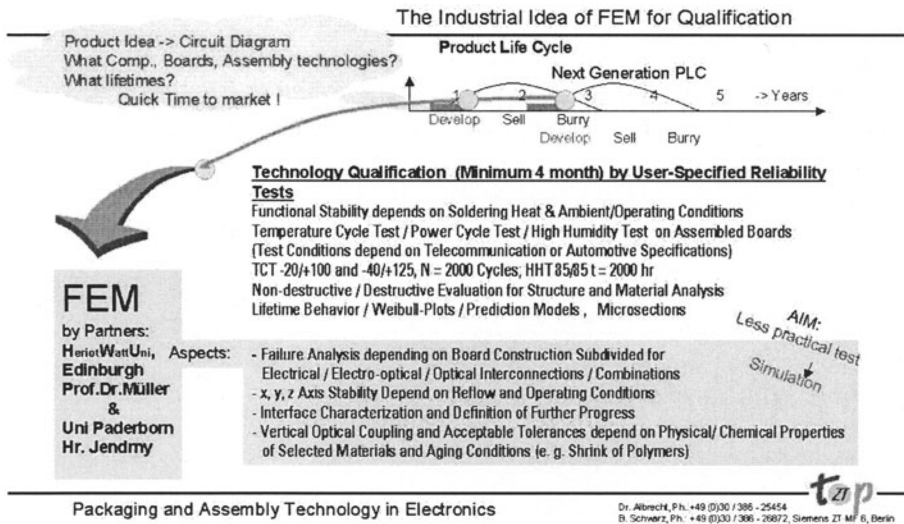
Boards today are nearly as complex and different as the assembled components. The Die (Chip) in this case here is attached face up and I/O-Pads are connected with bond wires to the rigid substrate interposer. This interposer redistributes the I/O-Lines to a Ball Grid Array.
SBU-boards are characterized by a multilayer kernel and additionally build up peripheral layers. The interconnections of layers are realized by Through Holes and µVia connections. No need to mention, that the fabrication technologies are vary and for the product manufacturer - the people in the production lines - those details mostly are unknown and hardly could be taken in consideration (have to be disregarded).

3. Examples of Practical Work

To help the production lines by there individual technology qualification, tests are necessary. The product life cycles are getting shorter and shorter and the standards for accelerated lifetime tests can not compete with the short times to market.
The aim is to support or even replace those practical tests by Simulation. Qualification tests for new components and technologies take at least 4 month. For accomplishing practical tests by Simulation and FEM a lot of parameters have to be taken in consideration.
The problem in Simulation is, to describe the wide field of assembly board technology variations and take in consideration all the possible Failure mechanisms.
Together with Prof. Dr. Müller from HWU Edinburgh and Mr. Jendrny from Uni Paderborn a lot of calculations have been made in the last ten years of cooperation. The results have been compared with the practically performed Temperature Cycle Tests (TCT), Power Cycle Tests (PCT) and Humidity & High Temperature Tests (HHT).

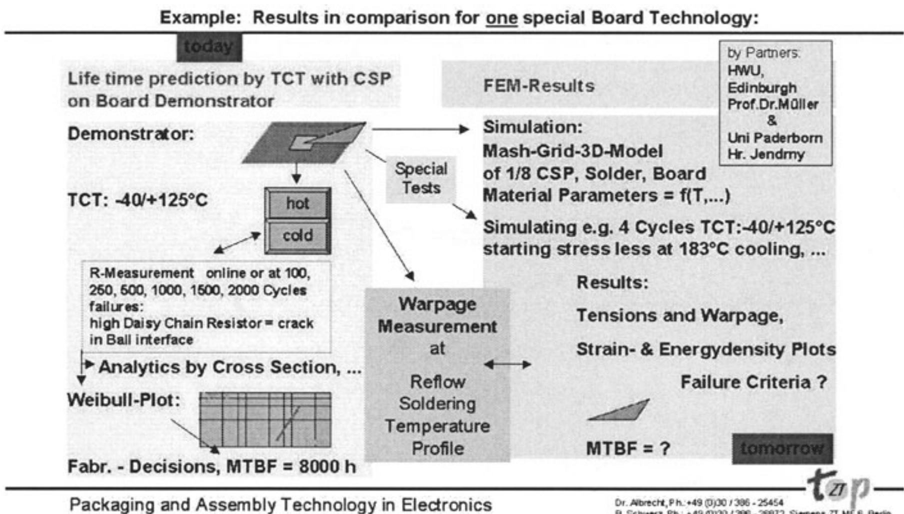
المنارة للاستشارات

ww



Simulation was started about 10 years ago with passiv components – the melf and chip size packages - and today we have sub-models even for the TH's and µVias in SBU boards.

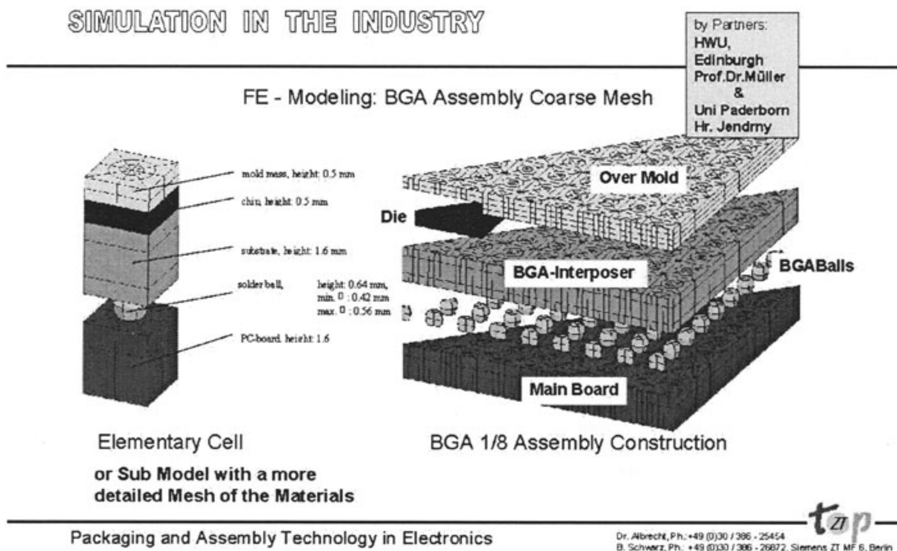
Siemens part was to perform the technological qualification on the traditional way by test vehicles under life time accelerating stress (left side of next diagram) and our University partners used the test vehicle parameters to feed ABACUS (right halfside).



المنارة للاستشارات

Results like the calculated warpage if a temperature stress cycle is performed (later also if a reflow temperature. process is performed) could be compared with the measurement results of our TherMoiré- Warpage measurement System. This system enables to determine the surface profile of DUT (BGA, boards, ...) for a programmed dynamic temperature profile. (system description later or will be placed in www)-
So it was noticed on the one hand, that calculations are right and suitable to practical measurements, but also sometimes different.

The following Picture shoes an example of our work: a BGA Coarse Mesh on a simplified homogeneous board. Today we have sub models for board details like µVias or TH's.



Details about modeling you can get directly from Prof. Dr. Müller, Harriot-Watt University Edinburgh (W.H.Muller@hw.ac.uk).

<u>FE-Simulation Tasks out of the View of SIEMENS & Partners:</u>

Collecting Material Data,

work out the characteristic parameters for the thermal mechanical and thermal dynamical behavior of Materials used for height density assembly boards and

realize FEM-Models to give recommendations for best materials and production technologies to use for high quality products or to predict, what happens with the lifetime, if more environment-friendly or cheaper materials are used or dimension shrinking for next generation products will be carried out.

المنارة للاستشارات
ww

To use Simulation in the Industry, the first main task is to collect sufficient and right material data.
Also it is very important to work out the characteristic parameters for the thermal mechanical and thermal dynamical behavior of materials used for high density assembly boards.
Realized FEM-Models has been compared to reality. So it is possible to prove the quality of the models:
Are they suitable to decide about material and geometric optimizations or changes (next generation products e.g. lead-less and more shrinking) or to predict acceleration factors and replace practical qualification tests?

4. Accelerated Life Time Tests with Demonstrator Boards versus Simulation

Comparing the results of practically performed stress tests and FEM simulation has shown, that the smaller the structures and the thinner the materials the more difficult is modeling, because the inter-metallic phases and the interacting connection regions do not behave as the simulation parameters for bulk material implement.

So there arise a lot of new tasks to improve the simulation models: the damage models, new materials, especially a better characterization of polymers, a model optimization in regard of failure mechanisms, the verification by demonstrator boards and the know-how dissemination.

المنارة للاستشارات

ww

SIMULATION IN THE INDUSTRY

The smaller the structures and the thinner the materials
the more difficult is the modeling of the reliability behavior of Assembly Boards.

Cause: Materials act not like their FEM-Parameters predict. For the inter metallic phases and the interacting connecting regions modeling parameters are missing often .

() Need of improved modeling,

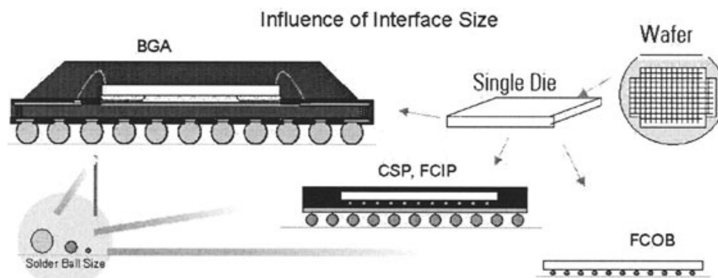
-> **Aims**
- virtual qualification
- selection guides for advanced packages and interconnection technologies
- Prediction for Acceleration Factors

-> **Tasks:**
1: **Development of appropriate damage test methodologies and failure criteria /** Comparison to Test Data Board Level Reliability
2: **Assessment of currently available Solder and adhesive material data and identification of data requirement for thermo-mechanical simulation / Life Time Prediction /** Acceleration Factor Test- / Field-Conditions
3: **Process dependent packaging polymer characterization**
4: **Modeling, optimization and verification w.r.t. thermo-mechanical failure mechanism**
5: Demonstrators based on mixed boards and/or high density solutions
6: **Formulation of methodologies for know-how dissemination**

Packaging and Assembly Technology in Electronics

Dr. Albrecht, Ph.:+49 (0)30 / 386 - 25454
B. Schwarz, Ph.: +49 (0)30 / 386 - 26872, Siemens ZT MF 6, Berlin

Week simulation items are:
- incomplete or wrong (unrealistic) material parameters,
- lack of know how about the processes in interconnection areas of different materials (inter-metallic Phases) and
- the corresponding simulation parameters (model lack),
- poor plastic (polymer) models,
- missing equations (transformation tools) to consider manufacturing parameters and environmental influences.
- no Database to handle all the parameters and disseminate them to all users.



The smaller the Solder Ball Diameter, the higher the Influence of Interfaces and Materials chosen for Component-Board Connection

This picture explains the influence of interface size. The trend in miniaturization of the component packages let the solder ball sizes decrease from 500µm for a BGA, to 200µm for a CSP and finally just 100µm for a

المنارة للاستشارات

FC-ball. The grain (corn) of the finest solder-past particles today are 15µm, so we just would have few volumes, higher tolerances by solder past printing and more interactions of the thin material volumes to there neighbor materials. Bulk material parameters are not suitable to make the correct thermal-mechanical lifetime simulations.

Also it should be mentioned, that we found out (measurements with TherMoiré-System, see appendix), that the humidity especially in the polymers has a large influence to the warpage behaviour during lifetime. Describing the used polymers (practical experiments) is a main task (requirement) to achieve realistic FEM results.

Example: CSP - FC -Testboard

CSP - FC - Testboard

Board (Technology) Specifications:

- FR 4: 220 x 290 x 1,6 mm (4-layer)
- 35 µm Cu-Metallization
- Pad Finish: Ni-Au
- Solder Mask: NSMD
- Layout for online and/or offline-Measurements
- Solder Paste: Sn62Pb36Ag2 (Powder Type 4), NC, Flux SMQ57, Metal Content: 89 %
- Solder Paste Thickness: 120 µm for CSP, 75 µm for FC
- Placement SIPLACE 80 F4 FS02
- Reflow (ERSA Hotflow 7), T-Profil
- Underfiller

"Some" Details:

What FR4, Tg
kind of SBU
fiber glass filling
fiber glass direction
humidity
state of curing
via technology (lasered, etched,...)
plating technique
kind of solder mask, adhesion
symmetrical layout
heat dissipation
temp. gradients
solder metallurgy
kind and extensions of intermetallic phases
stand off if soldered
warpage?
-> stress in connection
Flux pollution
influence to adhesion
geometrical tolerances
-> misplacement
underfiller adhesion
filled underfiller
settlement of fill material
completely filled
isotrop, ...

possible damage models

Packaging and Assembly Technology in Electronics

tZIp

Dr. Albrecht, Ph.: +49 (0)30 / 386 - 25454
B. Schwarz, Ph.: +49 (0)30 / 386 - 26872, Siemens ZT MF 6, Berlin

Above one of our test boards and a set of describing parameters (assembly technology parameters). There are much more details, which should be taken into account. The list is not complete but just should demonstrate the complexity of the task to achieve the explained aims by using simulation in the industry.

Customer needs are driving different PCB technologies:

المنارة للاستشارات
ww

Spectrum of Possibilities for a Demonstrator Assembly Board:

Die: active, passive (e.g. DC-Typ), attachment (Bond, Ball, Tape, ...)
Redistribution: Wafer-level, Interposer (f, r, c)
Package: Material
Board attachment: Ball Grid, LGA, Pitch, IO-No., solder-typ (Pb?)
Underfiller:
Board: Standard FR4, High-Tg, SBU (µVia-Geometry, Material, Platings,...)

Kind of Stress: HTS120°C, TCT-20/+100°C, TCT-40/+125°C, PCT, Vibration
Failure Model: Solder Joint Cracks, µVia-Cracks, Delamination
Test of Failure: R-Measurement (Value or go/nogo), ...
Type of Layout: DC-Channels, Logic on Board, I/O-Protection Circuits, Boundary Scan
Measurement: Online or Offline after e.g. 100, 250, 500, 1000, 1500, 2000 h or Cycles
Failure-Case-Procedure: take from Stress, Failure Analysis (x-Ray, Cross Section, ...)

Analytics at the components: Warpage measurement under reflow temp. profile
Collecting Model Parameters by asking supplier, literature and experiments

Packaging and Assembly Technology in Electronics

tzip

Dr. Albrecht, Ph.: +49 (0)30 / 386 - 25454
B. Schwarz, Ph.: +49 (0)30 / 386 - 26872, Siemens ZT MF 6, Berlin

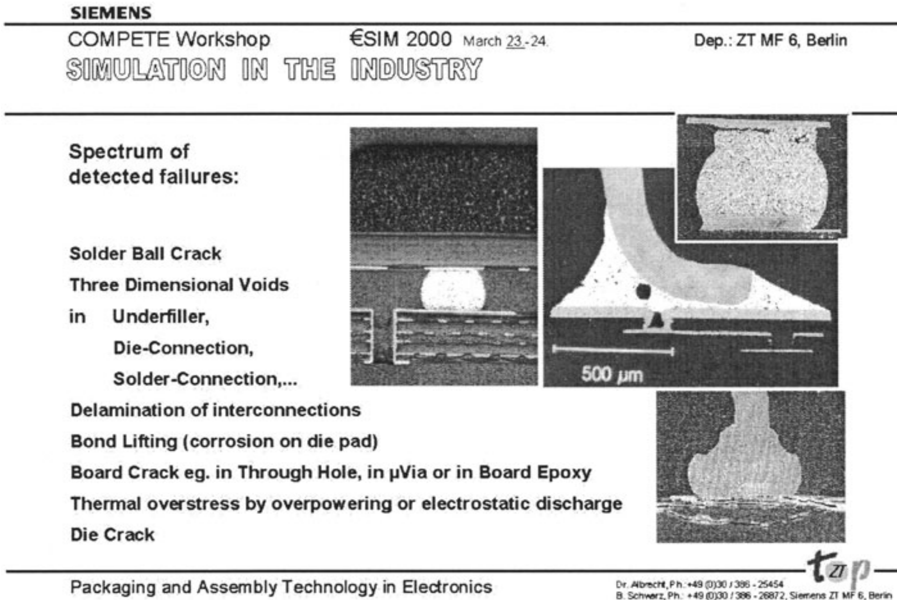
Here are the main technology variation parameters and an overview of the today typically used acceleration stress tests to predict product quality and product lifetime.

The measurement to detect fail mode during test is to control conductivity of a daisy chain test layout or the conductivity of the protection circuits, which exists at nearly each I/O of real (no Daisy Chain) IC's (components). With this test method we detect solder ball cracks, board line- or µVia-cracks and delamination which was so strong that electric connections opend.

It is very important to verify this failure mechanisms by failure analysis. So we have and we need x-ray, ultrasonic microscope, metallographic analysis (cross-sections), warpage measurement-system and temp. Measurement equipment to check, if the supposed failures really happened. Also it can be necessary to do material analysis with Spectroscopy or EDX included in our SEM.
The next picture shoes some examples of failures we detected after microsectioning.

المنارة للاستشارات

www



Cross-sections through the supposed failure area show e.g. Wire Bond Cracks or Si Cracks in the el. Structure of the die, solder ball cracks, voids in solder or delaminations in component packages, between board layers or at micro vias.

5. Reasons for more investigations in simulation (models)

It is obvious, that the practical experiments and also the modeling and simulation has to consider all three levels of assembly board manufacturing:

1. wafer level,
2. component level including the interposers for I/O redistribution and component packaging and
3. the second level connection, what means assembling the components to the product board (including underfiller, heat dissipation constructions and the stress during lifetime).

Today wide ranges of geometry, technology and material variations are used on the market for the mentioned three levels die, package and board. It is necessary to do the examinations separately and in detail but for the practical use in industry the behaviour and lifetime prediction for the whole product is of enormous importance. The interactions of the materials,

المنارة للاستشارات

material diffusions at the interfaces and the change of material properties makes it very difficult to make a realistic simulation.

Especially the boards are getting more and more complex, the structures are decreasing to enable the interconnection (wireing) of the small components. Experience and Examinations of the last years have shown, that the assembled printed circuit board failures are showing more and more failures in the boards. We have analyzed not only delaminations and conductor line cracks but also the lifting of µ-vias in the region of Area Array Components (Ball Grid Connections).
The Materials of the boards are polymers and epoxies and mostly the ingredients are unknown. So the Simulation parameters are unknown.

Those complex constructions, which are difficult to describe, we have twice:
- in the components in form of the interposers
- and as product board.

To have success in simulation it is important to have the right concept. What is neglectable and in what detail the modeling is necessary? To give an answer we should know already the results.

Especially the polymers and there visco-elastic behaviour are a problem. This materials are used in thousands of variations and even the component suppliers often don't know, how to describe, what they deliver.

As mentioned the Polymers are very different and we find them mainly on the 1st and 2nd packaging level of an assembly board. But also on wafer level it seems to be a matter of time until isolation layers for wafer level redistributions are made by polymers. What's about conductive polymers?
Not mentioned and considered are "Opto boards" (fiber lines between board layers).

Without to be an expert for polymer simulation it can be estimated, that the usually used simulation parameters for FEM are not sufficient.
By the way - also the simulation of metalisations and thin finishings are a problem, because simulation parameters from material databooks and literature cannot used and degradiations, dissolutions and diffusions to new alloys should be taken in the model.
For polymers it was observed, that the warpage of a PBGA is depending from it's humidity. So it's essential to consider the drypack instructions.

المنارة للاستشارات
ww

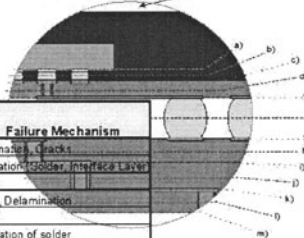
Last but not least the failures like delamination and initially cracks do not destroy the product (lifetime) at once but those processes during product lifetime are changing the simulation model.

The next table gives an overview of the locations and valences of occurring failures of the four main categories of component packaging (SMD, BGA, CSP, FC) and the today common Multilayer boards with sequential build up layers and partial micro contacts from layer to layer. How to simulate them?

SIMULATION IN THE INDUSTRY

Valence of Failure Layers:

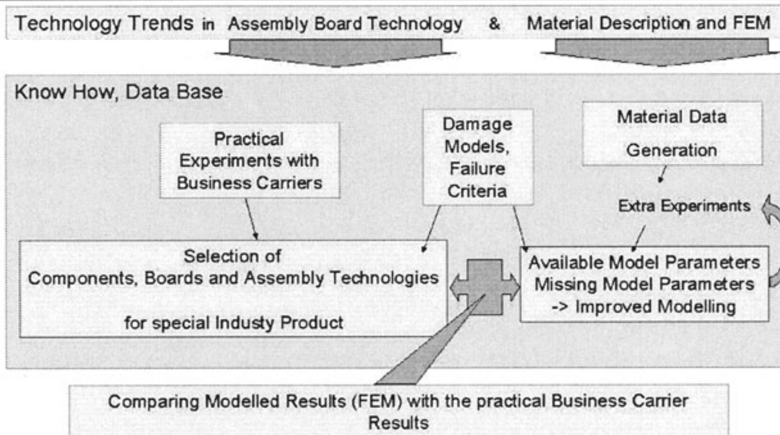
(Picture: FCIP on 2-layer Interposer assembled on 4-layer Board)



El relevant Failure Levels		Weightening				
Layer	Description	SMD	BGA	CSP	FC	Failure Mechanism
a	Struktures on Die IC (vert./lateral)	+	+	++	+++	Delamination, Cracks
b	1st Level Interconnection	+	+/++*	++/+++	+++	Degradation (Solder, Interface Layer)
c	Top Layer Interposer	-	+	+	/	Cracks
d	Interconnection Layer Interposer (Via)	-	+	++	/	Cracks, Delamination
e	Bottom Layer Interposer	-	+	+	/	Cracks
f	2nd Level Interconnection	+++	+++	+++	/	Degradation of solder
g	Top Layer Board	-	-	-/+	+	Cracks; Strukturstabilität
h	1. Interconnection Layer Board (Via)	+	+	+/++**	++	Cracks, Delamination
i	Mid-Layer 1 Board	-	-	-	-	Cracks
j	2. Interconnection Layer Board (Via)	+	+	+/++**	++	Cracks, Delamination
k	Mid-Layer 2 Board	-	-	-	-	Cracks
l	3. Interconnection Layer Board (Via)	+	+	+/++**	++	Cracks, Delamination
m	Bottom-Layer Board	-	-	-	-	Cracks

W (Weightening): - unrelevant; + relevant; * depending on Interconnection Typ; ** depending on manufacturing technology (standard, SBU)

The conclusion of all this facts is, that it is necessary to work together, to split the tasks and especially to create a common database about the material parameters and existing models.



المنارة للاستشارات
ww

On the one hand improvements of simulation in industry are necessary. FEM is not good enough to replace and accelerate practical reliability test. It is good for optimizing special materials with an established model.

On the other hand, it's very expensive and slow to use real testboards for the PCB Technology Qualification and the aim of the Industry is, to do it in simulation (virtual qualification).
New technology trends and materials should continually be added to this database and European or even worldwide Universities, Institutes and Companies should complement each other.

The idea is to push and realize those aims by an EU project called VITEMP.

All pictures published and documented in Lecture Notes SIM 2000, Eindhoven, March 23-24, 2000

6. Appendix:

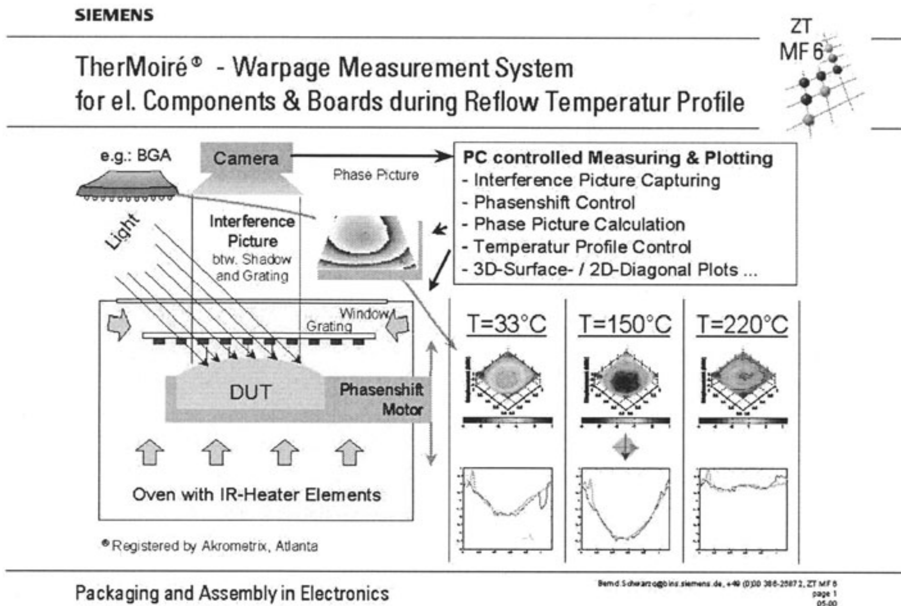
Introduction and Operation of TherMoiré-System – a warpage measurement system for surfaces (PCB, Components and other Materials) under Temperature Loading (e.g. Reflow soldering process)

Autumn 1998 – more than 2 years ago now – we got the first system prototyp from EPS – a company founded from Atlanta Georgia Tech Uni members. Today a lot of improvements are made, over ten systems are sold worldwide and the companies name changed to Akrometrix.

The picture explains working of TM-System and shows result examples for BGA.

المنارة للاستشارات

ww



From the 3D surface diagrams diagonal or cord plots can be extracted or typically values like bow, twist or center deflection of DUT (e.g. a BGA top and bottom side) in dependence of DUT-temperature can be derived.

How does it work:

Halogen light is illuminating the measurement surface of the DUT through the window of the oven chamber and through a 100 lpi grating (glass with lines on bottom side). The shadow of the grating in the surface of DUT and the grating lines itself produce an interference picture. The interference fringes are taken by camera and frame-grabber to PC. To determine Min and Max of the surface, phase-steps are performed and the set of interference pictures is calculated to a phase picture, containing all infos about topography of the DUT. Thermo Couples are measuring the surface temperature of DUT and make the temperature following the preset temperature profile. During the profile the measurement procedure of 2 seconds duration can be initializes as often as necessary. After measurement the phase-pictures are converted to 3D-surface plots and 2D-Diagonal- or Cord plots and the mentioned characteristic parameters can be calculated.

Temperature Range from RT to 240°C.
Maximum Measurement area is 200mm x 200mm
Resolution with 100 lpi grating is 2 µm, Accuracy ± 5 µm

المنارة للاستشارات

ww

Thermal and Mechanical Problems in Microelectronics

O.F.Slattery, G.Kelly and J.Greer
Computational Modelling Group, National Microelectronics Research Centre, Lee Maltings, Prospect Row, Cork, Ireland

Key words: Reliability, temperature, thermal management, thermomechanical stress, future challenges

Abstract: This paper aims to present a brief introduction to the topic of thermal and mechanical problems in microelectronics, specifically packaged devices. A description of typical reliability concerns at component and PCB level is given and the effects of temperature on the reliability of microelectronics devices are outlined. The origin of thermomechanically induced stresses is explained and finally, future technology requirements facing the assembly and packaging community are described and possible solutions are introduced.

1. INTRODUCTION

Management of thermal and mechanical issues in microelectronics is becoming an ever more important concern. Today, the thermal performance of electronics devices is being pushed to the limit by increasing heat fluxes, continued miniaturisation and higher speeds. An understanding of the impact of thermal issues on microelectronics reliability, as well as an ability to cater for increasing demands, is necessary for continued development.

From a thermomechanical viewpoint, electronics devices consist of a number of materials of differing mechanical properties assembled together during high temperature processes. The mismatch between the mechanical properties of the various materials coupled with high package processing

G. Q. Zhang et al. (eds.), Benefiting from Thermal and Mechanical Simulation in Micro-electronics, 17–26.
© 2000 *Kluwer Academic Publishers.*

المنارة للاستشارات

ww

temperatures leads to the introduction of thermomechanical stresses. Excessive levels of such stresses pose a serious reliability concern.

The subject of thermal and mechanical problems in microelectronics is vast and complex and cannot be discussed in detail in a format such as this. Thus, this paper aims only to provide a brief introduction to the topic of thermal and mechanical problems in microelectronics.

2. THERMAL AND MEHANICAL FAILURES IN MICROELECTRONICS

Figure 1 shows the primary electronics failures at component level. These figures include electrical failures as well as thermal and mechanical failures since many electrical failures are activated and accelerated by temperature. This is discussed in the next section. At the package level excessive thermomechanical stresses may lead to delaminations and package and passivation layer cracks. These provide paths for moisture ingress which may lead to corrosion [1]. Wirebond failures frequently occur as a result of problems with the molding process [2] but may also be a consequence of thermal cycling.

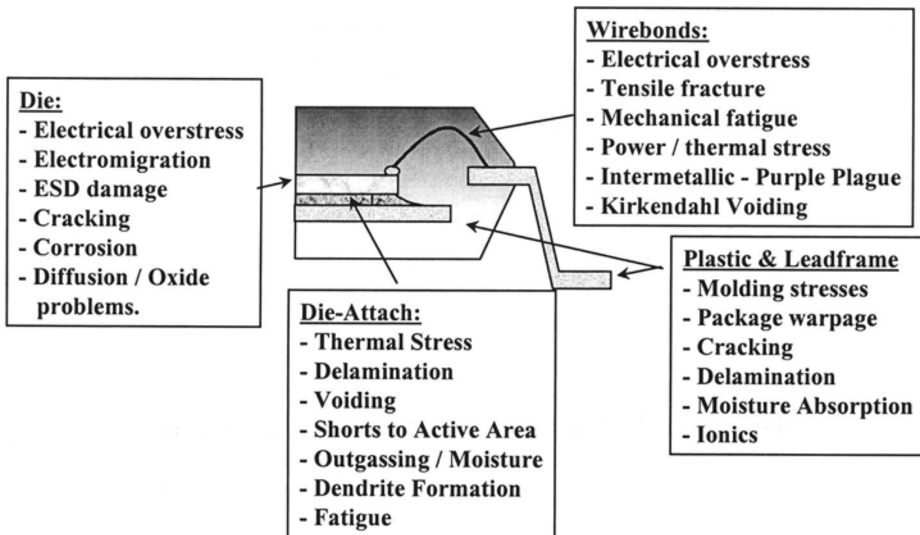


Figure 1. Component Level Failures

المنارة للاستشارات

ww

Die failures; such as electromigration occur at elevated temperatures under the influence of a temperature gradient. The metal thins and eventually separates completely causing an open circuit [3]. *Figure 2* shows the primary electronics failures at solder joint / PCB level. Under normal operating conditions mismatch between the coefficients of thermal expansion of a package and board mean that solder joints are subjected to cyclic stresses and strains which ultimately lead to the failure of the solder joints in fatigue [4]. Full and detailed descriptions of package and PCB /board level failures may be found in textbooks such as those referenced in this paper. [1, 3-7]

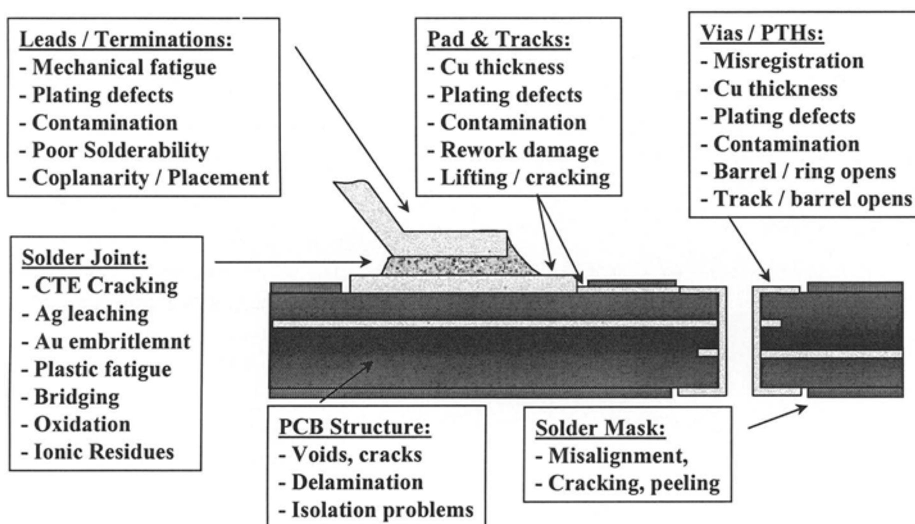


Figure 2. Solder Joint / PCB Level Failures

3. EFFECT OF TEMPERATURE ON THE RELIABILITY OF MICROELECTRONICS

Table 1 shows the influence of temperature on electronic device failure mechanisms for a few typical examples. The failures mechanisms are observed to be dependent on temperature (T), the magnitude of temperature change (ΔT), and temperature gradient (∇ T) in the range –55 to 125 °C and the rate of change of temperature with time (dT/dt). This table shows only a few typical examples; a complete review of this subject is contained in the work of *Bar-Cohen and Kraus* [6].

المنارة للاستشارات
ww

The temperature dependence of some electrical properties, and electron as well as dopant mobility in silicon, results in a link between temperature and electrical failure mechanisms. Similarly, many of the chemical processes occurring in corrosion and intermetallic formation, are accelerated by elevated temperature leading to failures in metallisation and bonded interfaces such as wire bond and die bond. Mechanical failures arise due to differential thermal expansion between the package materials.

Table 1. Influence of Temperature on Electronic Device Failure Mechanisms

Failure Site	Failure Mechanism	Dominant Temperature Dependence	Nature of Steady State Temperature Dependence
Wire Bond	Shear Fatigue	ΔT	Independent of steady state temperature
Die	Fracture	ΔT, ∇T	Primarily dependent on temperature cycle
Encapsulant	Cracking	ΔT	Independent of Steady State Temperature below the Tg of the encapsulant.
Package	Stress Corrosion	dT/dt	Mildly steady state temperature dependent under normal conditions
Die	Electrical overstress	T	Independent of steady state temperature below 160°C
Die Metallisation	Corrosion	dT/dt	Only occurs above dew point temperature, Mildly steady state temperature dependent under normal operation
Die Metallisation	Electromigration	∇T	Steady state temperature dependent above 150°C

4. THERMOMECHANICAL ISSUES IN MICROELECTRONICS

The mismatch between the mechanical properties of electronics packaging materials coupled with the temperature changes associated with packaging processes lead to the introduction of thermomechanical stresses. These stresses are a major reliability concern. *Table 2* gives typical mechanical properties for the materials in a plastic package.

المنارة للاستشارات
ww

Table 2. Typical Mechanical Properties of Plastic Package Materials

Material	Coefficient of Thermal Expansion (CTE) (ppm/°C)	Elastic Modulus (Gpa)
Silicon	3	188
Plastic	13-20	10-15
Leadframe	12-15	4-17
Die Attach (Epoxy)	40-60	1-5

The main stress inducing processes involved in package assembly are; die attach, encapsulation and solder reflow. *Figure 3* shows a sketch of a cross section through a plastic package indicating the package materials and the mechanism of stress transfer between the different materials. In this figure T refers to tensile stress, C to compressive stress and S to shear stress. The mechanism of stress transfer in plastic packages is very complex and a full explanation of this mechanism can be found in the work of *Kelly* [7].

After the die attach process, the mismatch between the mechanical properties of the die and die paddle result in the die paddle transferring compressive stress into the die at an angle of 45° from the bottom edge of the die.

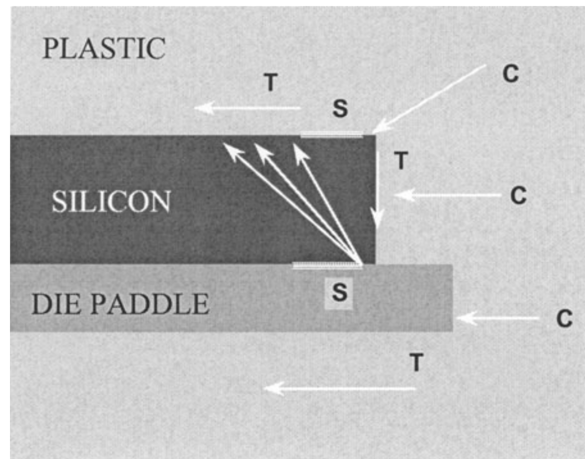


Figure 3. Packaging Induced Stresses in a Plastic Package[7]

There is a significant mismatch between the CTE of the plastic and that of the die. After encapsulation, which typically occurs at 180°C, contraction of the molding compound results in a build up of compressive and shear stresses on the die surface and tensile stress in the plastic. The corner of the

المنارة للاستشارات

ww

die represents the first point of restraint for molding compound contraction and, thus, the compressive stresses at this point are particularly high.

5. FUTURE TECHNOLOGY REQUIREMENTS

Tables 3 and *4* show the predicted assembly and packaging technology requirements for the next 14 years. These predictions were compiled for the Semiconductor Industry Association Technology Roadmap for Semiconductors [8]. *Table 3* shows thermal requirements for single chip packages (SCP) covering the complete range of applications form low cost to memory devices. This table indicates that power levels will continue to increase while maximum allowable junction temperatures and operating temperatures will remain largely unchanged. In this context, the need for improved thermal management techniques is obvious.

Table 3. Assembly and Packaging Technology Requirements (Thermal)

Year	2000	2004	2008	2014
Power SCP (Watts)				
Low Cost	n/a	n/a	n/a	n/a
Hand Held	1.6	2.3	2.5	2.7
Cost Performance	54	88	104	115
High Performance	100	150	170	183
Harsh	14	14	14	14
Memory	1.0	1.8	2.5	3.5
Junction Temperature Maximum (°C)				
Low Cost	125	125	125	125
Hand Held	115	115	115	115
Cost Performance	95	85	85	85
High Performance	95	85	85	85
Harsh	155	155	185	185
Memory	100	100	100	100
Operating Temperature Extreme Ambient (°C)				
Low Cost	55	55	55	55
Hand Held	55	55	55	55
Cost Performance	45	45	45	45
High Performance	45	45	45	45
Harsh	-40 to 150	-40 to 150	-40 to 180	-40 to 180
Memory	55	55	55	55

Among the most difficult challenges facing the assembly and packaging community over the next 5 years is the need for new and improved packaging materials. The roadmap identifies a requirement for improved

المنارة للاستشارات

www

organic substrates of lower CTE, lower moisture absorption and a T_g compatible with lead free soldering. The need for improved underfills was also highlighted.

Table 4 shows predicted trends in chip size, overall package profile and pincount. In logic chips with 800 pins or higher, conventional package design is inappropriate since the chip area must be increased for installation of the terminals alone. This trend coupled with a requirement for reduction in package size means that an area array configuration must be adopted. Here the package has terminals arranged in grid form on the entire surface of the chip. To achieve further reduction in size of packages and high density boards at low costs, flip chip connection in ball grid arrays (BGA's) is expected to produce good results. Conventional ceramic substrates will have to be replaced with low cost organic materials for most applications. The CTE of organic materials is higher than that of ceramics and extensive materials research is required to develop materials whose CTE more closely matches that of silicon.

Environmental concerns mean that conventional PbSn solder is currently being replaced by lead free materials. Lead free solders typically require higher processing temperatures in comparison with PbSn and thus packaging materials with higher glass transition temperatures are required. Furthermore, improved adhesion and moisture resistance for increased mechanical strength of connections, as well as improved hygroscopy for increased reliability, will be needed for the underfill materials for flip chips.

In addition to these technology requirements, there is a growing demand for today's' electronics products to be produced in less time and at a lower cost than previously. To reduce time to market and product costs, there is a requirement for a fast, cost-effective and accurate means of characterising electronics performance. Validated simulation tools can provide the solution to this requirement. However, there are a number of issues that need to be addressed to render the simulations tools currently available more accurate, effective and user friendly.

New simulation tools that couple thermal, mechanical and electrical effects are required. In electrical modelling, high frequency circuit modelling is a growing challenge as clock speeds continue to increase. Accurate modelling of interconnect parasitics is particularly challenging. More accurate thermal, mechanical and electrical material properties are required to increase simulation accuracy. Currently, simulation results cannot directly be translated into reliability predictions since the behaviour of materials and interfaces under operational and environmental stresses is poorly understood. Improved reliability models need to be developed to address this shortcoming.

المنارة للاستشارات

ww

Table 4. Assembly and Packaging Technology Requirements

Year	2000	2004	2008	2014
Chip Size (mm^2)				
Low Cost	55	63	72	90
Hand Held	55	63	72	90
Cost Performance	170	225	270	351
High Performance	450	595	713	937
Harsh	55	63	72	90
Memory	139	167	200	262
Overall Package Profile (mm)				
Low Cost	1.7	1.0	1.0	1.0
Hand Held	1.2	0.8	0.65	0.5
Cost Performance	1.2 - 1.7	0.8 - 1.0	0.65 - 0.8	0.65 - 0.8
High Performance	n/a	n/a	n/a	n/a
Harsh	1.4	1.0	1.0	1.0
Memory	1.2	0.8	0.65	0.5
Package Pincount				
Low Cost	86 - 313	118 - 426	160 - 580	254 - 920
Hand Held	138 - 397	188 -541	256 - 736	406 - 1167
Cost Performance	400 - 821	544 - 1247	740 - 1893	1174 -3541
High Performance	1792	2820	4437	8758
Harsh	40 - 259	40 - 353	40 - 480	40 - 761
Memory	44 - 128	48 - 160	48 - 182	44 - 220

6. SUMMARY AND SOLUTIONS

The objective of this paper is to present a brief introduction to the vast and complex topic of thermal and mechanical problems in microelectronics, specifically packaged devices. Current trends and future predictions show that management of thermal and mechanical issues is becoming increasingly more important due to increasing heat fluxes, continued miniaturisation and higher speeds.

Selection of thermal solutions is very much application driven. For hand held devices, use of higher thermal conductivity materials and reduction in component internal resistance provide the most viable solutions. With the emergence of area array flip chip technology, the backside of the chip provides a direct heat path for cooling. Thermal resistance of material interfaces can contribute significantly to overall package resistance. A lot of research effort is currently being applied to the development of interface materials that can be used to minimise interfacial solutions.

System level design solutions will also need to be developed to cater for increasing power levels and lower junction temperatures. System level solutions will include more advanced / efficient air cooling, boundary layer control, engineered surfaces and cost effective alternative cooling systems.

المنارة للاستشارات

ww

One of the main problems associated with prediction of mechanical reliability is the lack of availability of accurate and complete materials data. Improvements in material characterisation techniques and acceptance criteria, especially in areas such as interfacial adhesion, are critical. With increasing pressure on industry to release products into the market place in as short a time as possible, there is increasing emphasis on development of new, rapid test methods for materials, substrates and components.

To reduce package thermomechanical stresses it is necessary to reduce the CTE mismatch between the chip and the other package materials. Thus, plastics and substrates with lower CTEs (silicon CTE 3 ppm/°C) are currently being developed. Materials with increased glass transition temperature (T_g) will also contribute to lower package stresses.

Increased use and applicability of simulation tools is a key enabling factor for future product development. For all applications, thermal and mechanical considerations need to become an integral part of the overall component design. An integrated design tool in which modelling capabilities are incorporated into packaging CAD systems is one possible solution. Thermal, mechanical and electrical issues can no longer be treated separately. Higher power dissipation levels require that thermal simulations will need to be solved consistently with electrical behaviour. High frequency circuit modelling is a major challenge facing the electrical simulation community and improvements on existing techniques are currently being investigated.

The current state-of-the-art in modelling is that powerful simulation tools are available, along with low cost computing power, for electrical, thermal and mechanical performance. However, a high degree of user expertise is required to accurately simulate device performance and it is generally necessary that quite detailed models are built to accurately represent component geometries, materials and operating conditions. More widespread use of simulation techniques requires that compact models be derived from validated detailed models.

Within the electronics thermal community, there is a growing demand that component manufacturers supply compact thermal models of their parts in preference to the single thermal resistance value which is common today. Development of compact thermomechanical models is a more complex issue requiring improvements in material behaviour prediction and also significant developments in reliability predictions. Improved reliability models are necessary to be able to translate results of thermomechanical stress simulations into lifetime and reliability predictions.

Current technology trends pose significant challenges to those responsible for management of thermal and mechanical issues in electronics. The solution is to adopt a totally integrated design approach where thermal

المنارة للاستشارات

ww

and mechanical issues are considered at all stages of product development and where thermal and mechanical models are incorporated into package CAD systems. Achievement of this ideal hinges on development and widespread availability of accurate compact models.

ACKNOWLEDGEMENTS

The authors wish to thank Adrian Mulcahy and Finbarr Waldron of the Industrial Access Group, NMRC as well as John Barton of the Technology Research and Development for Industry Group, NMRC. The commission of the European Communities provided support and funding for much of the work presented during various EU funded projects.

REFERENCES

[1] Plastic Packaging of Microelectronic Devices, L.T.Manzione, 1990, Van Nostrand Reinhold

[2] "A Bond Failure Mechanism", T.Koch, W.Richling, J.Whitlock, D.Hall, Proceedings IEEE International Reliability Physics Symposium, 1986, pp 55 – 60

[3] Integrated Circuit Quality and Reliability, E.R.Hnatek, 1995, Marcel Dekker Inc.

[4] Thermal Stress and Strain in Microelectronics Packaging, J.H.Lau, 1993, Van Nostrand Reinhold

[5] Microelectronics Packaging Handbook, R.Tummala, 1997, Chapman and Hall

[6]"Advances in Thermal Modelling of Electronic Components and Systems", A.Bar-Cohen, A.D.Kraus, IEEE/ASME Press 1993, Volume 3, IBSN 0 – 7918 – 0018

[7]"The Simulation of Thermomechanically Induced Stresses in Plastic Encapsulated IC Packages", G.Kelly, Kluwer Academic Publishers, 1999, ISBN 0 – 7923 – 8485 - 7

[8]Semiconductor Industry Association, International Technology Roadmap for Semiconductors: 1999 Edition, Austin TX: International Sematech, 1999

المنارة للاستشارات

ww

SOLDER MATERIAL CHARACTERISATION AND MODELLING

Steffen Wiese, Frank Feustel, Ekkehard Meusel
Technische Universität Dresden
Institut f. Halbleiter- und Mikrosystemtechnik
wiese@ihm.et.tu-dresden.de

ABSTRACT

Constituive behaviour and fracture of micro solder joints are supposed to differ very much from that of bulky samples. The paper presents the results of reversible shear tests on flip chip solder joints under isothermal conditions. Two micro shear testers have been designed and built for this task. One tester is optimized to achieve high precision. The second tester works very similar but fits in a UHV chamber. This way, it enables in-situ SEM observations during the test. The experimental program included cyclic shear tests for elastic plastic material data as well as a new reversal creep and relaxation test procedure for time dependent deformation data. FEM simulation has subsequently been applied to evaluate the experimental raw data to extract material parameters according to the material models provided by ANSYS ®. Crack propagation was investigated by isothermal mechanical cycling.

1. INTRODUCTION

FEM simulations are widely used in order to estimate thermomechanical stresses in flip-chip-packages [1-7]. However the precision of quantitative FEM results (e.g. local strain values) depends above all on the assumed material behavior [8]. The application of material parameters that have been gained in standardized tests with bulk specimens suffers from the fact that a flip chip joint in real applications experiences different loading characteristics, has a

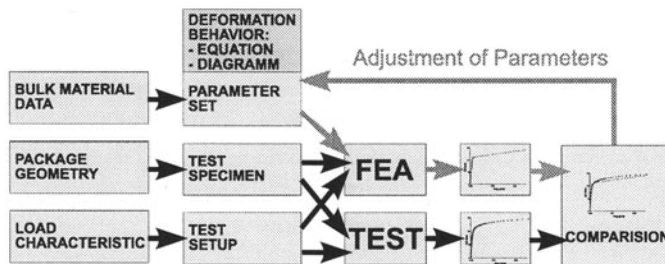


Figure 1: Calibration of Material Model

G. Q. Zhang et al. (eds.), Benefiting from Thermal and Mechanical Simulation in Micro-electronics, 27–36.
© 2000 *Kluwer Academic Publishers.*

المنارة للاستشارات
ww

much different geometry and consists of a different microstructure. Consequently, a new research method should take into account all characteristics of thermomechanical fatigue of flip chip joints. The simplest idea was to perform shear experiments on real flip chip joints. However, an evaluation of such experiments could not be carried out with simple analytical calculations, because the specimen geometry as well as the load characteristic are too complex. The key idea was to perform a 3D FE-Analysis of the experiment and to compare the experimental results with simulated results. Adjusting the parameters to commonly known material laws, the course of the simulated test is brought into line with the experiment (Figure 1).

2. SPECIMEN

The specimen (Figure 2) consisted of two silicon chips (3,3mm X 3,3mm) bonded vis à vis against each other by four solder joints (one on each corner). The geometry of a flip chip joint has the shape of a hyperbola, because then the majority of strain is accumulated inside the joint and away from the interface areas on the top and bottom of the joint. For all tests specimens a 100µm² footprint and about 150 µm height was used.

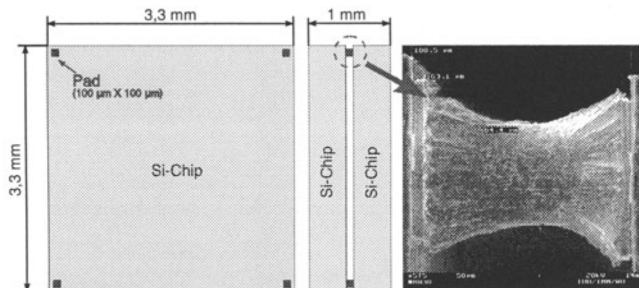


Figure 2: Test specimen

3. EXPERIMENTAL EQUIPMENT

The experimental setup has to satisfy two major requirements. One is, that it should be able to measure the force displacement behaviour of the very little flip chip joints correctly, and the other is, that it has to impose load conditions of simple shear on the specimen. On a more detailed view, the later requirement is completed by demand for appropriate strain rates and amplitudes, which means a well controlled movement of the two silicon chips of the specimen against each other in such a way that all bending moments and out of plane forces are eliminated or at least minimized. The realization of this requirements bases on idea of a symmetric grip configuration of two identical

المنارة للاستشارات
ww

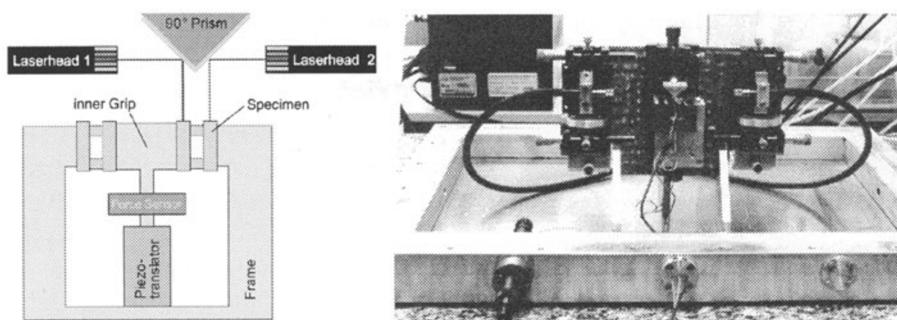


Figure 3: Microtester I

specimens as well as on the use of a piezoelectric translator (Figure 3). Such a translator, which drives the test setup, provides a smooth movement with subnanometer resolution for a wide range of velocities. A force sensor is placed between the inner grip and the piezoelectric translator. The translation direction is in line with the gravity axis, so that no bearings are necessary in order to eliminate out of plane forces due to equipment weights (grips, silicon chips). A high resolution displacement measurement is achieved by the application of a dual beam laser interferometer. Each laser beam is reflected at one of two neighboring edges of the specimen silicon chips. In order to minimize errors due to temperature changes, the laser interferometer heads are borne symmetrically to the specimen. This symmetric configuration compensates all thermal expansions of the aluminum positioning stages, so that a high accuracy in displacement measurement can be achieved, although the thermal chamber cannot be run with a stability smaller then 0.1K per day. Theoretically, measurement resolutions of 2 nm for displacement and 1mN for force can be achieved. Actual operational resolutions are 4 ... 10 times poorer. The available deformation rates are in a range between 10^{-5} ... 10^{3} µm/s, which refers to strain rates between 10^{-7} ... 10^{1} s^{-1} for the described specimen.

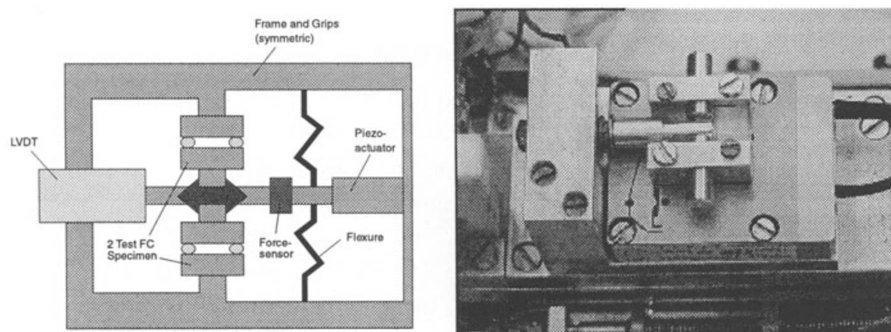


Figure 4: Microtester II (for SEM)

المنارة للاستشارات
ww

Although the first test setup (Figure 3) is capable to record very precise force displacement hysteresis during the course of the crack propagation experiment, it is impossible to observe the fracture evolution during the experiment. Therefore a second test setup had been designed (Figure 4). The second tester works basically like the first tester, but is much smaller in size, so that it can be used for in-situ experiments inside a UHV of an SEM. In contrast to the first tester, a LVDT is used for displacement measurements. The second tester reaches consequently poorer resolutions of 10 mN and 100nm, which is sufficient to carry out equivalent experiments on the first and the second tester.

4. EXPERIMENTAL PROGRAMM

The material behavior of metals is usually investigated by two tests. Both tests are carried out at a defined temperature and load the material with either a constant strain rate or a constant stress. These two types of experiments led to a division in the consideration of material responses into linear elasticity, nonlinear plasticity (time independent) and creep (time dependent behavior). But in a elementary view of physics there is no difference between plasticity and creep. Both responses are dealing with the kinetic processes in the material. However from test at constant strain rate it is easier to conclude to the

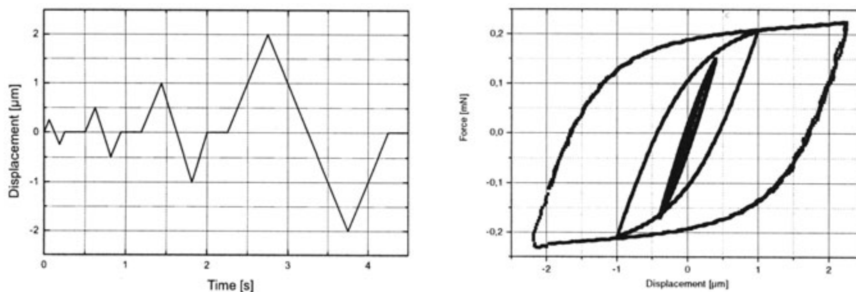


Figure 5: Incremental step test (load and result graph)

elastic and plastic behaviour, while the test at constant stress indicates very well a steady state condition of material creep. Conventionally tests are carried out with monotone loading. However, in order to receive information about the material response under true environmental conditions tests need to have reversible loading characteristics. The material behaviour at various cyclic reversible loads is characterized by the cyclic stress strain curve (CSSC), which represents the steady-state stress amplitude as a function of the plastic strain. The loading profile of the corresponding incremental step test (Figure 5) is characterized by the incline of displacement amplitude with the number

المنارة للاستشارات
ww

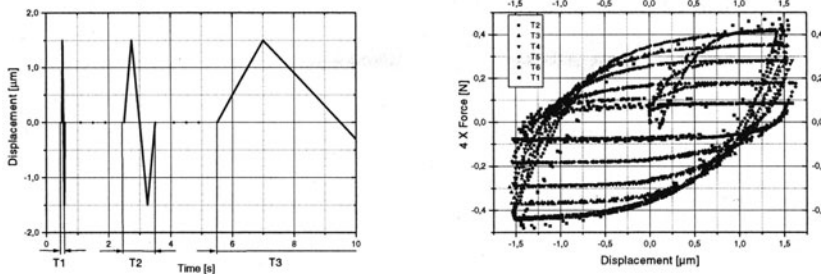


Figure 6: Incremental frequency test (load and result graph)

of cycles, which leads to a set of force displacement hysteresis. The turning points of the resulting simulated stress strain hysteresis are finally connected, in order to develop the CSSC, which provides all important information about the cyclic material behavior (e.g. cyclic material hardening or softening). The CSSC refers to only one displacement rate. In order to investigate to material response at different displacement rates an incremental frequency test (Figure 6) is used. The load characteristic consists of triangular displacement waves with constant amplitude but variable frequency. The wave frequency is increased stepwise (in decades). The result of this test is a set of force displacement hysteresis characterizing the material behavior at different strain rates.

More difficult than the reversible tests at constant strain rates is a reversible test at constant stress. The basic idea for such a test is shown in Figure 7. First a constant force ($+F_1$) is applied and the flip chip joint creeps until it reaches a defined shear angle ($+\gamma_{max}$). This shear angle is held until the force is relaxed to a certain value (e.g. $+F_0 = 1\%$ of $+F_1$). Then the shear force is reversed to a equal negative value ($-F_1$) and the joint creeps backwards until it reaches a defined negative shear angle ($-\gamma_{max}$), which is again held until the force is relaxed to a corresponding negative value (e.g. $-F_0 = 1\%$ of $-F_1$). Subsequently, the shear force is reversed and increased ($+F_2$), so that the joint

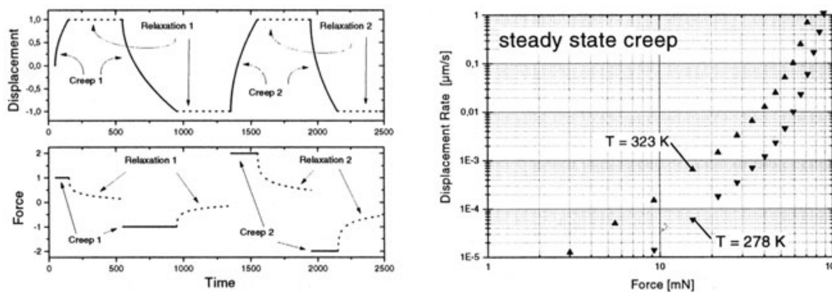


Figure 7: Incremental and reversible creep test (load and result)

المنارة للاستشارات
ww

creeps forward until it reaches $+\gamma_{max}$ again. The following relaxation is carried out until the force value reaches $+F_0$. This incremental and reversible creep test uses only one specimen for all forces, this avoids errors due to the scatter between different specimens. Since the maximum shear angle is predefined the total deformation of the joint corresponds with deformations in real applications. The determination of a forward and backward creep at the same applied force allows to determine the zero force offset due to noise, drift of the force cell and residual forces in the joint.

5. CONSTITUTIVE MODELLING

Three dimensional nonlinear FEM modeling was done with ANSYS ® version 5.3. A half flip chip solder joint was modeled by 500 elements. Element types have been either solid92 (+ user defined creep model) or visco107. The initial and the subsequent cycle was simulated, when cyclic experiments had been evaluated. The most effective model, which has shown good fit ca-

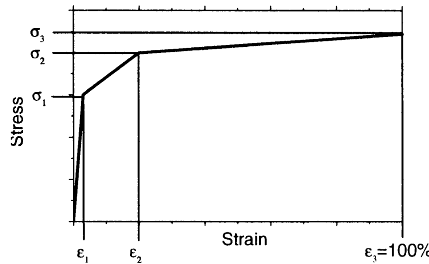


Figure 8: Elastic-plastic Model

pability consisted on a dual description divided into an elastic-plastic model and a creep model. The time independent elastic plastic description is characterized by 5 parameters (σ1, σ2, σ3, ε1, ε2).The model uses the multilinear kinematic hardening model of ANSYS ®. It is composed of 3 adjoining lines starting in the origin. Steady state creep was modeled by equation 1, which must be implemented to ANSYS ® by little modifications in source code.

$$\dot{\varepsilon} = C_1 \cdot [\sinh(C_3 \cdot \sigma)]^{C_2} \cdot \exp\left(-\frac{C_4}{T}\right) \quad (1)$$

This equation is already included in the Anand's model for viscoplastic elements, where the Parameters $C_1 ... C_4$ correspond with A, 1/m, ξ/s, Q/k. The parameter h_0 must be set to zero in order to skip all evolution equations, because SnPb37 Solder does not undergo cyclic hardening. The use of viscoplastic elements considers only elasticity and steady state creep, which leads to increasing errors, when deformations are little or strain rates are high.

المنارة للاستشارات

ww

PAR	ε_1	ε_1	σ_1	σ_2	σ_3	C_1	C_2	C_3	C_4
Unit	-	-	MPa	MPa	MPa	1/s	-	-	K
280K	7E-4	3E-3	21	41	600	10	2	0,2	5400
320K	7E-4	3E-3	19	31	200	E = 29 GPa			

Table 1: Parameters for SnPb37 Solder in Flip Chip Joints

6. CRACK PROPAGATION

All crack propagation experiments were carried out as cyclic tests at constant strain rate. A completely reversed symmetric triangular wave with constant frequencies ranging from $4 \cdot 10^{-4}$Hz ...$1 \cdot 10^{1}$Hz was used as load function for all tests. During these tests the specimens were subjected to constant total displacement range with amplitudes ranging from 0.5µm ... 3µm, which corresponds to shear angels of $\gamma = 0{,}18°$ ... 1,8° for the described joint geometry. Throughout the whole test every load versus displacement hysteresis was recorded with a rate of 100 scans per cycle. More Information about the experimental program is given in [9]. The force amplitude decreased with the number of cycles in the course of the test (Figure 9). The photographs (Figure 10) that were taken in the SEM (Microtester II) during the course of the crack propagation experiment show that the decrease in force amplitude corresponds with the crack growth in the solder joint. The fact that the plastic strain range increases due to the decrease of uncracked area might be considered as an incorrect test condition. This however corresponds with real operating conditions, because changes in material response of the solder do not influence the total displacement amplitude, since solder is the weakest component in the flip chip assembly.

In order to model crack propagation in FEM simulations there exists a variety of solutions, which differ from each other in terms of effort and preci-

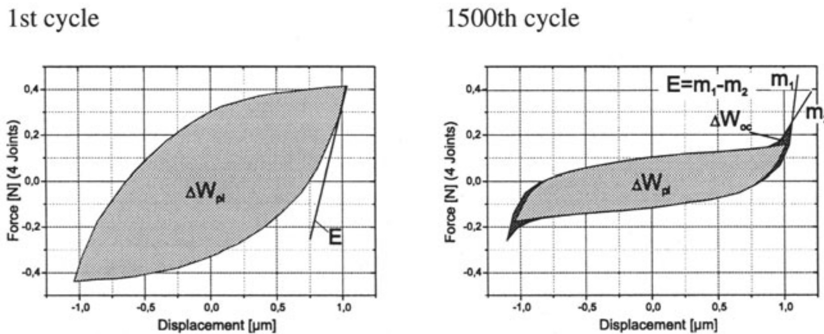


Figure 9: Subsequent force-displacement hysteresis

المنارة للاستشارات
ww

sion. The crack modelling approach which will be proposed in this paper might be not the most precise one, it is however very simple to understand and demands only a minimum of simulation effort (Two thermal cycles + the initial half cycle). Two different criteria have been. tested, to develop a simple crack propagation model. The first criterion was plastic strain energy ΔW_{pl} (eq. 2) and the second criterion was accumulated inelastic strain ε_{acc} (eq.3). Since most of the inelastic accumulated strain ε_{acc} of a entire cycle is creep

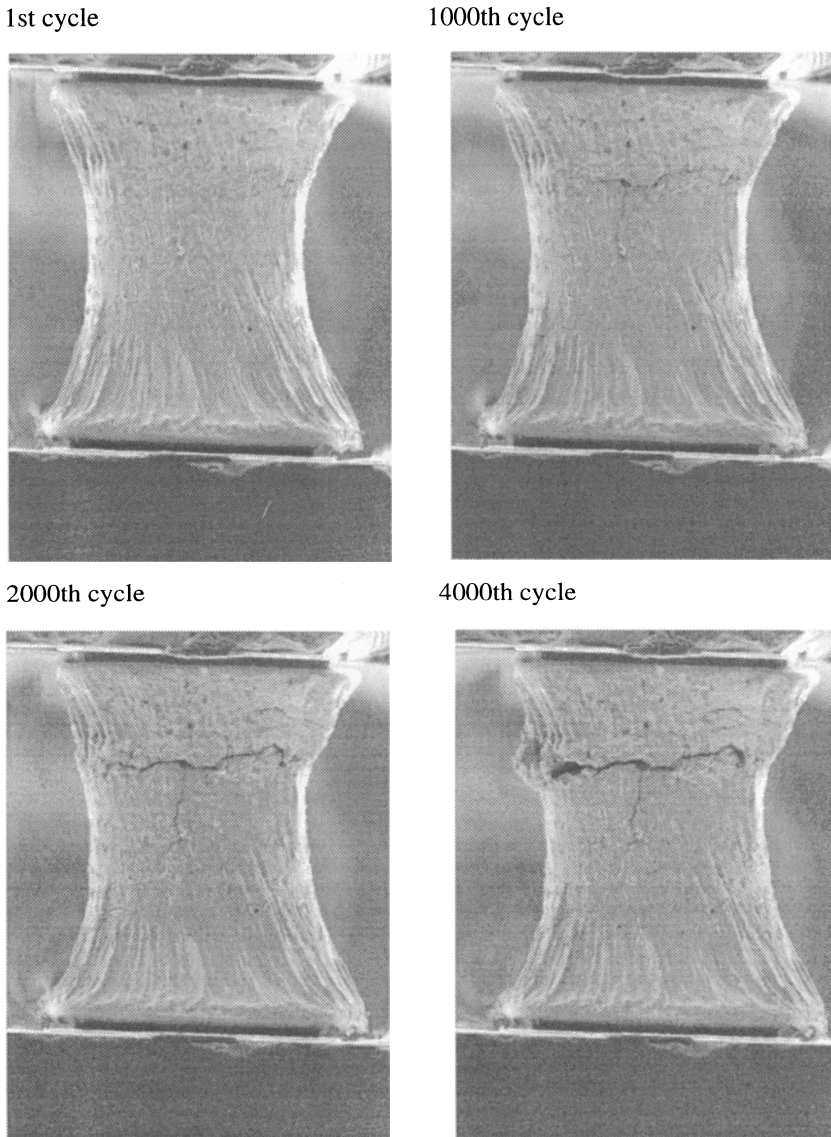


Figure 10: SEM shots from crack propagation experiment

المنارة للاستشارات
ww

strain the both criterias ΔW_{pl} and ε_{acc} work very similar for the calculation of crack propagation rate. The values for ΔW_{pl} and ε_{acc} were calculated by FEM Simulation. For this calculation an element inside a high stress region was chosen.

$$\frac{da}{dN} = 8 \cdot 10^{-7} \cdot \Delta W_{pl} \cdot \left(\frac{1Hz}{f}\right)^{0,2} \cdot \left(\frac{300K}{T}\right)^{c} \cdot \left[\frac{m^2}{MPa \cdot cycle}\right] \tag{2}$$

$$\frac{da}{dN} = 2 \cdot 10^{-7} \cdot \varepsilon_{acc} \cdot \left(\frac{1Hz}{f}\right)^{0,15} \cdot \left(\frac{300K}{T}\right)^{c} \cdot \left[\frac{m}{cycle}\right] \tag{3}$$

In contrast the Paris-Law like equation for crack propagation in [10], [11] the equations (2), (3) are corrected for cycle frequency and temperature. Until now only frequency dependence had experimentally investigated, which means that the current crack propagation model in equation 3 is only valid for room temperature.The experimental data for crack propagation is shown in (Figure 11) together with the fit equation lines.

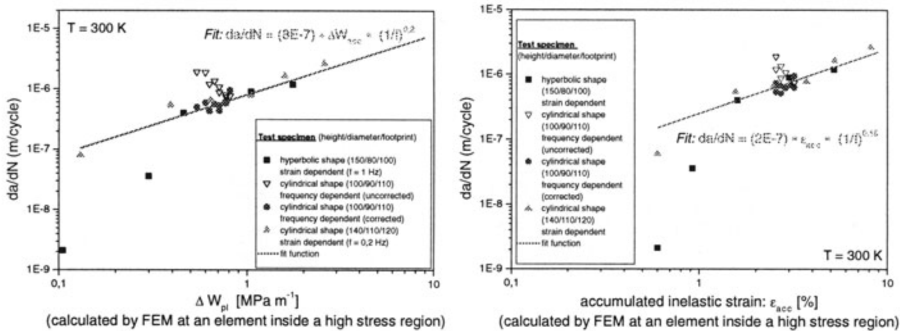


Figure 11: Experimental data for crack propagation rate

7. CONCLUSIONS

Two micro testers have been designed in order to carry out deformation as well as crack propagation experiments on real flip chip solder joints. The first tester was designed to record precisely the force displacement hysteresis of the flip chip joints. The second tester had poorer accuracy, but was able to work in a SEM vacuum chamber. Similar experiments were carried out on both testers, in order to receive a combined quantitative and spatial description of crack propagation in flip chip solder joints. The experimental raw data was evaluated by FEM simulation. Parameters of material models provided by ANSYS® have been determined. The crack growth rate for mode II loading was investigated in fully reversed shear on SnPb37 flip chip solder joints at room temperature. The crack growth rate was correlated to the dissipated energy at the crack tip as well as with total strain amplitude.

المنارة للاستشارات
ww

REFERENCES

1. Rzepka, S.; Meusel, E.: Flip Chips directly attached to FR4 Printed Circuit Boards: FEM Simulations and Experimental Tests, *Proc. ITAP(1995)*, San Jose, pp. 8-15

2. Auersperg, J.; Schubert, A.; Vogel, D.; Michel, B.; Reichl, H.: Fracture and Damage Evaluation in Chip Scale Packages and Flip-Chip-Assemblies by FEA and MicroDAC. *Proceedings of the Symposium on Application of Fracture Mechanics in Electronic Packaging at the 1997 ASME International Engineering Congress and Exhibition, Dallas*, Nov. 16-21, 1997, AMD-Bd. 222/EEP-Bd. 20, S. 133-138

3. Dudek, R.; Nylen, M.; Schubert, A.; Michel, B.; Reichl, H.: An Efficient Approach to Predict Solder Fatigue Life and its Application to SM- and Area Array Components. *Proceedings IEEE 47th Electronic Components and Technology Conference*, 1997, S. 462-471

4. Jendrny, J.; Müller, W.H.; Albrecht, H.-J.: Models and Concepts for Lifetime Prediction of SMT Solder Joints: A Comparison and Critical Evaluation. *Proceedings of the Micro Materials '97 Conference, Berlin, ddp Goldenbogen 1997*, S. 907-912

5. Feustel, F.; Meusel, E.:" Modellierung und Simulation von Einbaufällen". In: Direktmontage - Handbuch über die Verarbeitung ungehäuster IC's. Reichl, H. (Ed.). Berlin - Heidelberg - New York: Springer 1998, S. 255-261

6. Rzepka, S.; Feustel, F.; Meusel, E.: The Effect of Underfill Imperfections on the Reliability of Flip Chip Modules: FEM Simulations and Experiments, *Proceedings IEEE 50th Electronic Components and Technology Conference*, 25-28.05.1998, S. 362-370

7. Schubert, A.; Dudek, R.; Kloeser, J., Michel, B.; Reichel, H.; Hauck, T.; Kaskoun, K.: Experimental and Numerical Reliability Investigations of FCOB Assembblies with Process-induced Defects, *Proceedings IEEE 50th Electronic Components and Technology Conference*, 21-25.05.2000, S. 624-632

8. Feustel, F.; S.Wiese, S.; Meusel, E.: Time-Dependent Material Modeling for Finite Element Analyses of Flip Chips, *Proceedings IEEE 50th Electronic Components and Technology Conference*, 21-25.05.2000, S. 1548-1553

9. Wiese, S.; Feustel, F.; Rzepka, S.; Meusel, E.: Creep and Crack Propagation in Flip Chip SnPb37 Solder Joints, *Proceedings IEEE 49th Electronic Components and Technology Conference*, 01-04.06.1999, S. 1015-1020

10. Z. Guo; H. Conrad: Fatigue Crack Growth Rate in 63Sn37Pb Solder Joints, *Trans. of the ASME Journal of Electronic Packaging*, June 1993, Vol. 115, p.159-164

11. Wiese, S., Feustel, F., Rzepka, S., Meusel, E.: Crack propagation experiments on flip chip solder joints. *Electronic Packaging Materials Science X - MRS Symposium Proceedings*, April 14–16, 1998, San Francisco, Belton, D. J., Gaynes, M., Jacobs, E. G., Pearson, R., Wu, T. (Ed.), Materials Research Society, Warrendale, Vol. 515 (1998), pp. 227–232

المنارة للاستشارات

ww

POLYMER MATERIAL CHARACTERIZATION AND MODELING

Leo J. Ernst
Delft University of Technology
P.O. Box 5033, 2600 GA Delft, The Netherlands, Phone: +31 15 2786519
Fax: +31 15 2782150, E-mail: l.j.ernst@wbmt.tudelft.nl

Keywords: Polymers, Materials modeling, Constitutive relations, Electronic packaging, Curing, Viscoelasticity

Abstract: In computational prototyping of electronic packages 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. INTRODUCTION

When giving an overview of polymer material modeling and characterization, it can be realized that many models, from quite simple to very complicated, are presently available. For electronics packaging simulations the moderately simple linear viscoelastic models appear to be appropriate for most cases. Therefore these models will be discussed here.
Although viscoelasticity based models appear to be appropriate for thermal cycling simulations, up till now a severe obstacle remained for reliable computational prototyping of electronic packages, because of the deficiency to be able to adequately investigate the processing induced stress fields during and after fabrication. Possible damage originating from the fabrication process and the residual stress field directly after fabrication could not adequately be evaluated. This residual stress field is merely due to chemical shrinkage and simultaneous stiffness built-up in underfills and encapsulates 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. Recently, a cure dependent viscoelastic model was developed and applied in the electronic packaging field. Special attention will be given to this model and its application in sections 3.2-3.5.

G. Q. Zhang et al. (eds.), Benefiting from Thermal and Mechanical Simulation in Micro-electronics, 37–58.
© 2000 *Kluwer Academic Publishers.*

المنارة للاستشارات

ww

An overview of characteristic mechanical properties of polymers is given in section 2. The further discussion of mechanical modeling will be restricted to thermoset polymers. First of all, to introduce characteristic behavior of such polymers, in section 3.1 some simple models are discussed. A discussion of linear viscoelastic modeling for fully cured material is given in section 3.6. Further, a short overview of the more complicated non-linear viscoelastic models is finally presented in section 3.7.

2. POLYMERS

According to the mechanical behavior on their usual application temperature polymers often are split-up into 3 groups: "Thermoplastic Polymers", "Rubbers" and "Thermosetting polymers".
"Thermoplastic Polymers" can be shortly characterized as: -made up of linear polymer-chains having some branches, -being a viscous liquid (when heated above T_g) and solid through crystallization (below T_g), -the application temperature is below T_g, -reversible: solid ↔ viscous liquid -visco-elasto-plastic
"Rubbers" (elastomers) are shortly characterized as: -cross linked polymeric chains, -application temperature above T_g -allowing very large deformations, -rubber hysteresis (Mullins), -unstable for longer times: -chain breakage + chain growth -creep/relaxation, permanent set
"Thermosetting polymers" are characterized by: -cross linked polymeric chains, -application temperature below T_g, -stiff "glassy" solid state, below T_g, -"rubbery" state, above T_g
For practical purposes often additives are used within these polymers: For example: for high temperature protection (anti-oxidants), for influencing the stiffness, strength or fracture toughness and for reducing the price. Additives however can make the mechanical modeling much more complicated.
In the sequence, especially the mechanical behavior of thermosetting polymers will be considered.
In general, thermosetting polymers behave visco-elastic. Time-dependent behavior, such as "stress relaxation" under constant deformation and "creep" under constant stress level, is characteristic.

3. MECHANICAL MODELING

3.1 SIMPLE MODELS

Simple mechanical models, made up of "springs" and "dashpots", sometimes combined with "friction elements" are often used to illustrate simple 1D responses: Well known are the Kelvin element for the illustration "creep", the Maxwell element for illustration of stress relaxation and the Burgers model for a combination of both phenomena. For the understanding of

المنارة للاستشارات

subsequent modeling, some characteristics of the Maxwell element (a spring and dashpot in line) are shown:

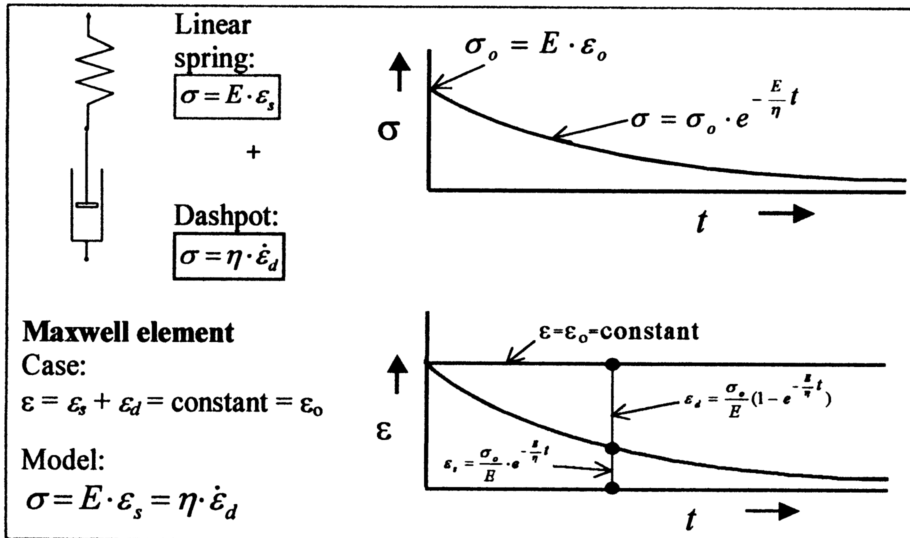


Fig.1: Illustration of the Maxwell-element

A constant strain, being applied at time 0, induces an initial stress σ_o, 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:

$$\sigma = \sigma_o \cdot e^{-\frac{E}{\eta}t} = \sigma_o \cdot e^{-t/\tau} \tag{1}$$

Here τ in the exponent is the so-called "relaxation time". This is the time to obtain an exponent –1. The stress is than reduced to 37%.

3.2 STRESS AND STATE DEPENDENT VISCOELASTICITY

For the presentation of more adequate and 3D-models first vectors of stress, strain and initial strain components are introduced.

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

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

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

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

المنارة للاستشارات

ww

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_o. (The actual application of this load increment is assumed to take place in an infinitesimal time increment $\Delta\xi$):

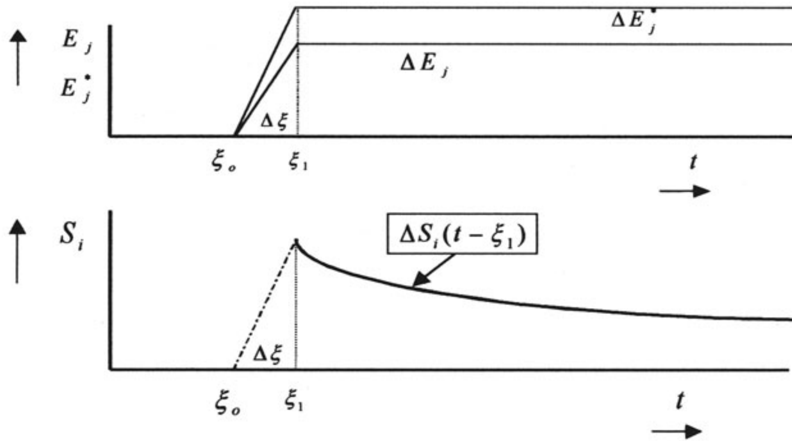


Fig. 2: Step relaxation response

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

$$\Delta S_i(t-\xi_1) = C_{ij}(t-\xi_1)\cdot(\Delta E_j - \Delta E^*_j) \qquad (2)$$

Here C_{ij} represent so-called relaxation moduli 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.

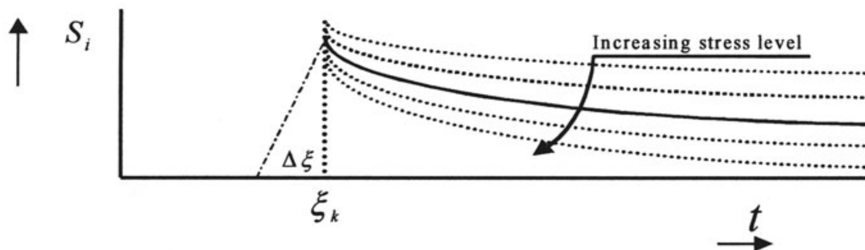


Fig. 3: Stress responses: Effect of increasing stress level

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. So we add

المنارة للاستشارات

ww

symbolically the stress state at load application time ξ_I to the relaxation moduli-functions: $C_{ij}=C_{ij}[\sigma(\xi_I), (t-\xi_I)]$.
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 (in strain level)" occurring over discrete load-application-time increments, $\Delta\xi$, so that subsequent "load application times" are:

$$\xi_1 = \xi_o + \Delta\xi,\ \xi_2 = \xi_o + 2\Delta\xi,\ \ldots\ldots\ldots\ \xi_k = \xi_o + k\Delta\xi, \ldots\ldots\ \ \xi_n = \xi_o + n\Delta\xi$$

Matching deformation and initial deformation increments than are:

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

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

For load-application time, number k, the stress increment - function is:

$$\Delta S_i^k(t-\xi_k) = C_{ij}^k(t-\xi_k)\cdot[(E_{j,\xi})_{\xi_{k-1}} - (E^*_{j,\xi})_{\xi_{k-1}}]\cdot\Delta\xi \qquad (3)$$

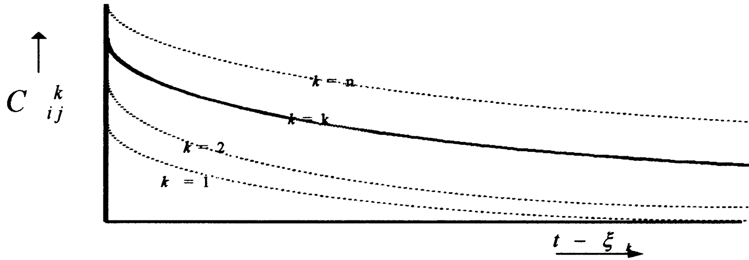


Fig.4: Curing material: Curing-time dependent relaxation functions

If we observe different relaxation-modulus functions for different load-application times (even for constant stress levels), the relaxation-modulus-functions of load-increment k should be considered to dependent on state functions $\alpha^l(\xi_k)$ (such as Temperature or Degree of cure):

$$C_{ij}^k = C_{ij}[\alpha^l(\xi_k), \sigma(\xi_k), (t-\xi_k)] \qquad (4)$$

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

$$S_i(t) = \int_{\xi=\xi_o}^{t} C_{ij}[\alpha^l(\xi), \sigma(\xi), (t-\xi)]\cdot\left\{\left(E_{j,\xi}\right)_\xi - \left(E^*_{j,\xi}\right)_\xi\right\} d\xi \qquad (5)$$

The experimental investigation of the stress and state dependent modulus functions $C_{ij}=C_{ij}[[\alpha^l(\xi), \sigma(\xi), (t-\xi)]$ for polymers being used in electronic packages is still a challenging task. Therefore, in section 3.4. the stress level dependency is assumed to be less important, resulting into the relations for "linear state dependent viscoelasticity".

المنارة للاستشارات

ww

3.3 INCREMENTAL RELATIONSHIP

If we want to implement the general constitutive relation (5) into a finite element program, we actually need an incremental stress-strain relation. According to this constitutive relation the stress-update $\Delta S_i = S_i(t+\Delta t) - S_i(t)$ for a time step Δt is represented by:

$$\Delta S_i = \int_{\xi=\xi_o}^{t+\Delta t} C_{ij}[\alpha_l(\xi),\sigma(\xi),(t+\Delta t-\xi)]\cdot\left\{\left(E_{j,\xi}\right)_\xi-\left(E^*_{j,\xi}\right)_\xi\right\}d\xi \quad -S_i(t) \qquad (6)$$

Here the integral represents the stress state at time $t+\Delta t$. The latter term represents the stress state at time t. Note that the time $t+\Delta t$ is not only present in the upper bound of the integral. It is also present in the kernel function. As a consequence, 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. This would result into a huge storage and data handling problem, making the model quite unattractive. The problem, however, can be circumvented or reduced by approximating the kernel function with Prony series. In particular the following Prony series approximation of the cure-dependent relaxation moduli is adopted:

$$C_{ij}[\alpha^l(\xi),\sigma(\xi),(t-\xi)] \approx \sum_{n=1}^{N} C^n_{ij}[\alpha^l(\xi),\sigma(\xi)]\cdot \mathrm{e}^{-(t-\xi)/\tau_n} \qquad (7)$$

Here the individual terms can be looked upon as responses of Maxwell elements with relaxation times τ_n (generalized for 3D). The participation factors, C^n_{ij}, are functions of the "load application time", ξ. On adequate experimental data, principally these participation factors and the relaxation times (τ_n) can directly be fitted. However, in the actual fitting procedure a number of relaxation times will be selected, while the fitting process is directed to the participation factors only.

The stress update procedure (based on the Prony-series approximation (7)) can be worked out as [Hof & Ernst, July 2000]:

$$S_i(t)=\sum_{n=1}^{N}\Re^n(t) \quad \text{and} \quad S_i(t+\Delta t)=\sum_{n=1}^{N}\Re^n(t+\Delta t) \qquad (8a)$$

$$\Re^n(t)=\mathrm{e}^{-t/\tau_n}\int_{\xi=\xi_o}^{t} C^n_{ij}[\alpha^l(\xi),\sigma(\xi)]\cdot \mathrm{e}^{\xi/\tau_n}\cdot\left\{\left(E_{j,\xi}\right)_\xi-\left(E^*_{j,\xi}\right)_\xi\right\}d\xi \qquad (8b)$$

$$\Re^n(t+\Delta t)=\mathrm{e}^{-\Delta t/\tau_n}\cdot\Re^n(t)+ \mathrm{e}^{-\Delta t/\tau_n}\cdot \mathrm{e}^{-t/\tau_n}\int_{\xi=t}^{(t+\Delta t)} C^n_{ij}[\alpha^l(\xi),\sigma(\xi)]\cdot \mathrm{e}^{\xi/\tau_n}\cdot\left\{\left(E_{j,\xi}\right)_\xi-\left(E^*_{j,\xi}\right)_\xi\right\}d\xi \qquad (8c)$$

This stress update only requires stored data for the functions $\Re^n(t)$ for the previous time. A stress update procedure, based on the above expressions can be implemented into a FEM program.

المنارة للاستشارات
www

3.4 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 stress levels remain relatively low (< 0.5 * ultimate strength). Without stress-level dependency we get so-called linear-state dependent viscoelasticity:

$$S_i(t) = \int_{\xi=\xi_o}^{t} C_{ij}[\alpha'(\xi),..,(t-\xi)] \cdot \left\{ \left(E_{j,\xi}\right)_\xi - \left(E^*_{j,\xi}\right)_\xi \right\} d\xi \qquad (9)$$

Focussing on linear state dependent viscoelasticity the handling of the state dependency can be different for the “curing situation”, where the chemical state is continuously changing, and the “fully cured” situation, where the physical aspects only give rise to changes in material behavior.
For the fully cured situation various methods are available to account for “temperature changes”, “Physical and Chemical aging”, “Influences of moisture”, etc. A simple method to account for the temperature dependency of fully cured material will be discussed in section 3.6. First, in section 3.5 the state dependency for the “curing situation“ is further discussed.

3.5 CURE DEPENDENT LINEAR VISCOELASTICITY

The linear state dependent viscoelastic formulation (9) will further be considered under the assumptions of “Isothermal curing” and “Isotropic behavior”. Then only one single state parameter $\alpha(\xi)$, the degree of cure, is necessary: $C_{ij}=C_{ij}[\alpha(\xi),..,(t-\xi)]$. Because of the isotropy the relaxation functions further depend on 2 parameter functions only, being the cure dependent bulk-relaxation-modulus and shear modulus functions $K=K[\alpha(\xi),..,(t-\xi)]$ and $G=G[\alpha(\xi),..,(t-\xi)]$, respectively:

$$C_{ij}[\alpha(\xi),..,(t-\xi)] = K[\alpha(\xi),..,(t-\xi)] \cdot V_{ij} + G[\alpha(\xi),..,(t-\xi)] \cdot D_{ij} \qquad (10)$$

with: $[V_{ij}] = \begin{vmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{vmatrix}$ $\quad [D_{ij}] = \begin{vmatrix} 4/3 & -2/3 & -2/3 & 0 & 0 & 0 \\ -2/3 & 4/3 & -2/3 & 0 & 0 & 0 \\ -2/3 & -2/3 & 4/3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 \end{vmatrix}$

$$K[\alpha(\xi),..,(t-\xi)] = \frac{E[\alpha(\xi),..,(t-\xi)]}{3(1-2\cdot\nu[\alpha(\xi),..,(t-\xi)])} \; ; \; G[\alpha(\xi),..,(t-\xi)] = \frac{E[\alpha(\xi),..,(t-\xi)]}{2(1+\nu[\alpha(\xi),..,(t-\xi)])}$$

Here $E=E[\alpha(\xi),..,(t-\xi)]$ and $\nu=\nu[\alpha(\xi),..,(t-\xi)]$ represent the cure dependent "Young's modulus” and “Poisson’s ratio” relaxation functions.
In the “fully cured” situation (without the $\alpha(\xi)$-dependency) these relaxation functions could simply be obtained from unit-step relaxation experiments on dog-bone specimens. Then we simply have to assume appropriate fitting

المنارة للاستشارات
ww

functions for the observed relaxation curves. However, for a "curing resin" dog-bone specimen can not be applied, as the resin is liquid at the start of curing. Further, the results of unit-step relaxation experiments should be affected by the ongoing chemical shrinkage (E^*_j can not be kept constant). In this case, it is more appropriate to use so-called Dynamic Mechanical Analyses (DMA). Here the (liquid) resin is applied between two parallel plates. A coin-shaped specimen is obtained in this matter.

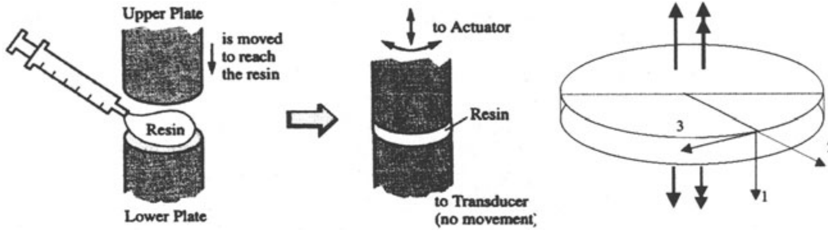


Fig.5: Coin-shaped resin sample + reference system.

This specimen can be loaded by longitudinal and by torsional deformation. Sinusoidal deformations with various frequencies can thus be applied. When the temperature of the parallel plates can be kept constant, then iso-thermal curing can be assumed. In reality isothermal curing cannot really be realized. A small temperature rise at the start of the curing reaction can be observed. However, generally this temperature rise occurs still in the liquid state.
As discussed in section 3.3. it is advantageous for the FEM-implementation to approxiamate the kernel functions by assuming Prony series for the modulus functions. Thus $K=K[[\alpha(\xi),.,(t-\xi)]$.and $G=G[[\alpha(\xi),.,(t-\xi)]$ will be aproximated by:

$$G[\alpha(\xi),(t-\xi)] \approx \sum_{n=1}^{N} G^n[\alpha(\xi)] \cdot e^{-(t-\xi)/\tau_n} \tag{11}$$

$$K[\alpha(\xi),(t-\xi)] \approx \sum_{n=1}^{N} K^n[\alpha(\xi)] \cdot e^{-(t-\xi)/\tau_n} \tag{12}$$

The participation factors $G^n = G^n[\alpha(\xi)]$ and $K^n = K^n[\alpha(\xi)]$ (the stiffness functions) are then subject to parameter investigation, when a number of appropriately spaced relaxation times is chosen.

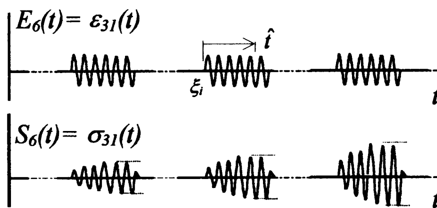


Fig. 6b: Intermittent torsional DMA
Specimen: $\phi 8 \times 1.1$ (mm)

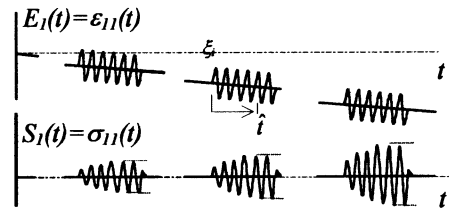


Fig. 6c: Intermittent longitudinal DMA
Specimen: $\phi 25 \times 1$ (mm)

المنارة للاستشارات

ww

In the torsional DMA-testing the longitudinal deformation remains completely free, while at certain (curing) times a sinusoidal torsional deformation, with amplitude E_o and angular frequency ω is applied. The observed response for the torsional stress component is periodical; A combination of a cosine and a sine with angular frequency ω, with increasing amplitudes, which in turn stabilize after a number of cycles. In the stabilized situation the amplitudes depend on the so-called "storage shear modulus" $G' = G'[\alpha(\xi_i),.,\omega]$ and "loss shear modulus" $G'' = G''[\alpha(\xi_i),.,\omega]$. Both are dependent on the angular frequency being applied and on the "(first) load application time" or "curing-time" (ξ_i). Using the Prony series approximation (11), for the "stabilized situation", the following relations can be established for the storage and loss shear modulus functions:

$$G'[\alpha(\xi_i),,\omega) \approx \sum_{n=1}^{N}[G^n[\alpha(\xi_i)] \cdot \frac{\omega^2 \cdot (\tau_n)^2}{\omega^2 \cdot (\tau_n)^2 + 1}] \tag{13}$$

$$G''[\alpha(\xi_i),,\omega) \approx \sum_{n=1}^{N}[G^n[\alpha(\xi_i)] \cdot \frac{\omega \cdot \tau_n}{\omega^2 \cdot (\tau_n)^2 + 1}] \tag{14}$$

In the longitudinal DMA-testing the axial deformation is kept free between the intermittent DMA-loading, such that the curing-shrinkage is not hindered. At certain (curing) times a sinusoidal axial deformation is superposed. The observed response for the axial stress component is also periodical; A combination of a cosine and a sine with (angular) frequency ω, with increasing amplitudes, which in turn stabilize after a number of cycles. In the stabilized situation the amplitudes depend on the so-called "storage longitudinal bulk modulus" $M' = M'[\alpha(\xi_i),.,\omega]$ and "loss longitudinal bulk modulus" $M'' = M''[\alpha(\xi_i),.,\omega]$. Both are dependent on the angular frequency being applied and on the "(first) load application time" or "curing-time" (ξ_i). Using the Prony series approximations (11+12), for the "stabilized situation" the following relation can be established for the storage and loss longitudinal bulk modulus functions:

$$M'[\alpha(\xi_i),,\omega) \approx \sum_{n=1}^{N}[M^n[\alpha(\xi_i)] \cdot \frac{\omega^2 \cdot (\tau_n)^2}{\omega^2 \cdot (\tau_n)^2 + 1}] \tag{15}$$

$$M''[\alpha(\xi_i),,\omega) \approx \sum_{n=1}^{N}[M^n[\alpha(\xi_i)] \cdot \frac{\omega \cdot \tau_n}{\omega^2 \cdot (\tau_n)^2 + 1}] \tag{16}$$

where: $$M^n[\alpha(\xi)] = K^n[\alpha(\xi)] + \frac{4}{3}G^n[\alpha(\xi)] \tag{17}$$

The evolution of chemical shrinkage can also be obtained from the longitudinal test as the (free) axial deformation is monitored. As the upper and lower surfaces of the sample are fixed to parallel rigid plates, where the shrinkage in radial (2-) direction and in circumferential (3-) direction is hindered, plane strain conditions can be assumed. Under these assumptions

المنارة للاستشارات

ww

and assuming isotropic initial deformations, E_1*= E_2*= E_3*= E*{α(ξ)}, following relation is obtained (Ernst, *et al.*, May 2000):

$$\left(\frac{\partial E_3}{\partial \xi}\right)_\xi = \frac{9K}{3K+4G}\left(\frac{\partial E^*}{\partial \xi}\right)_\xi \equiv \frac{1+\nu}{1-\nu}\left(\frac{\partial E^*}{\partial \xi}\right)_\xi \qquad (18)$$

The parameter identification process is illustrated for a selected epoxy (an acid anhydride system with a filler content of 63% amorphous SiO_2). Prony series approximations according to (11) and (12) with n=8 are used. The relaxation times τ_n are logarithmically equally spaced (see the table below).

n	1	2	3	4	5	6	7	8
τ_n	0.0002	0.0018	0.0133	0.098	0.7576	5.5556	41.667	2.10^9

Fig. 7: Chosen relaxation times in minutes.

Torsional DMA measurements were performed on a standard Rheometrics Dynamic Analyser RDA II, with sample diameter Ø8 [mm] and sample thickness of 1 [mm]. Fig. 8 shows the angular frequencies being applied.

k	1	2	3	4	5	6
ω_k	0.1	1	3.16	10	31,6	100

Fig.8: Chosen DMA angular frequences (ω_k) in rad/s.

Longitudinal DMA measurements were performed on a standard tensile/compression test machine, Zwick 1486, with sample diameter Ø25 [mm] and sample thickness of 1 [mm]. A special temperature controlled parallel plate construction was designed for application and loading of the resin specimen. Angular frequencies, being applied are presented in Fig. 9.

k	1	2	3
f_k	0.1	1	2

Fig. 9 Chosen DMA longitudinal frequencies f_k (=ω_k/(2π)) in s^{-1}

For each intermittent torsional DMA test, the steady state storage shear modulus is calculated from the established amplitude τ_o and phase shift δ according to:

$$G'(\xi_i',,\omega_k) = \tau_o(\xi_i',,\omega_k)\cdot\cos\{\delta(\xi_i',,\omega_k)\}/2E_o(r) \qquad (19)$$

A continuous dependency of the curing time is constructed by fitting the intermittent test steady state storage shear moduli on the following function:

$$G'(\xi,,\omega_k) \approx q_1(\omega_k)\cdot e^{-\{q_2(\omega_k)/\xi\}^{q_1(\omega_k)}} + q_4(\omega_k)\cdot e^{-\{q_5(\omega_k)/\xi\}^{q_6(\omega_k)}} \qquad (20)$$

The fitting results are presented in Fig. 10.

المنارة للاستشارات
ww

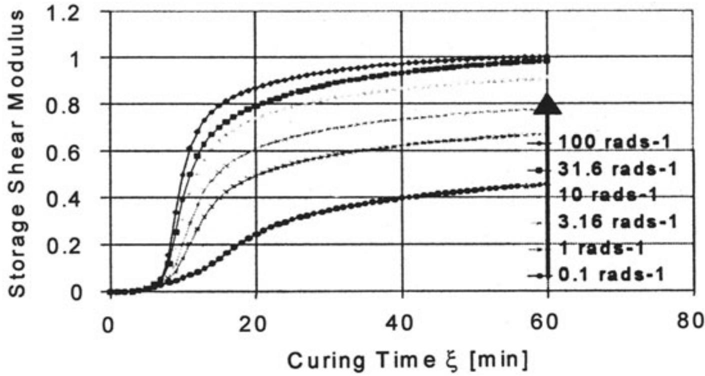


Fig. 10: Fitted results for the Storage shear moduli $G'(\xi,,\omega_k)$ in GPA.

As the intermittent test steady state storage shear moduli fit well on expression (20), a similar representation is assumed for the participation factors in the Prony series (11):

$$G^n\{\alpha(\xi)\} = G^n(\xi) \approx g_1^n \cdot \mathrm{e}^{-(g_2^n/\xi)^{g_3^n}} + g_4^n \cdot \mathrm{e}^{-(g_5^n/\xi)^{g_6^n}} \tag{21}$$

This representation of the stiffness coefficients $G^n\{\alpha(\xi)\}$ is substituted into the steady state expression (13):

$$G'(\xi,,\omega_k) = \sum_{n=1}^{N}\left[\{g_1^n \cdot \mathrm{e}^{-(g_2^n/\xi)^{g_3^n}} + g_4^n \cdot \mathrm{e}^{-(g_5^n/\xi)^{g_6^n}}\} \cdot \frac{(\omega_k)^2 \cdot (\tau_n)^2}{(\omega_k)^2 \cdot (\tau_n)^2 + 1}\right] \tag{22}$$

The intermittent test storage shear moduli data, as presented in Fig. 10, are fitted on this expression. The established fitting coefficients g_1^n, g_2^n, g_3^n, g_4^n, g_5^n, g_6^n, are subsequently substituted into expression (21) for the stiffness coefficients $G^n\{\alpha(\xi)\}$. The graphical representation of these stiffness coefficients is presented in Fig. 11

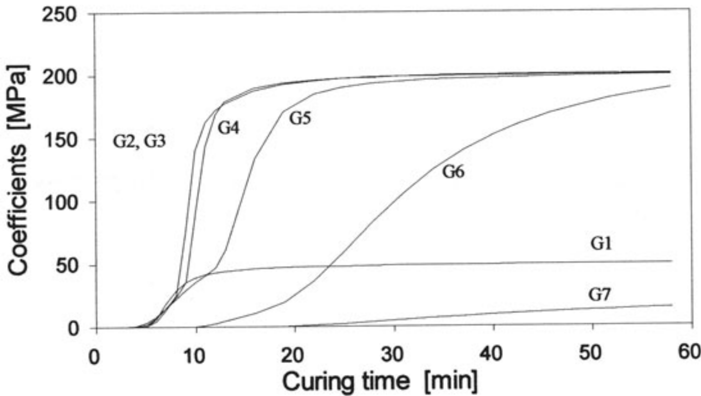


Fig. 11: Cure dependent stiffness coefficients $G^n\{\alpha(\xi)\}$; $G^8\{\alpha(\xi)\}$~0}

المنارة للاستشارات

ww

For each intermittent longitudinal DMA test, the storage longitudinal bulk modulus is calculated from the established amplitude σ_o and phase shift δ:

$$M'(\xi_i,,\omega_k) = \sigma_o(\xi_i,,\omega_k)\cdot\cos\{\delta(\xi_i,,\omega_k)\}/2E_o(r) \tag{23}$$

A continuous dependency of the curing time is constructed by fitting the intermittent test data on to the following function:

$$M'(\xi,,\omega_k) \approx r_1(\omega_k)\cdot e^{-\{r_2(\omega_k)/\xi\}^{r_1(\omega_k)}} + r_4(\omega_k)\cdot e^{-\{r_5(\omega_k)/\xi\}^{r_6(\omega_k)}} \tag{24}$$

The result of the fitting procedure is presented in Fig. 12 (the dots originating from the origin). Curves from different specimen were being shifted along the (curing) time axis to ensure the same gel point as in the shear-DMA tests.

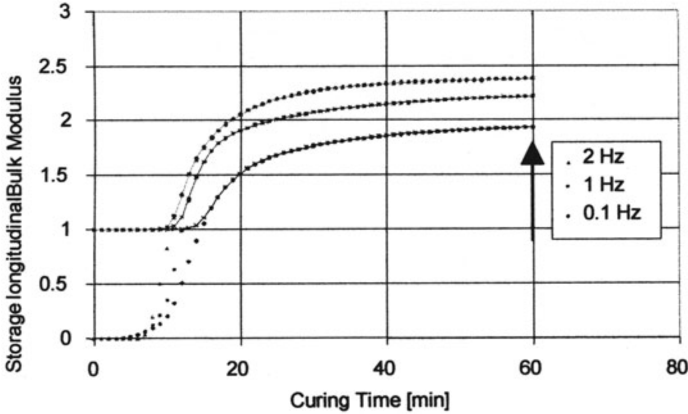


Fig. 12: Longitudinal storage bulk modulus $M'(\xi,,\omega_k)$ in GPa

Since for small curing times no axial stiffness is measured, the "dots" for the longitudinal storage bulk modulus, $M'(\xi,,\omega_k)$, originate from the origin, such that for small curing times: $M^n\{\alpha(\xi)\}\sim 0$. Further from Fig. 11 it follows that for small curing times $G^n\{\alpha(\xi)\}\sim 0$. According to (17), which is a consequence of the "plane strain assumption", the bulk modulus coefficients should also be approximately zero for small curing times: $K^n\{\alpha(\xi)\}\sim 0$. This, according to (12), would result in a zero bulk modulus K for small curing times. This of course is not physically correct. The minimum should be the "liquid-bulk modulus"-value. This is corrected, by modifying the results for the longitudinal storage bulk modulus, $M'(\xi,,\omega_k)$, using an appropriate lower limitation. The "dotted-curves" in Fig. 12 represent the thus modified results. The modified curves can be well fitted through following expression:

$$M'(\xi,,\omega_k) \approx C + \hat{q}_1(\omega_k)\cdot e^{-\{\hat{q}_2(\omega_k)/\xi\}^{\hat{q}_1(\omega_k)}} + \hat{q}_4(\omega_k)\cdot e^{-\{\hat{q}_5(\omega_k)/\xi\}^{\hat{q}_6(\omega_k)}} \tag{25}$$

A similar representation is assumed for the stiffness coefficients $M^n\{\alpha(\xi)\}$:

المنارة للاستشارات

ww

$$M^n\{\alpha(\xi)\} = M^n(\xi) \approx m_7^n + m_1^n \cdot e^{-(m_2^n/\xi)^{m_3^n}} + m_4^n \cdot e^{-(m_5^n/\xi)^{m_6^n}} \qquad (26)$$

This representation is substituted into expression (15):

$$M'(\xi,,\omega_k) = \sum_{n=1}^{N}\left[\{m_7^n + m_1^n \cdot e^{-(m_2^n/\xi)^{m_3^n}} + m_4^n \cdot e^{-(m_5^n/\xi)^{m_6^n}}\}\cdot\frac{(\omega_k)^2\cdot(\tau_n)^2}{(\omega_k)^2\cdot(\tau_n)^2+1}\right] \qquad (27)$$

The (corrected) longitudinal storage bulk modulus data according to Fig. 12 are fitted on this expression. The established fitting coefficients m_1^n, m_2^n, m_3^n, m_4^n, m_5^n, m_6^n, are subsequently substituted into expression (26) for the stiffness coefficients $M^n\{\alpha(\xi)\}$. The cure dependent stiffness coefficients $K^n\{\alpha(\xi)\}$ are than found from (17):

$$K^n\{\alpha(\xi)\} = M^n\{\alpha(\xi)\} - \frac{4}{3}G^n\{\alpha(\xi)\} = K^n\{\alpha(\xi)\} = K^n(\xi) \approx k_7^n + k_1^n \cdot e^{-(k_2^n/\xi)^{k_3^n}} + k_4^n \cdot e^{-(k_5^n/\xi)^{k_6^n}} \qquad (28)$$

The result of the fitting procedure is presented in Fig. 13.

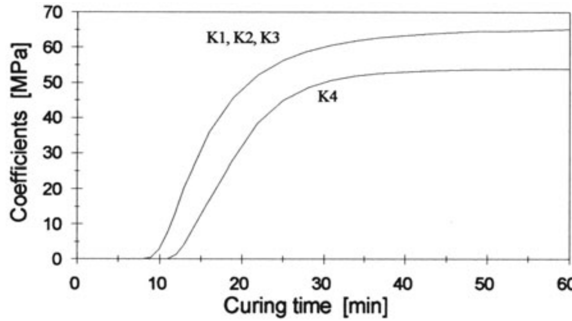


Fig. 13: Cure dependent stiffness coefficients $K^n\{\alpha(\xi)\}$, (K^8=1 in GPa)

Finally, since the cure dependent relaxation moduli K and G, and consequently also the cure dependent Poisson's ratio ν is established, the initial strain evolution is obtained according to expression (18) from the (free) axial deformation. The established axial deformation for the sample resin is given in Fig. 14.

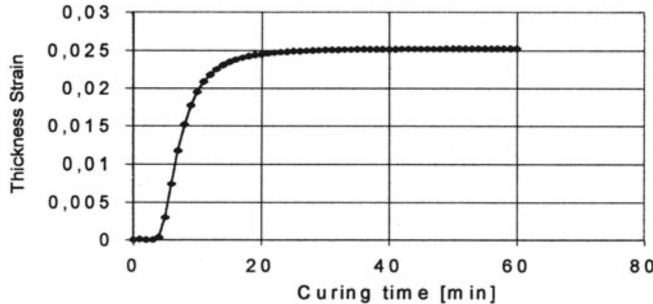


Fig. 14: Cure dependent axial deformation

المنارة للاستشارات
ww

As an illustration of the application of the cure dependent viscoelastic model on a practical situation, the curing stress evolution in a flip chip configuration (Fig. 15) as obtained from a 2D-FEM simulation is presented.

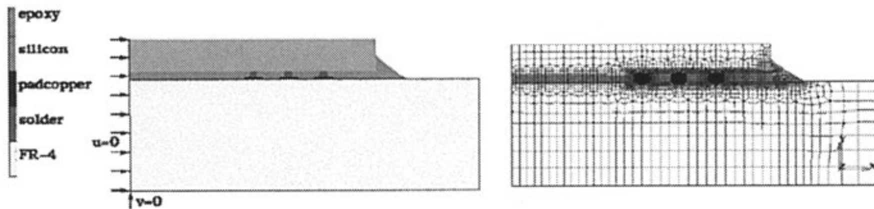


Fig. 15: Flip chip demonstrator configuration + FEM mesh used

The configuration is assumed to be stress free at the start of the curing process (ξ=0). The curing temperature is 130 °C. The curing epoxy underfill (FP4526) is described with the present constitutive model. For the solder (Pb95Sn5) an elastic-plastic rate dependent creep model is used. Isotropic linear material properties are used for the chip (silicon) and the copper pad. The (FR-4) substrate is described with an orthotropic linear model.

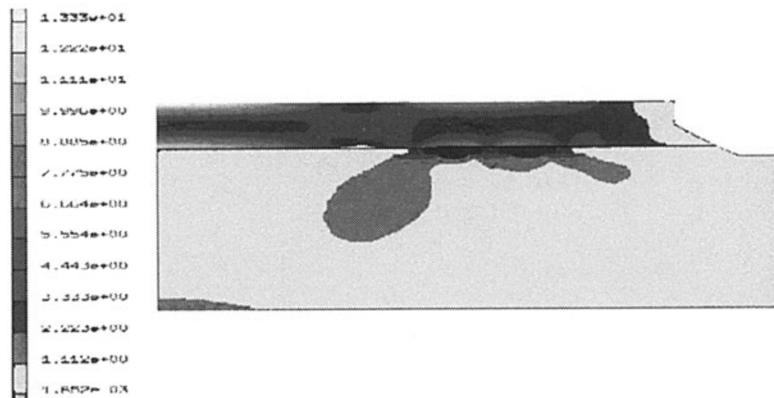


Fig. 16: Von Mises stress distribution (60 min)

The Von Mises (residual-) stress distribution just after the curing process (60 min) is presented in Fig. 16. The evolution of the stress and strain components and the equivalent stress and strain (σ_i, ε_i) during the curing process are presented in Fig. 17 for a chosen point in the epoxy at the right hand side of the right solder bump. The maximum stress level is quite high (2.4 Mpa), in view of the low ultimate strength during the curing process. In a "constrained shrinkage"-test on curing epoxy in the test facility, failure occurred in about 16-17 minutes at an axial stress level of about 2.3 Mpa.

المنارة للاستشارات
ww

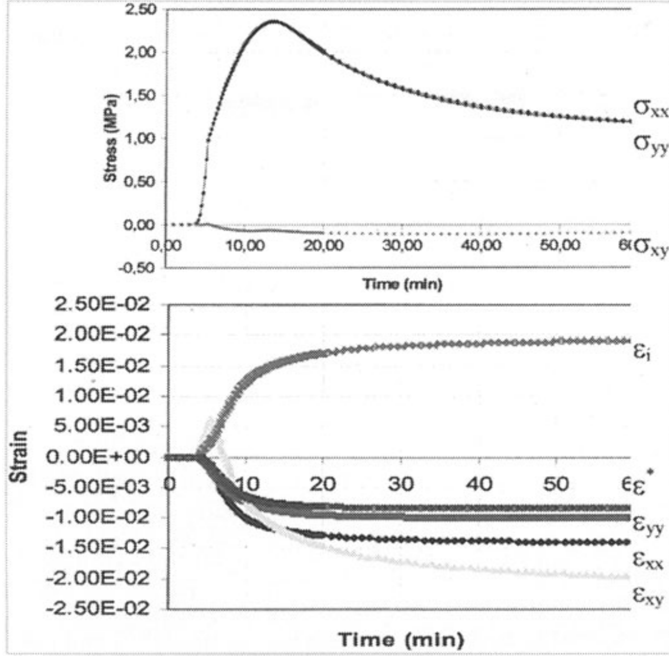


Fig. 17: Stress and strain evolution in epoxy (point A).

3.6 CURE INDEPENDENT LINEAR VISCOELASTICITY

Let we now further focus on cured polymers. We then could apply the earlier derived expression for "Linear state dependent visco-elasticity" (9), where (if we again assume isotropy) the relaxation functions, (C_{ij}), depend on 2 parameter-functions only:

$$S_i(t) = \int_{\xi=\xi_o}^{t} C_{ij}[\alpha^l(\xi),..,(t-\xi)] \cdot \left\{ \left(E_{j,\xi}\right)_\xi - \left(E^*_{j,\xi}\right)_\xi \right\} d\xi \tag{29}$$

$$C_{ij}[\alpha^l(\xi),..,(t-\xi)] = K[\alpha^l(\xi),..,(t-\xi)] \cdot V_{ij} + G[\alpha^l(\xi),..,(t-\xi)] \cdot D_{ij} \tag{30}$$

In principal, changes in state, such as due to temperature fluctuations and aging, could be described by the state parameters $\alpha^l(\xi)$ and possible changes in initial strain E^*_j. Further, in case that no changes in state, such as temperature, degree of cure, are necessary, the well-known expression for "linear viscoelasticity" is obtained:

$$S_i(t) = \int_{\xi=\xi_o}^{t} C_{ij}[.,.,(t-\xi)] \cdot \left\{ \left(E_{j,\xi}\right)_\xi - \left(E^*_{j,\xi}\right)_\xi \right\} d\xi \tag{31}$$

A very important state parameter influencing the viscoelastic behavior is the temperature.

المنارة للاستشارات

ww

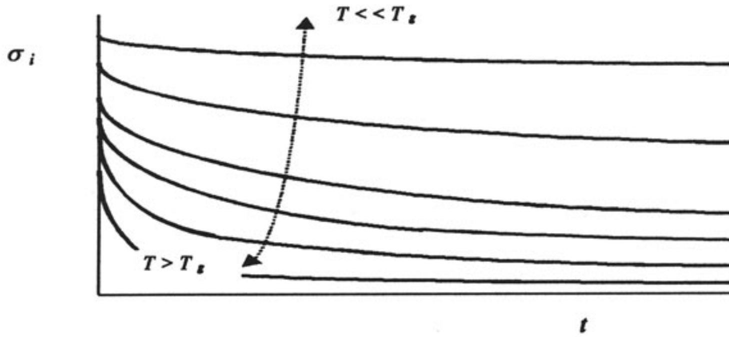


Fig. 18: Relaxation curves at different temperatures

Fig. 18 illustrates characteristic relaxation behavior at different temperatures (on the same step deformation). The upper curves demonstrate possible behavior for low temperatures (below T_g): Characteristics are high initial stiffness and relatively large relaxation times. The lower curves demonstrate possible behavior for high temperatures (above Tg) with characteristics as low initial stiffness and relatively (very) small relaxation times.
Many polymers behave so-called "rheologically simple". The Prony series of the relaxation moduli can then be written as follows:

$$C_{ij}(t,T) \approx \sum_{n=1}^{N} C_{ij}^{n}(T) \cdot \mathrm{e}^{-t/\tau_n(T)} \tag{32}$$

The initial stiffness of cause is temperature dependent. This dependency can be described by the participation factors: $C^n{}_{ij}=C^n{}_{ij}(T)$. The relaxation times are also temperature dependent; The relaxation times for temperature T can here be obtained from the relaxation times for a reference temperature T_o by simply scaling with a factor a, which is a function of (T, T_o):

$$\tau_n(T) = \frac{\tau_n(T_o)}{a(T,T_o)} \tag{33}$$

If we consider the (normalized) relaxation modulus versus log (time), this scaling of relaxation times results into a shift on the log (time) axis.

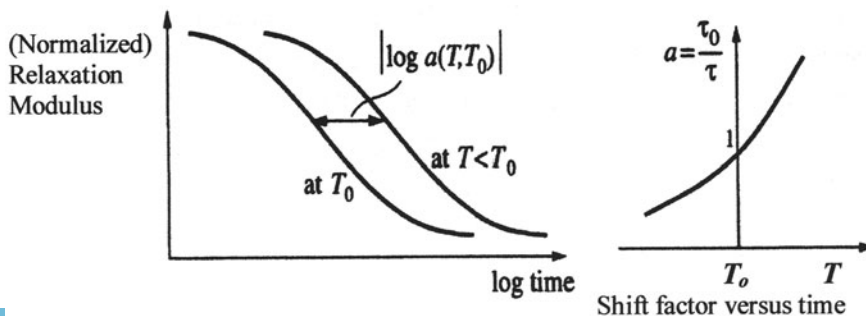


Fig. 19: Time shift of relaxation modulus

المنارة للاستشارات

This shifting behavior makes it possible to construct a so-called Master curve for the reference temperature. For this purpose relaxation moduli functions are experimentally obtained for different temperatures, over a limited period of time. These functions can be shifted over the time axis to construct a relaxation curve, which can be used for a much larger time period. The actual procedure for constructing the Master curve and using the concept can be found in many textbooks on viscoelasticity.

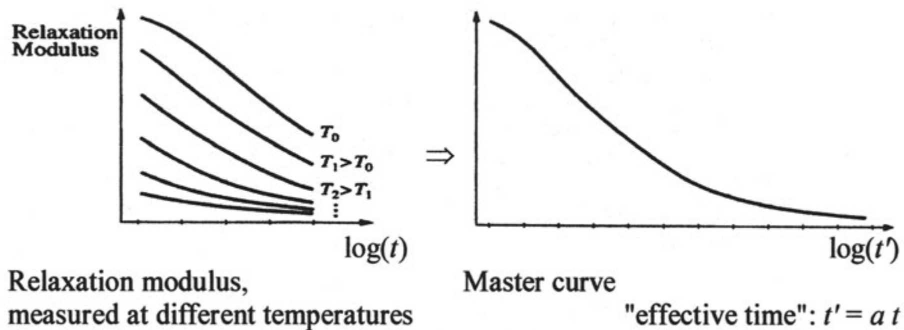


Fig. 20: Master curve concept

An analogous "master curve" construction was proposed by Struik to account for "ongoing physical aging" (Struik, 1977). For various polymers the relaxation functions appear to reach "equilibrium values" after some period of stress relaxation.

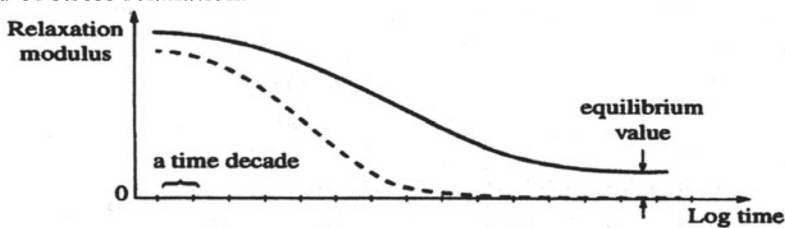


Fig. 21: Equilibrium modulus concept.

These "equilibrium values" of the relaxation modulus are often referred to as "the equilibrium moduli". In rough approximations of reality these "equilibrium moduli" are sometimes used in "linear elastic" simulations. The effects of the transient behavior are then completely omitted.
Quite the opposite of using the rough approximations with "equilibrium moduli" is the application of "non-linear" viscoelastic models. Some notes on such models are given in section 2.6.

3.7 NON- LINEAR VISCOELASTICITY

Non-linear models are much more complicated. Further they require much more experimental data for the parameter investigation. The practical application of non-linear modeling for packaging materials therefore is very

المنارة للاستشارات
ww

limited. However, the use of non-linear viscoelastic models could be advisable as: •High stresses or strains (> 50% of the ultimate strength) occur within the polymers. •A relatively high filler percentage is present (> 50 % volume). •Cyclic loading with many cycles has to be considered. However, in these cases not all non-linear models appear to give correct results.
Mostly non-linear modeling starts out from a "creep formulation" rather than a "relaxation formulation" as we have used for the linear-viscoelastic modeling. It is beyond the scope of this paper to discuss all possible non-linear viscoelastic theories being available. We will restrict the discussion to few well-known non-linear models.
One of these is Shapery's model (Schapery, 1969 and later): This model is rather complicated since it uses 4 non-linearizing functions:

$$\varepsilon(t) = g_o J_o \sigma(t) + g_1 \int_0^t \Delta J[\psi(t) - \psi'(\xi)] \cdot \left\{ \frac{d(g_2 \sigma)}{d\xi} \right\}_{\xi} d\xi \tag{34}$$

J_o represents the elastic compliance. The non-linearizing functions g_o, g_1, g_2, and a_σ are stress- and stress history dependent. They can be determined experimentally. The kernel ΔJ is a function of shifted times $\psi(t)$ and $\psi'(\xi)$:

$$\psi(t) = \int_0^t \frac{dt'}{a_\sigma} \qquad \psi(\xi) = \int_0^{\xi} \frac{dt'}{a_\sigma} \tag{35}$$

Generally used kernel functions are a generalized Kelvin model and a power-law creep formulation:

$$\Delta J[\psi] = C \cdot \psi^n \quad \text{(power-law creep)} \tag{36}$$

$$\Delta J[\psi] = \sum_{i=1}^{N} J_i (1 - e^{-\psi/\tau_i}) + J_s \cdot \psi \tag{37}$$

These formulations appear to be less adequate for unloading situations (Hadley and Ward, 1975, Ward and Onat, 1963, Rendell *et al.* 1987). The parameter function investigation requires a large number of creep-recovery experiments at different load levels. Borderline cases are:
•Bolzman's superposition integral when: $g_o = g_1 = g_2 = a_\sigma = 1$
•Leaderman's Modified Superposition Method (MSM): $g_o = g_1 = a_\sigma = 1$, $g_2 \neq 1$
Zhang presented a less complicated and more accurate non-linear model (Zhang, et al., 1997 and later). It originally is based on Shapery's model, but it has 2 non-linearizing functions only ($g_1 = 1$, $a_\sigma \approx 1$). The kernel function is exponential:

$$\Delta J[\psi] = C[e^{\psi^n} - 1] \tag{38}$$

Struik also used such an exponential expression in his studies on aging implementation (Struik, 1977 and later). Zhang's model uses an elastic and a (constant) creep Poisson's ratio for the 3D extension of his theory. Simple

المنارة للاستشارات

1D creep-recovery tests (at different stress levels) are sufficient to establish the parameters of the visco-elastic model.

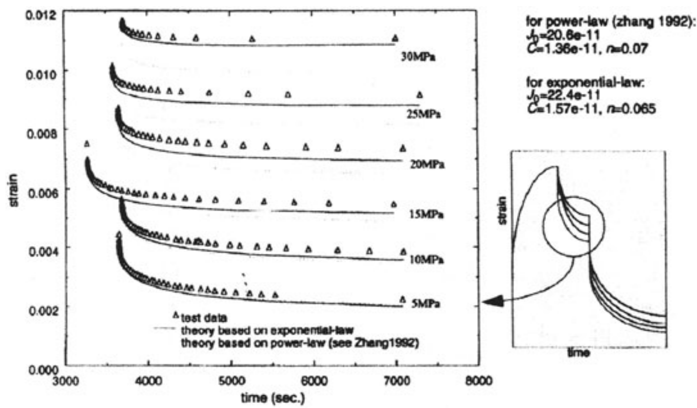


Fig. 22: 2-step creep recovery data on polyester (1st step 40 Mpa)

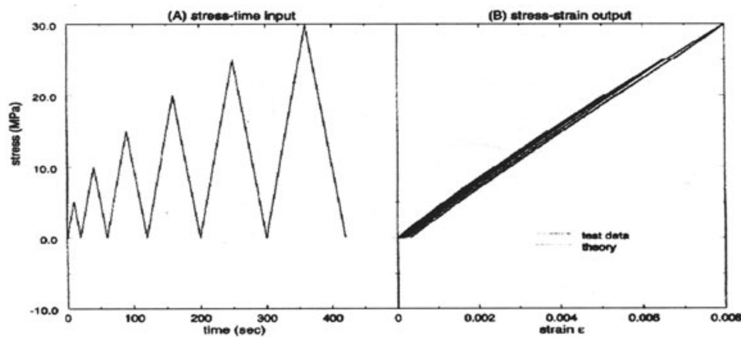


Fig. 23: Cyclic loading: Comparison between experiment and model

The model gives an excellent description of cyclic loading situations. The investigation of the elastic and creep-Poisson's ratio's requires unit-step creep recovery experiments where both the longitudinal and the lateral strains are measured (see Fig. 24 and 25).

Zhang's model was combined with Struik's aging theory to include the effects of ongoing aging. Details can be found in (Zhang, 1997). An illustration is presented in Fig. 26-27.

Finally we mention Lai's theory (:ai,1995), where plasticity was combined with Schapery's non-linear viscoelastic model, in order to appropriately describe elasto-visco-plastic material behavior.

المنارة للاستشارات
ww

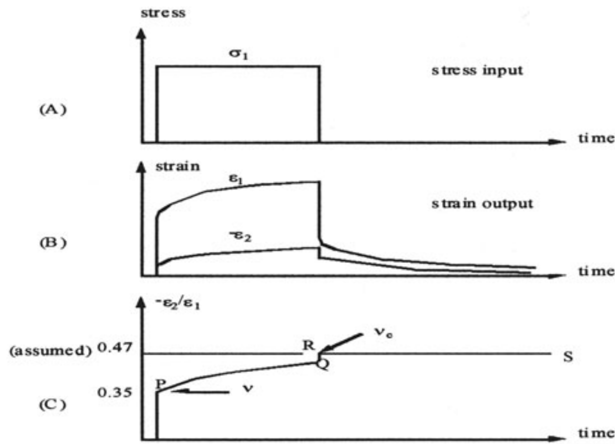


Fig. 24: Principle of investigation of elastic- and creep Poisson's ratios

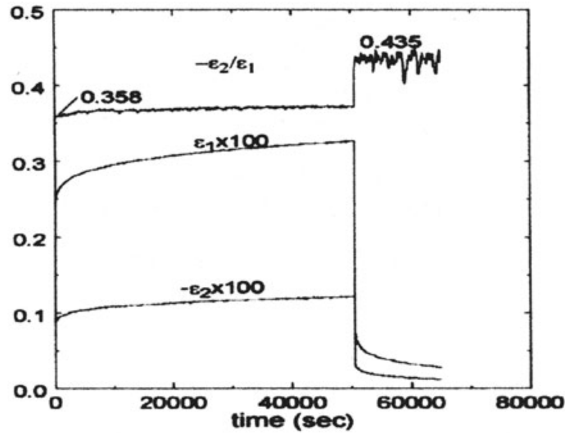


Fig. 25: Measurement results for the creep Poisson's ratio investigation.

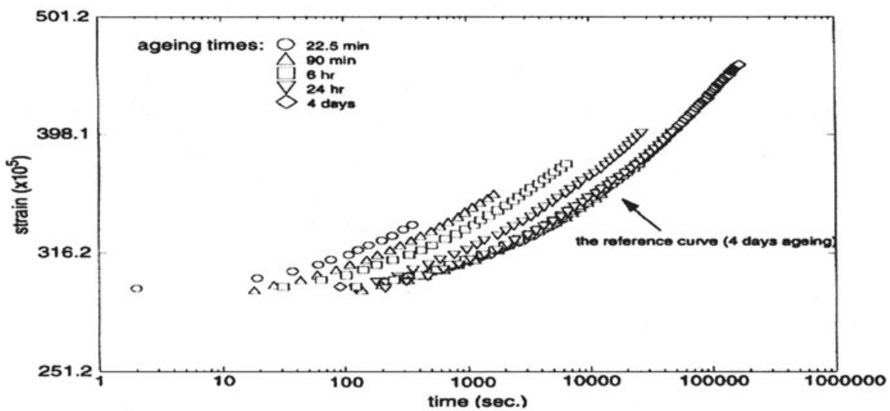


Fig. 26 Horizontal shifting of creep data for different aging times forming a mastercurve, corresponding to 4 days aging time. (10 Mpa)

المنارة للاستشارات

ww

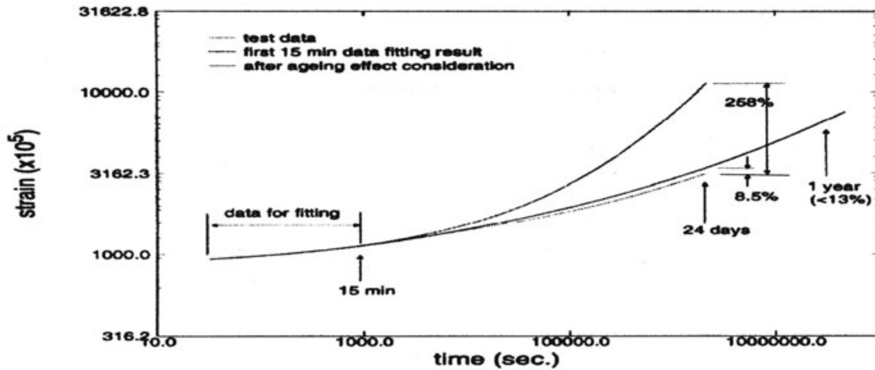


Fig. 27 Aging influence on long-term creep (30 Mpa)

REFERENCES

Brouwer, R., *"Nonlinear Viscoelastic Characterization of Transversely Isotropic Fibrous Composites under Biaxial Loading",* Ph.D. Dissertation, Free University of Brussels, Belgium 1986.

Ernst, L.J., "*Polymer material characterization and modeling*", Lecture notes of EuroSim2000 workshop, Organized by Compete, Eindhoven, March 2000

Ernst, L.J., Hof, C. van 't , Yang, D.G., Kiasat, M.S., Zhang, G.Q., Bressers, H.J.L., Caers, J.F.J., Boer, A.W.J. den, Janssen, J., *"Mechanical characterization and simulation of curing packaging polymers"* , 3rd International Conference and Exhibition on Micro Materials, MicroMat 2000,.Berlin, April 2000.

Ernst, L.J., Hof, C. van 't , Yang, D.G., Kiasat, M.S., Zhang, G.Q., Bressers, H.J.L., Caers, J.F.J., Boer, A.W.J. den, Janssen, J., *"Determination of visco-elastic properties during the curing process of encapsulate materials",* 50th Electronic Components and Technology Conference (ECTC), Las Vegas, May 2000

Hof, C. van 't , Yang, D.G., Kiasat, M.S., Ernst, L.J., Zhang, G.Q., Bressers, H.J.L., Caers, J.F.J., Boer, A.W.J. den, Janssen, J., *"Thermo-Mechanical Characterization of Packaging Polymers during the Curing Process",* Poly'99, Paris, December 1999

Hof, C. van 't, Ernst, L.J., *"Constitutive Modelling of the curing process of encapsulate polymers for electronic packages".,* 8th Int. UMBC Symp. "Plasticity 2000", Canada, July 2000.

Kiasat, M.S., "shrinkage and stress build up in unsaturated polyester resin during curing", In Polymers and Ceramics, vol. 2, 1997, pp. 95-102

المنارة للاستشارات

ww

Kiasat, M.S., "Curing Shrinkage and Residual Stresses in Viscoelastic Thermosetting Resins and Composites", Ph.D. thesis, Delft University of technology, Delft 2000.

Kiasat, M.S., Marissen, R, *"Modeling the cure-dependent viscoelastic behavior of a thermosetting resin and residual curing stresses"* , 3rd International Conference and Exhibition on Micro Materials, MicroMat 2000,.Berlin, April 2000.

Lai, Junbiao, *"Non-linear Time-dependent Deformation Behavior of High Density Polyethylene"*, Thesis Delft University of Technology, 1995, ISBN 90-407-1136-4, NUGI 841.

Leaderman, H., *"Elastic and Creep Properties of Filamentous Materials and Other High Polymers"*, The Textile Foundation, Washington, DC, 1943.

Schapery, R.A., *"On the Characterization of Nonlinear Viscoelastic Materials"*, Polymer Eng. & Sci., Vol. 9, No. 4, pp 295-310, 1969

Struik, L.C.E., *"Physical aging in amorphous polymers and other materials"*, Ph.D. Thesis, Delft University of Technology, Delft 1977, The Netherlands

Struik, L.C.E., *"Physical aging in amorphous polymers and other materials"*, Elsevier Sci. Publ., Amsterdam 1978

Struik, L.C.E., "Internal Stress, Dimensional Instabilities and Molecular Orientations in plastics", John Wiley & Sons Ltd., England, p. 31

Ward, I.M., Onat, E.T., *"Non-linear mechanical behavior of oriented polypropylene"*, J. Mech. Phys. Solids 19, pp. 201-214, 1963

Zhang, L., Ernst, L.J., "*Implementation of a Nonlinear Viscoelastic Theory*", Proc. of the 2nd Int. DIANA Conference on Finite Elements in Engineering and Science, Amsterdam June 1997, Balkema, Rotterdam, ISBN 9054108835

Zhang, L., Ernst, L.J., Brouwer, H.R., "*A study of nonlinear viscoelasticity of an unsaturated polyester resin, part 1, uniaxial model*", In Mechanics of Materials, Vol. 26, No. 3, October 1997, pp. 141-166,

Zhang, L., Ernst, L.J., Brouwer, H.R., "*A study of nonlinear viscoelasticity of an unsaturated polyester resin, part 2, 3D-model*", In Mechanics of Materials, Vol. 26, No. 3, October 1997, pp. 167-195.

Zhang, L., Ernst, L.J., Brouwer, H.R., Transverse Behavior of Unidirectional Composite (Glass Fibre Reinforced Polyester), *Part 1., Fiber Packing Geometry Influence*", In Mechanics of Materials, Vol. 27 (January 1998), pp. 13-36.

Zhang, L., Ernst, L.J., Brouwer, H.R., "Transverse Behavior of Unidirectional Composite (Glass Fiber Reinforced Polyester), *Part 2., Initial Strain Influence*", In Mechanics of Materials, Vol. 27 (January 1998), pp. 37-61.

المنارة للاستشارات

ww

GENERIC ISSUES IN NUMERICAL MODELLING

S.H.Pulko
School of Engineering
University of Hull
Cottingham Rd
Hull HU6 7RX
UK

INTRODUCTION.

Numerical simulation of thermal and mechanical behaviour of microelectronic systems and subsystems is an increasingly important aspect of their design; appropriate simulation is a vital component of any 'right first time' philosophy. Many excellent numerical modelling packages exist which enable simulation of thermal and mechanical behaviour, but the usefulness of simulation results depends crucially on the users' understanding of the problem to be solved, the numerical issues involved and the context in which the simulation results are to be used. This paper summarises issues to be considered before simulations are performed and then proceeds to consider specific issues in the choice of discretisation of space and time.

ISSUES FOR CONSIDERATIONS.

Steady State and Transient Models.

One issue to be considered very early in the simulation process is whether, for any particular purpose, a steady state or a transient simulation is appropriate. Steady state simulations are appropriate to systems which have fully responded to the stimuli applied. Transient simulations relate to situations in which the experience of the system as it moves from initial conditions to its steady state is of interest; a major class of problems is those for which results are required at a particular time and there is no guarantee that, at the time of interest, the steady state will have been reached. Examples of steady state problems might include the temperature distributions in electronic circuits under continuous operation and some

G. Q. Zhang et al. (eds.), Benefiting from Thermal and Mechanical Simulation in Micro-electronics, 59–73.
© 2000 *Kluwer Academic Publishers.*

المنارة للاستشارات
ww

deformation problems. Examples of situations requiring transient analysis might be thermal run-away, analysis of stresses as the result of a system's being dropped, and thermal transients during the range of soldering processes.

It is possible to identify a further class of problems which require a series of steady state analyses to be performed, each analysis taking at its starting point the field distribution resulting from the steady state response predicted by the previous simulation. An example here might be developed stresses and deformations during the processing sequence of a microelectronic system, so long as the system responds fully in terms of stresses an deformations to each stage of the sequence before the next stage takes place.

The Number of Spacial Dimensions Needed to Represent the Situation.

Although a physical situation may be 3-dimensional symmetry can render it either 1- or 2-dimensional. This has enormous advantages in terms of run time since the number of spacial elements is drastically reduced and the computation associated with each element is also reduced; the precise reductions in run time are dependent upon the nature of the numerical core of the software used.

A situation in which an obvious and valid reduction in the dimension of a problem is the dissipation of heat generated in a cylindrical component. Here, the 3-dimensional geometry, because of its radial symmetry can validly be represented in 2-dimensions. However, it is important to consider not only the symmetry of the geometry but also the symmetry of spacial boundary conditions. In some situations run time considerations mean that it is not feasible to model a situation in 3-dimensions, and no symmetry can be identified. In this type of situation the influence of 'edge effect' on simulation results should be born in mind.

Even when it is imperative to model a 3-dimensional situation in 3-dimensions it is sometimes possible to identify planes of symmetry which, while not reducing the dimension of the problem, significantly reduce the physical size of the problem and the number of elements. Figure 1 illustrates the geometry of a situation in which two regions of metallic film on an insulating substrate are sequentially illuminated by a laser. The temperature some distance from the heated regions is known to be constant and interest

المنارة للاستشارات
ww

lies in the temperature-time history of both illuminated regions and of the unilluminated region between them. Although a 3-dimensional simulation is necessary there is a clear plane of symmetry, which runs through the centre of both illuminated regions, and recognition of this allows the number of elements involved in the simulation to be approximately halved.

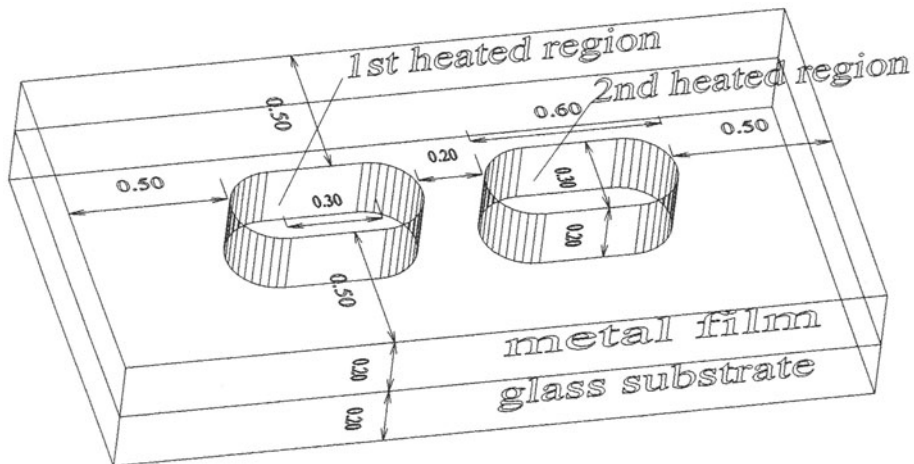


Figure 1. Situation with a plane of symmetry which reduces the number of elements necessary in the 3-dimensional model.

Linear and Nonlinear Simulations

In linear simulations the physical properties and geometry of the materials being simulated remain constant; in nonlinear simulations they are not constant and may even be dependent upon the field being modelled eg thermal conductivity may be dependent on temperature. Nonlinear models of transient situations can have particularly long run-times. This is highly dependent on the numerical approach taken and the explicit finite difference and TLM (transmission line matrix) approaches tend to be more efficient for this type of problem than the implicit finite element approaches. Coupled problems such as the diffusion of moisture are, perhaps, a very extreme case of a nonlinear problem. Here it can be necessary to model both heat and moisture transfer, with appropriate coupling between the processes. This type of model is beyond the scope of many standard packages but can be accomplished, explicit finite difference and TLM based approaches being particularly appropriate.

المنارة للاستشارات

ww

Another difficulty with nonlinear problems is that, although the problem may be known to be nonlinear, the precise nature of the nonlinearity may not be known. In this instance it is at least possible to perform simulations to establish how sensitive the system is to the nature of the nonlinearity and hence obtain increased insight into the situation.

Discretistion.

One of the tasks facing the modeller is choice of discretisation of space and, for transient problems, of time. Many modelling packages have an element of automatic spacial meshing, and the user has the option of enhancing the mesh manually. There are, perhaps, two aspects of spacial meshing to be considered

- Representation of boundaries and interfaces
- Representation of the field of interest throughout the modelling volume.

The importance of good boundary and interface representation depends very much in the physical process being modelled. If we are simulating heat transfer, then precise representation of the boundary is probably not very important, although it is important that mismatching the boundary is not bringing about a significant change in the volume being modelled. This can be particularly important when cylindrical symmetry has been used to reduce the dimension of the problem, since outer elements have very large volumes compared to inner volumes and volume discrepancies are not always immediately obvious by visual inspection. In terms of heat transfer across boundaries and interfaces it is important to consider the area of contact. However, discrepancies in contact area can often be compensated by appropriate modification of the heat transfer coefficient. Good matching of spacial boundaries is, however, essential for wave based models such as models of stresses, since mismatching is associated with improper representation of wave phase and stress concentrations tend to occur in the region of boundaries and interfaces.

For transient models choice of timestep is important. Some simulation packages involve automatic timestepping procedures which allow the timestep to change during the simulation. However it is important that the timestep is small enough to represent the most rapidly changing part of the transient if this is of interest. If, on the other hand, this rapid part of the transient is not of direct interest, care must still be taken to ensure that errors accumulating do not affect results in a subsequent period which is of interest.

المنارة للاستشارات

ww

If properties are nonlinear it is also important to be aware that the speed of response will vary with properties and that the most rapid response will not necessarily occur at the start.

We now present an analysis of the effect of discretisation on the solution to heat transfer problems.

Analysis.

For the thermal case the 2-dimensional diffusion equation is

$$\frac{\partial^2 \Phi}{\partial x^2} + \frac{\partial^2 \Phi}{\partial y^2} = \frac{s\rho}{K_T}\frac{\partial \Phi}{\partial t} \tag{1}$$

where Φ is the temperature, s the thermal conductivity, ρ the density, and K_T the thermal conductivity. Using the familiar concept of separation of variables we have

$$X'' + \alpha^2 X = 0 \tag{2}$$

$$Y'' + \gamma^2 Y = 0 \tag{3}$$

and

$$\dot{T} + \sigma^2 \frac{K_T}{s\rho} T = 0 \tag{4}$$

Since equations 2, 3 and 4 are ODEs, their solutions are conveniently calculable and, subject to the relevant boundary conditions, give solution in space and time of the diffusion equation. The continuous solution is determined by imposing boundary conditions on equations 2 to 4.

For the x-direction, the unforced solution is of the form

$$X(x) = \sum_{n_x=0}^{\infty} A_{n_x} \sin \alpha x + B_{n_x} \cos \alpha x \tag{5}$$

المنارة للاستشارات

ww

where α has an infinite series of values. These values are specified by the integer n_x and given by

$$\alpha = \sin \frac{n_x \pi}{l_x} \tag{6}$$

where $0 \le n_x \le \infty$ and l_x is the physical dimension of the region being represented. The general solution is, therefore, a summation of terms having the form

$$A_{n_x} \sin \frac{n_x \pi}{l_x} x + B_{n_x} \cos \frac{n_x \pi}{l_x} x \tag{7}$$

The solution is has the form of a Fourier series and a maximum value of n_x must be chosen consistent with the required accuracy of the solution.

We now derive general discrete solutions to ODEs having the form of equations 2 to 4. The general form of the solution to a second order ODE is

$$mD^2x + bDx + kx = F \tag{8}$$

where m, b and k are constants, F is a driving term, and D the differential operator. Factorising equation 8 gives

$$(D - z)(D - \bar{z})x = \frac{F}{m} \tag{9}$$

where z and $\bar{z}$ are complex roots

$$z = -\frac{b}{2m} + j\frac{\sqrt{4km - b^2}}{2m} \tag{10}$$

and

$$\bar{z} = -\frac{b}{2m} - j\frac{\sqrt{4km - b^2}}{2m} \tag{11}$$

Let

المنارة للاستشارات

ww

$$d = \frac{b}{2m} \tag{12}$$

and

$$\omega = \frac{\sqrt{4km - b^2}}{2m} \tag{13}$$

Equation 9 can be treated as two first order ODEs,

$$(D - \bar{z})x = y \tag{14}$$

and

$$(D - z)y = \frac{F}{m} \tag{15}$$

Considering the first order solution of equation 15, the homogeneous, unforced solution is

$$y(t) = y(0)e^{zt} \tag{16}$$

and complete solution is

$$y(t) = y(0)e^{zt} + e^{zt}\int_0^t e^{-z\Gamma}\frac{F(\Gamma)}{m}d\Gamma \tag{17}$$

Equations 16 and 17 are continuous solutions. To find the discrete solution we need a recurrence formula to relate the solution at any time to the solution at a previous time.

To find the recurrence formula, let $t = n\Delta t$ so that equation 16 can be rewritten

$$y(n\Delta t) = y(0)e^{zn\Delta t} \tag{18}$$

This gives the recurrence relationship

$$y(0\ ) = const$$

المنارة للاستشارات

ww

and $$y((n+1)\Delta t)=e^{z\Delta t}y(n\Delta t) \tag{19}$$

or $$y((n+1)\Delta t)-e^{z\Delta t}y(n\Delta t)=0 \tag{20}$$

The effect of the driving term between nΔt and $(n+1)\Delta t$ can be found from equation 20 by substituting for $y(n\Delta t)$ and $y(n+1)\Delta t$ from equation 17. Hence,

$$y((n+1)\Delta t)-e^{z\Delta t}y(n\Delta t)=\left[e^{z(n+1)\Delta t}y(0)+e^{z(n+1)\Delta t}\int_0^{(n+1)\Delta t}e^{-z\Gamma}\frac{F(\Gamma)}{m}d\Gamma\right]$$
$$-e^{z\Delta t}\left[e^{zn\Delta t}y(0)+e^{zn\Delta t}\int_0^{n\Delta t}e^{-z\Gamma}\frac{F(\Gamma)}{m}d\Gamma\right] \tag{21}$$
$$=e^{z(n+1)\Delta t}\int_{n\Delta t}^{(n+1)\Delta t}e^{-z\Gamma}\frac{F(\Gamma)}{m}d\Gamma$$

The recurrence relationship becomes

$$y((n+1)\Delta t)=e^{z\Delta t}y(n\Delta t)+\left[e^{z(n+1)\Delta t}\int_{n\Delta t}^{(n+1)\Delta t}e^{-z\Gamma}\frac{F(\Gamma)}{m}d\Gamma\right] \tag{22}$$

if a driving term is present. Equation 22 is an exact representation since $F(\Gamma)$ is within the integral. If we hold $F(\Gamma)$ constant from time nΔt to $(n+1)\Delta t$ then

$$F(\Gamma)=F_c(n\Delta t)$$

so that the driving term in equation 22 is

$$\frac{F_c(n\Delta t)}{m}e^{z(n+1)\Delta t}\int_{n\Delta t}^{(n+1)\Delta t}e^{-z\Gamma}d\Gamma$$

Simplifying

المنارة للاستشارات
www

$$\frac{F_c(n\Delta t)}{m}\frac{e^{z\Delta t}-1}{z}$$

Substituting for the second term on the right hand side of equation 22, gives a discrete recurrence relationship of

$$y((n+1)\Delta t) = e^{z\Delta t}y(n\Delta t) + \frac{F_c(n\Delta t)}{m}\frac{e^{z\Delta t}-1}{z} \qquad (23)$$

which is the discrete first order solution of equation 15.

Considering equation 14, the solution is

$$x(t) = x(0)e^{\bar{z}t} + e^{\bar{z}t}\int_0^t e^{-\bar{z}\Gamma}y(\Gamma)d\Gamma \qquad (24)$$

and, by analogy with the solution to equation 15, the recurrence formula is

$$x((n+1)\Delta t) = e^{\bar{z}\Delta t}x(n\Delta t) + e^{\bar{z}(n+1)\Delta t}\int_{n\Delta t}^{(n+1)\Delta t} e^{-\bar{z}\Gamma}y(\Gamma)d\Gamma \qquad (25)$$

To find the complete solution of equation 8 we need to evaluate the second term on the right hand side of equation 25. Here, $y(\Gamma)$ is part of the integral and, if we assume linearity with respect to time over a period of $2\Delta t$, we can apply the integral to equation 22, the recurrence equation for y. Hence

$$e^{\bar{z}(n+1)\Delta t}\int_{n\Delta t}^{(n+1)\Delta t} e^{-\bar{z}\Gamma}y(\Gamma)d\Gamma = e^{z\Delta t}e^{\bar{z}n\Delta t}\int_{(n-1)\Delta t}^{n\Delta t} e^{-\bar{z}\Gamma}y(\Gamma)d\Gamma + e^{\bar{z}(n+1)\Delta t}\int_{n\Delta t}^{(n+1)\Delta t} e^{-\bar{z}\Gamma}e^{z\Gamma}\int_{(\Gamma-\Delta t)}^{\Gamma} e^{-z\mu}\frac{F(\mu)}{m}d\mu d\Gamma$$

(26)

From equation 25,

$$e^{\bar{z}(n+1)\Delta t}\int_{n\Delta t}^{(n+1)\Delta t} e^{-\bar{z}\Gamma}y(\Gamma)d\Gamma = x(n\Delta t) - e^{\bar{z}\Delta t}x((n-1)\Delta t)$$

so that equation 26 becomes

المنارة للاستشارات

ww

$$e^{\bar{z}(n+1)\Delta t}\int_{n\Delta t}^{(n+1)\Delta t} e^{-\bar{z}\Gamma} y(\Gamma)d\Gamma = e^{z\Delta t}\left[x(n\Delta t) - e^{\bar{z}\Delta t}x((n-1)\Delta t)\right]$$
$$+ e^{\bar{z}(n+1)\Delta t}\int_{n\Delta t}^{(n+1)\Delta t} e^{-\bar{z}\Gamma} e^{z\Gamma}\int_{(\Gamma-\Delta t)}^{\Gamma} e^{-z\mu}\frac{F(\mu)}{m}d\mu d\Gamma \quad (27)$$

Substituting for the second term on the right hand side of equation 25 gives

$$x((n+1)\Delta t) = e^{\bar{z}\Delta t}x(n\Delta t) + e^{z\Delta t}x(n\Delta t) - e^{(z+\bar{z})\Delta t}x((n-1)\Delta t)$$
$$+ e^{\bar{z}(n+1)\Delta t}\int_{n\Delta t}^{(n+1)\Delta t} e^{-\bar{z}\Gamma} e^{z\Gamma}\int_{(\Gamma-\Delta t)}^{\Gamma} e^{-z\mu}\frac{F(\mu)}{m}d\mu d\Gamma \quad (28)$$

Equation 28 is the full continuous solution of equation 9. The discrete equivalent of equation 28 requires that $F(\mu)$ is constant across two timesteps, the period of the double integration. Hence,

$$x((n+1)\Delta t) = e^{\bar{z}\Delta t}x(n\Delta t) + e^{z\Delta t}x(n\Delta t) - e^{(z+\bar{z})\Delta t}x((n-1)\Delta t)$$
$$+\left[\frac{\left(e^{\bar{z}\Delta t}-1\right)}{\bar{z}}\frac{\left(e^{z\Delta t}-1\right)}{z}\right]\frac{F(n\Delta t)}{m} \quad (29)$$

We have now obtained both the continuous and the discrete versions of the solution of the first and second order ODEs which, together, can represent solutions to the diffusion equation.

From the continuous and discrete solutions to equation 8 we can write discrete solutions to equation 2. Considering the solution to equation 2 and using equations 8, 9, 10, and 11 with the definitions of equations 12 and 13, gives

$z = j\alpha$ and $\bar{z} = -j\alpha$

Substituting for z and $\bar{z}$ in equation 29 and writing in terms of X(x) rather than x(t) gives

المنارة للاستشارات
ww

$$X((n+1)\Delta x) = e^{-j\alpha\Delta x} X(n\Delta x) + e^{j\alpha\Delta x} X(n\Delta x) - X((n-1)\Delta x)$$

$$+\left[\frac{\left(e^{-j\alpha\Delta x}-1\right)}{-j\alpha}\frac{\left(e^{j\alpha\Delta x}-1\right)}{j\alpha}\right]\frac{F_X(n\Delta x)}{m} \qquad (30)$$

A similar expression can be derived for the solution of equation 3 which deals with the y-direction.

For equation 4, the single derivative with respect to time, equations 10 and 11 yield

$$z = +\frac{\sigma^2 K}{s\rho}$$

Substituting for z in equation 23 and writing in terms of T(t)

$$T((n+1)\Delta t) = e^{\frac{\sigma^2 K\Delta t}{s\rho}} T(n\Delta t) + F_T(n\Delta t)\frac{e^{\frac{\sigma^2 K\Delta t}{s\rho}}+1}{\frac{\sigma^2 K}{s\rho}} \qquad (31)$$

The discrete solutions derived represent the impulse invariant transformations of the continuous solutions, and correspond, therefore, to the best possible fit between the discrete and continuous solutions at the sampling points in space and time. They can also be presented in a form from which they can be expressed in terms of shift operators and, once this has been done, the recurrence relations can be compared directly with the differential equations they correspond to. For example, for X we can rearrange equations 30 so that

$$X((n+1)\Delta x) - \left(e^{-j\alpha\Delta x} + e^{j\alpha\Delta x}\right)X(n\Delta x) + X((n-1)\Delta x)$$

$$= \left[\frac{\left(e^{-j\alpha\Delta x}-1\right)}{-j\alpha}\right]\left[\frac{\left(e^{j\alpha\Delta x}-1\right)}{j\alpha}\right]F_x(n\Delta x) \qquad (32)$$

المنارة للاستشارات

ww

In operator notation this becomes

$$\left[I-\left(e^{-j\alpha\Delta x}+e^{j\alpha\Delta x}\right)+I^{-1}\right]X=\left[\frac{\left(e^{-j\alpha\Delta x}-1\right)}{-j\alpha}\right]\left[\frac{\left(e^{j\alpha\Delta x}-1\right)}{j\alpha}\right]F_X \quad (33)$$

or

$$\left[\frac{\left(I-2-I^{-1}\right)}{(\Delta x)^2}\right]+\left[\frac{\left(2-e^{-j\alpha\Delta x}-e^{j\alpha\Delta x}\right)}{(\Delta x)^2}\right]X$$

$$=\left[\frac{\left[\frac{\left(e^{-j\alpha\Delta x}-1\right)}{-j\alpha}\right]\left[\frac{\left(e^{j\alpha\Delta x}-1\right)}{j\alpha}\right]}{(\Delta x)^2}\right]F_X \quad (34)$$

where I is the shift operator in the x-direction. The corresponding continuous ODE is

$$X''+\alpha^2 X=F_X \quad (35)$$

which is equation 2 in the presence of a driving term. A similar expression can be derived for the y direction to represent the solution for to equation 3.

Finally, for the derivative with respect to time, if Q is the time shift operator, equation 31 becomes

$$\left[\frac{Q-1}{\Delta t}+\frac{1-e^{\frac{-\sigma^2 K_T\Delta t}{s\rho}}}{\Delta t}\right]T=\frac{\left[\frac{e^{-\frac{\sigma^2 K_T\Delta t}{s\rho}}-1}{-\frac{\sigma^2 K_T}{s\rho}}\right]}{\Delta t}F_T \quad (36)$$

where Q is the time shift operator. Equation 36 corresponds to

المنارة للاستشارات

ww

$$\dot{T} + \frac{\sigma^2 K_T}{s\rho} T = F_T \tag{37}$$

Comparing equation 34 with 35 we have

$$\alpha^2 \approx \frac{\left(2 - e^{-j\alpha\Delta x} - e^{j\alpha\Delta x}\right)}{(\Delta x)^2} \tag{38}$$

Similarly, for the y direction, we have

$$\gamma^2 \approx \frac{\left(2 - e^{-j\gamma\Delta y} - e^{j\gamma\Delta y}\right)}{(\Delta y)^2} \tag{39}$$

and comparing equation 36 with 37 gives

$$\sigma^2 \frac{K_T}{s\rho} \approx \frac{1 - e^{\frac{-\sigma^2 K_T \Delta t}{s\rho}}}{\Delta t} \tag{40}$$

Plots of α^2, γ^2 and $\sigma^2 \frac{K_T}{s\rho}$ against $\frac{1}{\Delta x}$, $\frac{1}{\Delta y}$, and $\frac{1}{\Delta t}$ respectively can be used to track the relationship between the discrete and continuous coefficients as discretisation changes. The plots have the general form suggested schematically in figure 2, and the vertical distance between the curve and the horizontal line to which it is asymptotic relates to the accuracy of solution, being a direct reflection of the error in the interpolation coefficient of the discrete model. The use of this approach in devising discretisation routines appropriate to particular problems is described elsewhere (1). However, it is clearly desirable to operate in a region in which the straight line representing the continuous coefficient is reasonably close to the curve corresponding to the impulse invariant response.

المنارة للاستشارات
ww

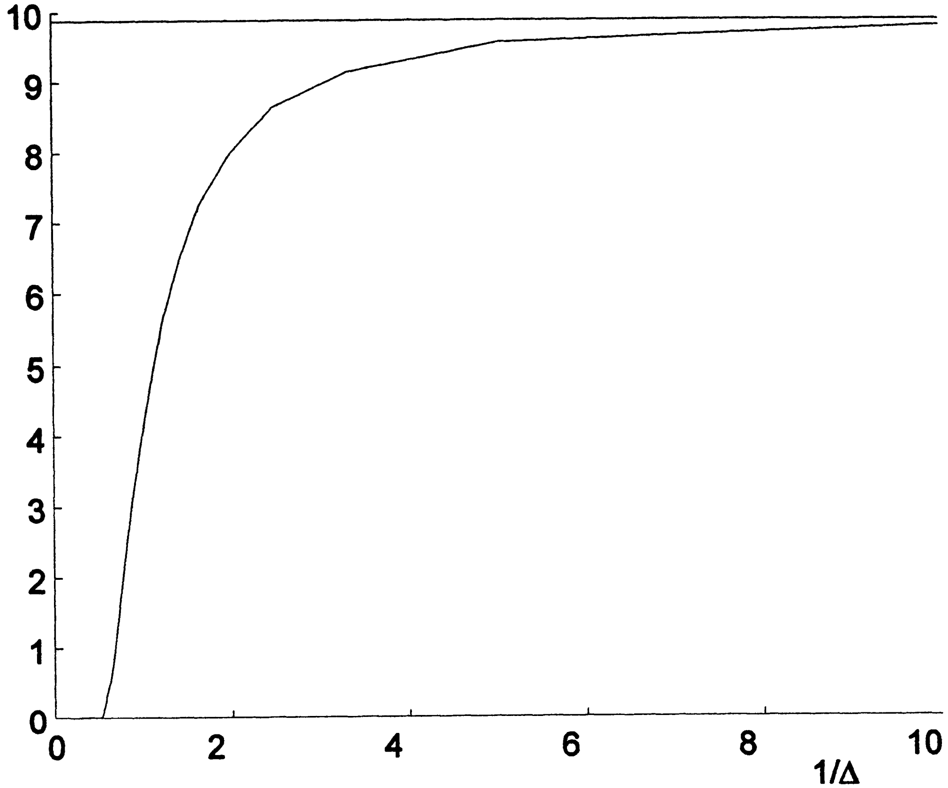


Figure 2. Typical form of the curves predicted by equations 38-40

DISCUSSION

A brief overview has been given of the issues to be considered when performing numerical modelling. They can, perhaps, be summed up in Confrontation 5 as devised by Bruce Irons and Nigel Shrive (2) and humorously expresses in 'old English':

'The virtue of numbers surpasseth not the thought that lieth behind them'.

المنارة للاستشارات

ww

Considering the effect of discretisation on the 'virtue' of results has yielded a method of monitoring the numerical error associated with a particular choice of discretisation. It is clear from the shape of the error curves that very large errors can exist if the spacial or temporal discretisation is inappropriate. In deciding what is appropriate the geometry, the physical properties and the nature of the excitation need to be considered (1). On the other hand too fine a mesh or, for a transient model, too short a timestep, can result in unreasonably long simulation run-times which in turn can make the whole approach of modelling less feasible. It is very much a matter of making reasonable, and informed, judgements based on the criteria of the overall simulation strategy.

The analysis presented here makes use of separation on variables and this limits the validity to models mounted on orthogonal meshes. However, as a matter of judgement, it might also yield a useful starting point for defining a suitable discretisation regime for models where this is not the case.

REFERENCES

1. Wilkinson AJ., Pulko SH, Saidane A, 'Discretisation effects in TLM diffusion models', Int. J. Numer. Modelling, 2000, 13: 59-74.

2. Irons BM, Shrive NG, *Numerical methods in engineering and applied science*, Ellis Horwood Ltd. (Wiley), 1987

المنارة للاستشارات

ww

MODELING OF VAPOR PRESSURE DURING REFLOW FOR ELECTRONIC PACKAGES

Xue-Jun Fan
Institute of Microelectronics
11 Science Park Road
Singapore Science Park II
Singapore 117685
Phone: (65) 7705448, Fax: (65) 7745747
Email: xuejun@ime.org.sg

Abstract

The plastic materials in electronic packages have high porosity and are susceptible to the moisture absorption. In order to investigate the moisture vaporization during reflow, the representative volume element (RVE) approach, by which the micro-void effect can be taken into consideration, is introduced in this paper. A theoretical model is established to calculate the whole-field vapor pressure in plastic materials. FEA models are built for both flip chip BGA (FCBGA) and wire bond PBGA packages to predict the moisture distribution, followed by the calculation of vapor pressure distribution in the package. Results show that the vapor pressure saturates much faster than the moisture diffusion, and a near uniform vapor pressure is reached in the package. The vapor pressure can't exceed the saturated pressure at the corresponding temperature, even when more moisture is added in. The vapor pressure introduces additional mismatch to the package, which is directly related to the vapor pressure distribution, rather than the moisture distribution. Moisture desorption during reflow is also studied and it has significant effect on the moisture distribution, but not on the vapor pressure distribution. A complete solution for the vapor pressure subjected to the interface after delamination is also derived.

1. INTRODUCTION

The moisture-induced failures, e.g., popcorn and delamination, of IC packages are common phenomenon during solder reflow. The failures are

G. Q. Zhang et al. (eds.), Benefiting from Thermal and Mechanical Simulation in Micro-electronics, 75–92.
© 2000 *Kluwer Academic Publishers.*

ww

due to sudden vaporization of moisture absorbed by the package at high temperature condition. Therefore, it is critical to evaluate the strength of internal vapor pressure generated in the package during reflow. The popcorn failure was first postulated by Fukuzawa et al. [2] in 1985, and later supported by many publications [1, 3, 6-8].

JEDEC standard [5] is widely used to conduct reliability test on moisture sensitivity of the electronic packages. Kitano et al. [6] showed that the package cracking is not controlled by the absolute water weight gain, rather it is due to the local moisture concentration at the critical interface. Therefore, the moisture diffusion modeling is required. However, the modeling of ensuing vapor pressure within the package during the reflow is the key element in understanding the failure mechanism. Previous researchers [3, 6-8] assumed that the delamination exists before the reflow, and considered the vapor pressure as traction loading subjected to the delaminated interfaces. Since the vapor pressure is generated anywhere in the package, it is necessary to investigate the whole field vapor pressure distribution before the package delamination.

2. VAPOR PRESSURE BEFORE DELAMINATION

For plastic materials such as mold compound, the saturated moisture concentration (C_{sat}) is a few orders larger than the corresponding saturated ambient water vapor density during the moisture preconditioning, e.g., ρ_g, at 85°C/85%RH (see Table 1). This implies that the moisture absorbed by plastic materials must be condensed into water in the micro-voids or free-volume of the materials. During the reflow, the moisture vaporizes at high temperature and produces internal vapor pressure. The vapor pressure, however, can not go beyond its saturated pressure as long as the moisture in the voids is not fully vaporized.

In order to estimate the vapor pressure generated inside the material, the Representative Volume Element (RVE) approach is introduced here. Let's take a very small representative material sample of volume dV, termed RVE as shown in Fig.1 [4]. From the microscopic level, the RVE is large enough to be statistically representative of the material properties at this location. Assume that the total volume of voids in this RVE is dV_f. Since the initial micro-voids are distributed randomly but uniformly in the material, the ratio $f_0 = dV_{f0}/dV$, defined as initial void volume fraction, is considered as a material property. The void volume fraction f at current state during reflow thus can be defined by

$$f = \frac{dV_f}{dV} \quad (1)$$

where dV_f is the current void volume and dV is the element volume.

المنارة للاستشارات

ww

The moisture in voids can be either at the single vapor phase or the mixture of water and vapor phases. At the single vapor-phase state, the vapor pressure p in voids obeys the ideal gas law, given as follows

$$\frac{p\mathrm{d}V_f}{T} = \text{const} \tag{2}$$

At the two-phase state, where moisture is present in the mixture of water and vapor, the vapor pressure p in voids is the saturated vapor pressure, as follows

$$p = p_g(T) \tag{3}$$

where, the saturated vapor pressure $p_g(T)$ as function of temperature, can be obtained from the steam table.

When the initial moisture state is at the mixture of water and vapor after the moisture preconditioning, the moisture may become a single vapor phase when temperature rises, or may still remain at the mixed vapor-liquid phase even at peak reflow temperature if the moisture content is large enough. A useful quantity, the moisture density in the voids to determine the moisture state, can be introduced as

$$\rho_m = \frac{dW_m}{dV_f} = \frac{dW_m / dV}{dV_f / dV} = \frac{C}{f_0} \tag{4}$$

where C is the local moisture concentration and dW_m is the moisture weight in a RVE.

The transition temperature, T_1, can be defined [1] as the temperature at which the moisture in the voids is fully transformed to vapor phase, as follows

$$\rho_m(x_i, T_0) = \rho_g(T_1) \tag{5}$$

where $\rho_g(T_1)$ is the saturated vapor density at temperature T_1, and T_0 is the preconditioning temperature at which the moisture is absorbed.

There are three distinct cases to compute the vapor pressure. The first case is when the moisture density in the voids is low enough such that all the moisture is at the single vapor phase at preconditioning temperature, T_0. In this case, $T_1 \le T_0$. The initial vapor pressure p_0 at T_0 can be determined as (see Appendix)

$$p_0(T_0) = \frac{p_g(T_0)C}{f_0 \rho_g(T_0)} \tag{6}$$

where ρ_g and p_g are the saturated vapor density and pressure respectively. Thus the ideal gas law is applicable from T_0 to reflow temperature T, at which the vapor pressure is

$$p = \frac{p_0(T_0)T}{T_0} = \frac{C\,p_g(T_0)T}{f\,\rho_g(T_0)T_0} \qquad \text{when } T_1 \le T_0 \tag{7}$$

In the second case, the moisture is initially two-phase state and will be fully vaporized at a temperature T_1 between the preconditioning temperature,

المنارة للاستشارات

T_0, and the peak reflow temperature T, i.e, $T_0 \le T_1 \le T$. At the transition temperature T_1, the moisture is fully vaporized and the vapor pressure is

$$p(T_1) = p_g(T_1) \quad (8)$$

From T_1 to T, the moisture is at the single-phase state. Therefore, the ideal gas law is used from T_1 to T, and the vapor pressure at T is

$$p = \frac{p_g(T_1)T}{T_1} \quad \text{when } T_0 \le T_1 \le T \quad (9)$$

For the last case, the moisture is not fully vaporized even at reflow temperature. This means $T_1 \ge T$. In this case, the vapor pressure is

$$p = p_g(T) \quad \text{when } T_1 \ge T \quad (10)$$

The initial void volume fraction, f_0, may be estimated from the material temperature-dependent moisture property. According to eq. (4), when saturated, the local concentration C is replaced by C_{sat}. So, the initial void volume fraction can be expressed as following

$$f_0 = \frac{C_{sat}}{\rho_m} \quad (11)$$

Since $\rho_m < 1.0 \text{g/cm}^3$, we have

$$f_0 > C_{sat}\big|_{100/100} = \rho_g \psi_0 \exp(Q_\psi / RT)\big|_{100/100} \quad (12)$$

Eq. (12) provides a simple way to predict the approximate magnitude of the voids existing in materials. The estimation is at the lower-limit since the moisture usually exists as a mixture of water and vapor in voids. Table 2 lists the results of the initial void volume fraction for some commonly used plastic materials in IC packages, computed by eq. (12), using the moisture property data given by Galloway et al. [3]. It shows that the values are usually between 1% and 5 %.

3. FINITE ELEMENT MODEL

In order to calculate the vapor pressure during reflow after the moisture preconditioning, it is necessary to know the moisture concentration distribution. The transient moisture diffusion can be described by Fick's Law as

$$\frac{\partial^2 C^2}{\partial x^2} + \frac{\partial^2 C^2}{\partial y^2} + \frac{\partial^2 C^2}{\partial z^2} = \frac{1}{\alpha_D}\frac{\partial C}{\partial t} \quad (13)$$

where C is the local moisture concentration (g/cm^3), x, y, z are the spatial coordinates (cm), α_D is the moisture diffusivity (cm^2/s), and t is the time (s). 2-D finite element models are generated for both unmolded FCBGA with depopulated array (156 I/O, 10x10mm chip, and 27x27mm BT substrate) and molded wire bond PBGA (35x35mm, 388 I/O, 2 layer). The element plots are shown in Fig 2 & 3 respectively. The moisture properties of

المنارة للاستشارات

www

diffusivity (D) and C_{sat} used in JEDEC level 1, 85%°C/85%RH, are listed in table 3. Die, copper, gold plate, and solder bump do not absorb moisture, and are assigned with very small values of diffusivity and solubility in analysis.

Moisture desorption is considered because moisture content in the package is lost during reflow, and this may affect the moisture and vapor pressure distributions. The desorption conditions analyzed is 2 minutes at 220°C, which is the worst case in the typical reflow temperature profile of moisture sensitivity test. The diffusivity values in desorption are assumed to be 100 times faster than in the moisture absorption at 85°C/85%RH [3].

After the moisture distribution is known, the vapor pressure can be calculated by the model proposed in section (2). The procedures to calculate the full-field vapor pressure during the reflow is summarized in Fig. 4.

4. RESULTS AND DISCUSSIONS

4.1 Transient Vapor Pressure Modeling

The finite element results of vapor pressure for FCBGA at level 1, reflow temperature of 220°C, are shown in Fig. 5, with different times of moisture absorption. The corresponding moisture diffusion distributions are shown in Fig. 6.

It is found that moisture diffusion and vapor pressure have different distributions. The vapor pressure in the package saturated much faster than the moisture diffusion. At level 1 condition (168 hours), the package is almost fully saturated with vapor pressure of 2.32 MPa, which corresponds to p_g at 220°C. Eq. (10) shows that the vapor pressure will remain its saturated pressure at 220°C when the transition temperature, T_1, is larger than 220°C. According to eq. (4), T_1 will be above 220°C as long as $\rho_m(T_0) \geq \rho_g(220°C) = 0.0116$ g/cm^3. Such a condition can be easily satisfied even though less moisture is absorbed. Similarly, the vapor pressure distributions at 220°C for wire bond PBGA at level 3 are given in Fig. 7, with different times of moisture absorption. The corresponding moisture diffusion distributions are shown in Fig.8.

4.2 Effect of Moisture Desorpion

The effect of moisture desorption during reflow on the vapor pressure distribution is studied. At high temperature, the moisture diffusivity for desorption is a few orders higher than the case for absorption at level 1. The consideration of moisture loss during reflow gives a more realistic vapor pressure calculation. Figs. 9-10 and 11-12 show the vapor pressure and moisture diffusion distributions of FCBGA and WBBGA before and after 2

المنارة للاستشارات

ww

minutes of desorption at 220°C. It is found that the moisture desorption during reflow affects the moisture distribution, but not the vapor pressure distribution, especially in the region near the package exterior. Though there is significant amount of moisture loss in the package during reflow, the moisture content is still large enough to produce the saturated vapor pressure at the reflow temperature.

4.3 Effect of Initial Void Volume Fraction

The results above are based on the estimation of initial void volume fraction (f_0) by eq. (12) which underestimates the f_0 values. It is important to understand the effect of f_0 on the vapor pressure distribution. Fig. 13 & 14 show the results for FCBGA and wire bond PBGA respectively, with different values of f_0. Although the f_0 is doubled, the vapor pressure distributions in both packages at 220°C after level-1 moisture absorption, are not much affected

When f_0 is doubled, the moisture properties (D and C_{sat}) are assumed to be unchanged. The saturated vapor pressure can still be maintained in the packages because there is enough moisture content in the micro-voids, even the moisture density is halved by doubling the f_0. By using eq. (4) with f_0 of 10% (worst case), the minimum C_{sat} required in the package to maintain the saturated pressure of 2.32MPa is 1.16 x 10^{-3} g/cm^3, which is much smaller than the C_{sat} for typical mold compound listed in Table 1. Even though the vapor pressure is not sensitive to f_0, but the local moisture concentration in the package increases with higher f_0, which weakens the interfacial adhesion and makes it more susceptible to failures.

4.4 Failure Mechanism

The moisture affects the package reliability at reflow from two aspects: generation of vapor pressure and degradation of interfacial adhesion. Fig. 15 shows the relative effect of moisture absorption on the interfacial adhesion and vapor pressure in the package. The interfacial adhesion is weakened with higher moisture level.

Previous results of vapor pressure estimation show that the saturated vapor pressure can be reached even with less moisture absorbed. For FCBGA and wire bond PBGA modeled here, the saturated pressure can be reached at the critical interface, even at the level 3 condition. However, the interfacial adhesion will be significantly decreased with more moisture absorption. When the adhesion strength is reduced to the level below the vapor pressure, delamination will occur. Therefore, the knowledge of material interfacial adhesion strength with moisture effect at high temperature condition is important in determining the failure criteria.

4.5 Vapor Pressure-Induced Expansion

المنارة للاستشارات
ww

The previous FEA results conclude that the vapor pressure saturated much faster than the moisture diffusion. This implies that the vapor pressure may be uniformly distributed in the plastic material regardless of moisture saturation. The Young's modulus of plastic material drops a few orders at the reflow temperature, thus the vapor pressure-induced expansion may become as important as thermal expansion. For instance, assume the Young's modulus of a typical underfill at 220°C is 500 MPa, and Poisson ratio is 0.3. Therefore, the volume change caused by vapor pressure, P_g(220°C) of 2.32 MPa, can be estimated as

$$\frac{\Delta V}{V} = \frac{3(1-2\nu)}{E} p = 5.568e-3 \qquad (14)$$

which is comparable with expansion due to the CTE thermal mismatch. It is obvious that the vapor pressure-induced expansion introduces additional mismatch. It must also be pointed out that such an expansion is directly related to the vapor pressure distribution, rather than the moisture distribution.

5. VAPOR PRESSURE AS EXTERNAL LOADING IN DELAMIANTATED AREA

The internal vapor pressure in voids at the interfaces instantaneously become an external pressure subjected to the delaminated interfaces when delamination is formed at reflow. By equation (1), the delamination is defined when f=1. Four cases can be identified to calculate the vapor pressure subjected to the delaminated area. The first case is when the moisture density in the voids is low enough before delamination such that all the moisture becomes vaporized at preconditioning temperature, T_0., i.e $T_1 \le T_0$.Thus ideal gas law in equation (2) is applicable from T_0 to reflow temperature T as follows

$$\frac{p_0 dV_{f0}}{T_0} = \frac{p dV_f}{T} \text{ or } \frac{p_0 f_0}{T_0} = \frac{p f}{T} \qquad (15)$$

where p_0 and dV_{f0} are the vapor pressure and void volume at preconditioning temperature T_0, respectively. When delamination takes places at T, the dV_f=dV or f=1. From equation (A-2) [15], the vapor pressure

$$p = \frac{p_0 f_0 T}{T_0} = \frac{C p_g(T_0) T}{\rho_g(T_0) T_0} \text{ when } T_1 \le T_0 \qquad (16)$$

In the second case, the moisture is not fully vaporized before delamination until temperature T_1, which are between preconditioning temperature, T_0, and the peak reflow temperature, T. Thus the ideal gas law is used from T_1 to T as

المنارة للاستشارات
ww

$$\frac{p_g(T_1)dV_{f_0}}{T_1}=\frac{pdV_f}{T} \text{ or } \frac{p_g(T_1)f_0}{T_1}=\frac{pf}{T} \quad (17)$$

With f=1, equation (17) becomes

$$p=\frac{p_g(T_1)Tf_0}{T_1} \text{ when } T_0 \le T_1 \le T \quad (18)$$

For the third case, the moisture is not fully vaporized until the voids grow up. Therefore, before voids grow up, the vapor pressure at T is its saturated vapor pressure at T, as shown in equation (10). When voids grow up, the void volume increases, and more moisture will be vaporized. Assume that the moisture will be fully vaporized at $f_s<1$. From f_s to f=1, the moisture is at single vapor phase, and therefore, the ideal gas law can be applied as follows

$$\frac{p_g(T)dV_{f_s}}{T}=\frac{pdV_f}{T} \text{ or } \frac{p_g(T)f_s}{T}=\frac{pf}{T} \quad (19)$$

The f_s can be determined by the following (see Appendix II)

$$f_s = CRT/p_g = C/\rho_g \quad (20)$$

Equation (19) becomes, when f=1

$$p=p_g(T)f_s = Cp_g(T)/\rho_g(T) \quad \text{when } T_1 \ge T \text{ and } f_s \le 1 \quad (21)$$

In the last case, the saturated vapor pressure remains even the delamination occurs, i.e.,

$$p=p_g(T) \quad \text{when } T_1 \ge T \text{ and } f_s>1 \quad (22)$$

Fig. 16 plotted the results of the initial vapor pressure at interface after delamination at reflow (220°C) as function of local moisture concentration *C*, when a typical value of initial void volume fraction f_0 is taken as 1.5. It showed that the equations (21) and (22) are dominant over a wide range of the moisture concentration. For cases 1 and 2, since the vapor pressure at reflow is rather low, the delamination may never occur.

The vapor pressure at the delaminated area will be immediately uniform when the delamination is complete, although the local moisture concentrations at interfaces are different with the location. Thus, the average local concentration can be defined as following

$$C^{ave}=\frac{\int_A CdA}{A} \quad (23)$$

where *A* is the delaminated area. In calculating the vapor pressure after delamination by above equations, the concentration *C* should be replaced by C^{ave}.

6. CONCLUSIONS

The package vapor pressure distribution during reflow is the key factor in understanding the failure mechanism. FEA moisture diffusion models are constructed for both FCBGA and wire bond PBGA packages to predict the

المنارة للاستشارات

local moisture concentration at the critical interfaces, which determines the magnitude of vapor pressure. The vapor pressure is found to be saturated much faster than the moisture diffusion. At reflow temperature, the moisture may not be fully vaporized after the level-1 moisture preconditioning, and thus saturated pressure is reached in most area of the FCBGA and wire bond PBGA packages. The vapor pressure generated can never go beyond the saturated pressure at the corresponding temperature, e.g., pressure of 2.32 MPa at 220°C. The vapor pressure is strongly temperature dependent. Moisture desorption affects the moisture distribution, rather than the vapor pressure distribution.

The moisture affects the package reliability at reflow from two aspects: generation of vapor pressure and degradation of interfacial adhesion. Although the vapor pressure remains at its saturated pressure when more moisture is absorbed, the adhesion strength may continuously deteriorate with additional moisture. When the interfacial adhesion is reduced to the level below the vapor pressure, the delamination will occur. The initial void volume fraction has insignificant effect on the vapor pressure distribution because there is always sufficient moisture concentration in the package to maintain the saturated pressure. However, larger void volume fraction along the interface will weaken the adhesion strength and make it more susceptible to failures.

The vapor pressure induces additional mismatch to the package, which is of the same order as the CTE thermal mismatch. This vapor pressure-induced expansion is directly related to the vapor pressure distribution, rather than the moisture distribution. Therefore, it is important to consider the mismatch caused by vapor pressure in the stress modeling.

The internal vapor pressure in voids at interfaces become an external pressure subjected to the delaminated interfaces when delamination is formed at reflow. A complete solution for the vapor pressure after delamination is derived and four different cases are identified. It showed that the case 3 and 4 are dominant in a wide range of the moisture concentration.

REFERENCES

1. Fan X.J. & Lim T.B., Mechanism Analysis for the Moisture Induced Failures in IC Packages, ASME 1999 International Mechanical Engineering Congress, 11th Symposium on Mechanics of Surface Mount Assemblies, Nashville, Tennessee, 14-19 November, 1999.
2. Tee, T.Y., Fan, X.J., and Lim, T.B. Modeling of Whole Field Vapor Pressure During Reflow for WB BGA and FC BGA Packages, 1st International Workshop on Electronic Materials and Packaging, September 29 – October 1, 1999 in Singapore.

المنارة للاستشارات

3. Fukuzawa, I., Ishiguro, S., and Nanbu, S., Moisture Resistance Degradation of Plastic LSI's by Reflow Soldering, Proc. IRPS, 1985, pp. 192-197.
4. Galloway, J.E. and Miles, B.M., Moisture Absorption and Desorption Predictions for Plastic Ball Grid Array Packages, IEEE Tranasction on Component,Packaging, and Manufacturing, A, 1997, 20(3), pp. 274-279.
5. Gurson, A.L., ASME Journal of Applied Mechanics, 1977, 99, pp. 2.
6. JEDEC Standard, 1994, Test method A 112, Moisture Induced Stress Sensitivity for Plastic Surface Mount Devices, Electronic Industries Association.
7. Kitano, M., Nishimura, A., and Kawai, S., Analysis of Package Cracking during Reflow Soldering Process, Proc. IRPS, 1988, pp. 90-95.
8. Tay, A.A.O. and Lin, T., Moisture Diffusion and Heat Transfer in Plastic IC Packages, IEEE Tranasction on Component, Packaging, and Manufacturing, A, 1996, 19(2), pp. 186-193.
9. Yi, Sung, Goh, J. S. and Yang, J. C.,Residual Stresses in Plastic IC Packages During Surface Mounting Process Preceded by Moisture Soaking Test, IEEE Transactions on Components, Packaging, and Manufacturing Technology, B, 1997, Vol. 20, No. 3, pp. 247-255.
10. B.Gebhart, *Heat Conduction and Mass Diffusion,* New York, McGraw Hill, 1993

APPENDIX I

Assume the moisture content in voids is dm. When $T_0 \geq T_1$, the moisture is at single vapor phase. Thus the ideal gas law can be written as

$$p_0 dV_{f0} = dmRT \qquad \text{(A-I-1)}$$

Consider another state when the moisture in voids is just saturated, with the moisture content dm_g

$$p_g dV_{f0} = dm_g RT \qquad \text{(A-I-2)}$$

Equations (A-I-1) and (A-I-2) give

$$\frac{p_0}{p_g} = \frac{dm}{dm_g} = \frac{\frac{dm}{dV}\frac{dV}{dV_{f0}}}{\frac{dm_g}{dV_{f0}}} = \frac{C}{\rho_g f_0} \qquad \text{(A-I-3)}$$

Thus

$$p_0 = \frac{Cp_g}{\rho_g f_0} \qquad \text{(A-I-4)}$$

المنارة للاستشارات

ww

APPENDIX II

When voids grow up to f_s, the moisture is just vaporized, so

$$p_g dV_{fs} = dmRT \qquad \text{(A-II-1)}$$

Equation (A-II-1) can be further written as

$$p_g dV_{fs} / dV = dm / dVRT \qquad \text{(A-II-2)}$$

so

$$f_s = CRT / p_g = C / \rho_g \qquad \text{(A-II-3)}$$

Table 1. Properties of moisture preconditioning for a typical mold compound

Moisture preconditioning	Saturated vapor density ρ_g (g/cm^3)	Thermal diffusivity α_D (cm^2/s)	Saturated concentration C_{sat} (g/cm^3)
30°C/60%RH	3.04×10^{-5}	3.13×10^{-9}	7.86×10^{-3}
85°C/60%RH	3.58×10^{-4}	2.85×10^{-8}	8.84×10^{-3}
85°C/85%RH	3.58×10^{-4}	2.85×10^{-8}	1.25×10^{-2}

Table 2. Initial void volume fraction for some common materials in IC packages

Material	BT	Die Attach	Mold Compound	Solder Resist
f_0 (%)	3.46	3.29	1.46	5.05

Table 3. Diffusivity and C_{sat} used in the moisture diffusion modeling

Material	Diffusivity D (cm^2/s)	C_{sat} (g/cm^3)
BT	8.55e-9	2.40e-2
Die Attach	1.68e-7	5.30e-3
Mold Compound	5.40e-8	4.00e-3
Solder Resist	2.47e-8	3.88e-2
Underfill	5.60e-9	2.47e-2

المنارة للاستشارات

ww

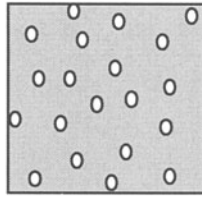


Fig. 1. A representative volume element (RVE)

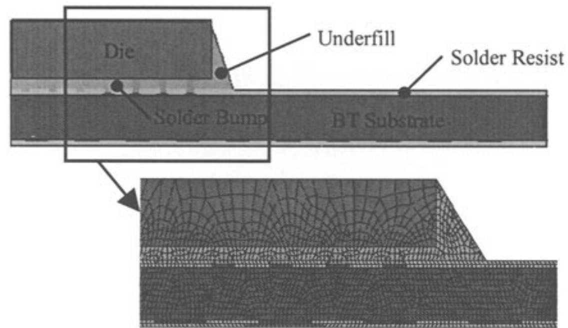


Fig 2. FEA model of FCBGA

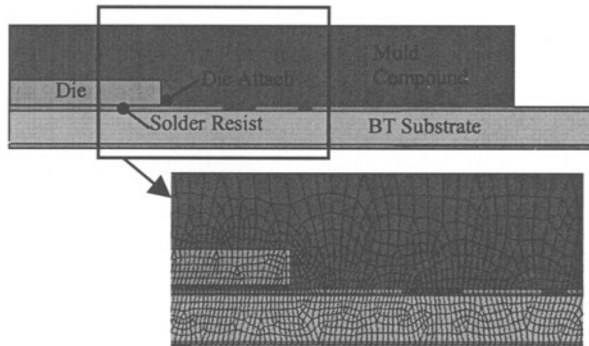


Fig 3. FEA model of wire bond PBGA

المنارة للاستشارات
ww

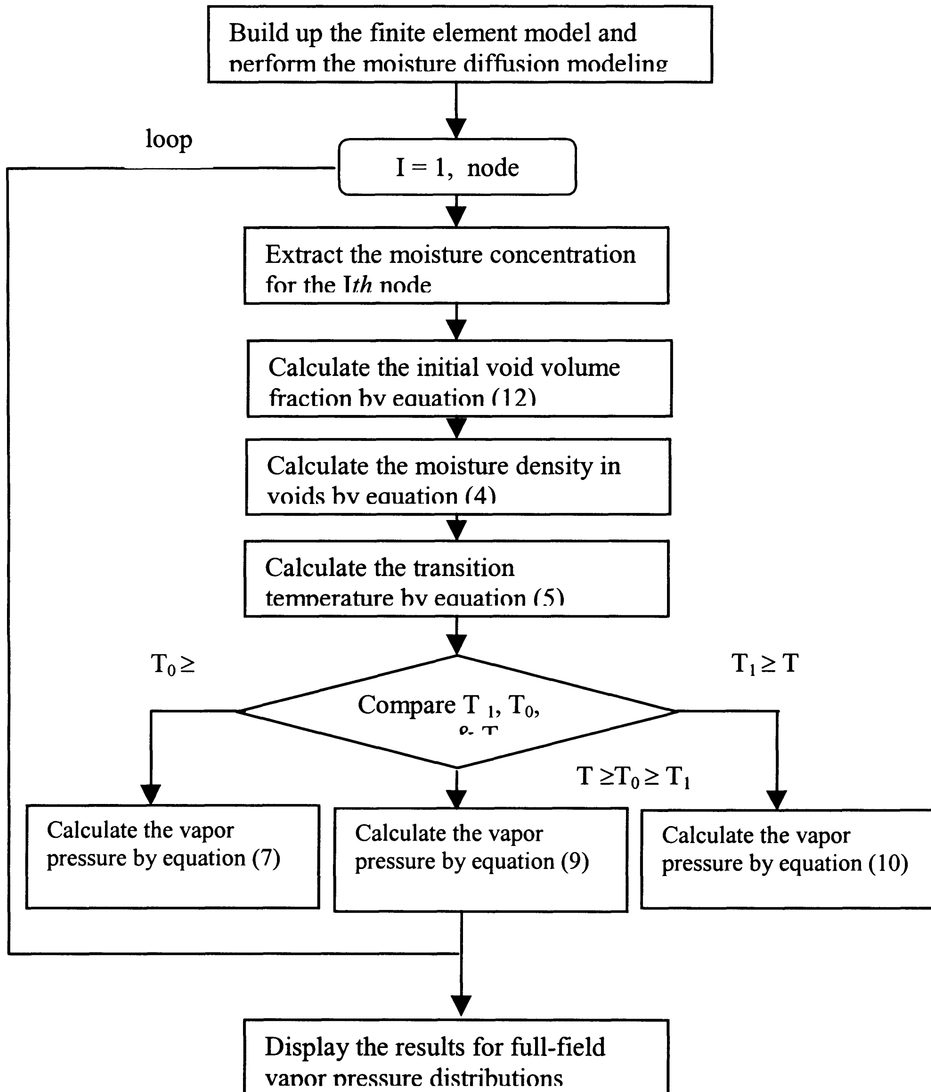


Fig. 4 procedure to calculate the vapor pressure

المنارة للاستشارات

ww

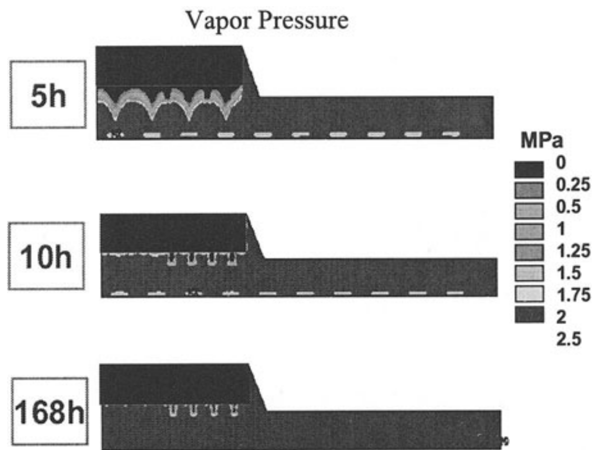


Fig. 5. Transient vapor pressure distribution in FCBGA at level 1, 220°C

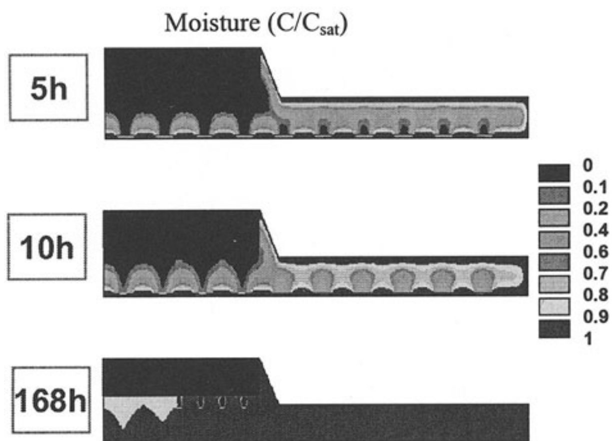


Fig. 6. Transient moisture distribution in FCBGA at level 1

المنارة للاستشارات

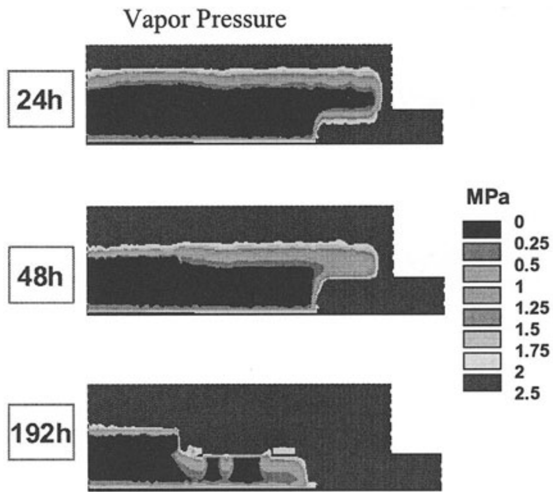


Fig. 7. Transient vapor pressure distribution in wire bond PBGA at level 3, 220°C

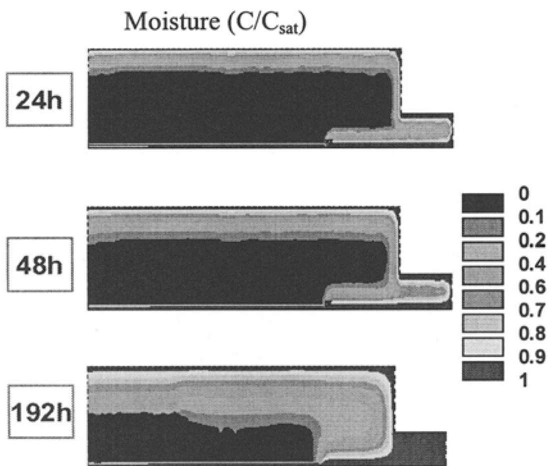


Fig. 8. Transient moisture distribution in wire bond PBGA at level 3

المنارة للاستشارات

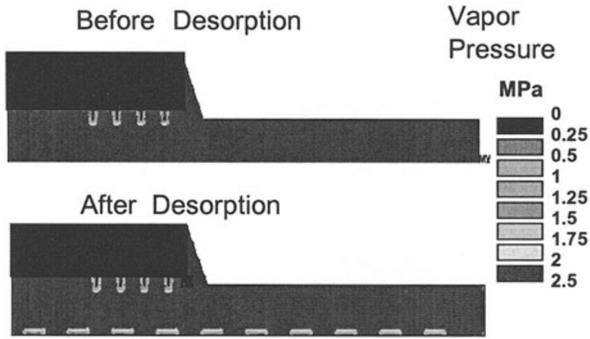


Fig. 9. Vapor pressure distribution in FCBGA before and after desorption

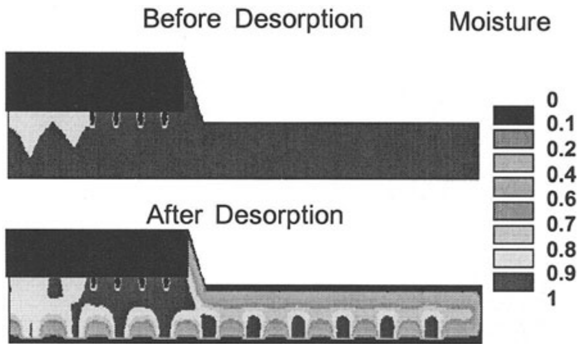


Fig. 10. Moisture distribution in FCBGA before and after desorption

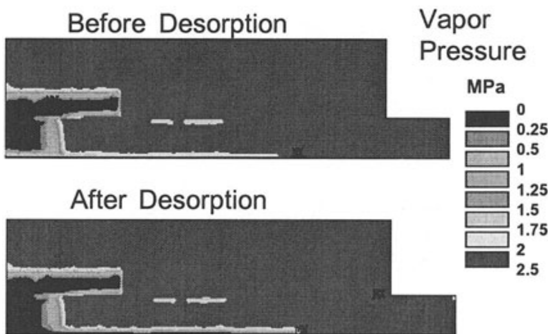


Fig. 11. Vapor pressure distribution in wire bond PBGA before and after desorption

المنارة للاستشارات

ww

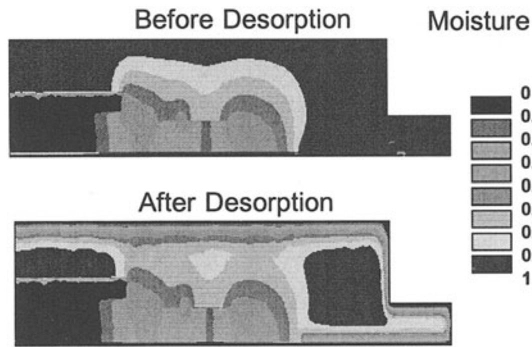


Fig. 12. Moisture distribution in wire bond PBGA before and after desorption

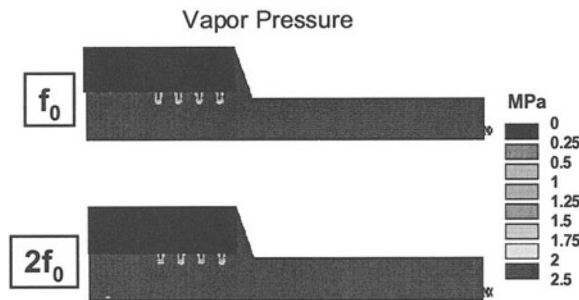


Fig 13. Effect of f_0 on the vapor pressure distribution of FCBGA

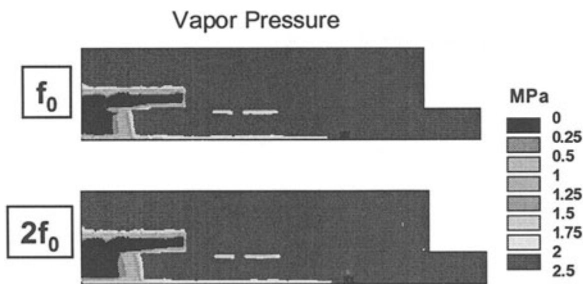


Fig 14. Effect of f_0 on the vapor pressure distribution ot wire bond PBGA

المنارة للاستشارات

ww

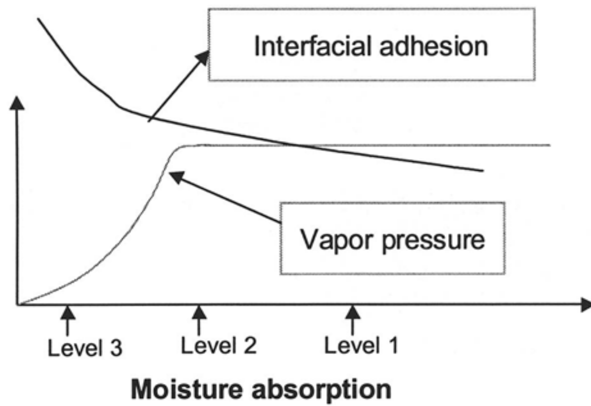
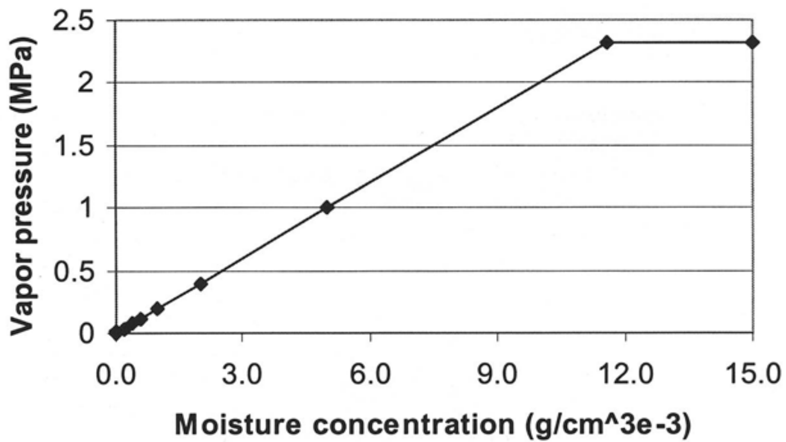


Fig. 15. Relative effect of moisture absorption on interfacial adhesion and vapor pressure



Case 1: 0 < C < 4.65 e-6 g/cm^3, case 2: 4.65 e-6 < C > 0.174 e-3 g/cm^3
case 3: 0.174 < C <11.6 g/cm^3, case 4: C>11.6 g/cm^3

Fig. 16 vapor pressure at interface after delamination at 220°C

المنارة للاستشارات

Simulation for fatigue, cracks and delamination

Rainer Dudek, Jürgen Auersperg, and Bernd Michel

Fraunhofer-Institut für Zuverlässigkeit und Mikrointegration (IZM), Gustav-Meyer-Allee 25,D-13355 Berlin, e-mail: dudek@che.izm.fhg.de

Key words: finite elements, interface fracture, delamination criteria

Abstract: Computer based thermo-mechanical design and performance optimization are in widespread use and are mainly based on finite-element (FE-) analyses. The theoretical investigations of stresses within electronic material compounds induced by environmental conditions, especially temperature changes, require both the characterization of materials properties and materials interface properties. Measurement results on typical commercially available electronic polymers are reported, which have been investigated by DMA and TMA measurements as well as tensile tests. The potential and application of failure models (e.g. integral fracture approaches, interfacial fracture modeling of several polymer/polymer interfaces) are discussed. The investigations are focused on the failure criteria at polymeric interfaces. FE-simulation results on an interface fracture experiment are compared to micrographs, which observe growing interface delaminations by using micro deformation measurements on the basis of a gray scale correlation method. Both numerical and experimental investigations provide the basis for a better understanding of failure mechanisms, and should contribute to an increase of applications in order to enhance the thermo-mechanical reliability of advanced electronic packages..

1. INTRODUCTION

Growing demands on performance, cost and the advancement in IC technology have drastically influenced packaging and interconnection technology. New packaging technologies and advanced materials have to be

G. Q. Zhang et al. (eds.), Benefiting from Thermal and Mechanical Simulation in Micro-electronics, 93–111.
© 2000 *Kluwer Academic Publishers.*

المنارة للاستشارات

ww

developed to handle larger dies, higher input/output (I/O) counts, lower operating voltages, high power consumption and high clock frequencies.

The computational design of reliable packages can minimize prototype development and testing, especially with the growing use of area array packages. Accordingly, finite element (FE-) modeling is widely used to perform parametric studies on their thermomechanical behavior. Since the packages of interest are material compounds, fatigue, delamination and interface cracking are the most frequently observed failure modes. Solder fatigue is one failure frequently observed in electronic assemblies. The application of a Coffin-Manson type criterion based on the cyclic equivalent creep strain or strain energy density to that type of failure has shown acceptable failure prediction capabilities, see e.g. [1].

Note that processing induced residual stresses generally overlay the loading induced stresses and, therefore, affect the critical stress states. For example, encapsulant/die-passivation or encapsulant/solder-mask cracking are typical failure modes within flip chip assemblies. The use of the FE-method requires a certain amount of input information, in particular concerning the proper geometry description and the constitutive modeling of the materials used in the assemblies.

It was pointed out earlier [2] that epoxy-based adhesives and encapsulants exhibit temperature, time and moisture dependent mechanical characteristics. Many other investigations deal with the mechanical characterization of electronic polymers, e.g. Harper et al. [3], who studied the effect of temperature and moisture upon the viscoelastic behavior of epoxy molding compounds, Chew [4], who discussed the effects of post-mold cure on thermal and viscoelastic characteristics of transfer-molded epoxy encapsulants, Chambers et al. [5], who developed a constitutive theory that can describe the behavior of glassy polymers under arbitrary histories and Qian et al. [6], who applied a viscoplastic material law to underfill materials.

Another general question for theoretical analyses of material compounds is linked to the stresses at material interfaces, which exhibit a strongly discontinuous character caused by the action of interfacial discontinuities like interfacial free edges and corner edges, see e.g. [7]. Due to the assumption of sharp interface edges stress singularities arise which might be accounted for only approximately in the FE-calculation.

2. MATERIAL CHRACTERISTICS FOR POLYMERIC MATERIALS USED IN ELECTRONICS

A broad variety of polymers containing different kinds of fillers with variable content has been investigated [8]. Obviously, the importance of a

المنارة للاستشارات

ww

mechanical property depends strongly on the use specifications of the material, i.e. the temperature range or applied strain range under service conditions. Accordingly, the typical application of the materials was assumed to be in flip-chip or other COB applications, where mechanical stress is mainly caused by thermal expansion mismatch. That means, deformations remain small but temperature dependence of the mechanical properties becomes essential. The experimental investigations include therefore TMA measurements of the thermal expansion and related glass transition temperature, DMA measurements of the storage and loss modulus and tensile tests on miniature dog-bone specimen. Different kinds of tensile tests were performed to record both the temperature dependent stress-strain curves and the relaxation behavior.

The thermal expansion behaviors of all investigated materials exhibit temperature dependence, which can be fitted well by a bi-linear approximation. The point where slope changes can be identified as the Tg temperature, which is mostly approximately 10 % lower than that measured by DMA analysis.

Fig. 1 shows TMA measuring results for three underfill materials with different filler content. The effect of lower thermal expansion with increasing filler concentration is clearly visible from the figure. The glass transition points taken from the changes in slope of the CTEs are only slightly influenced by the different filler contents.

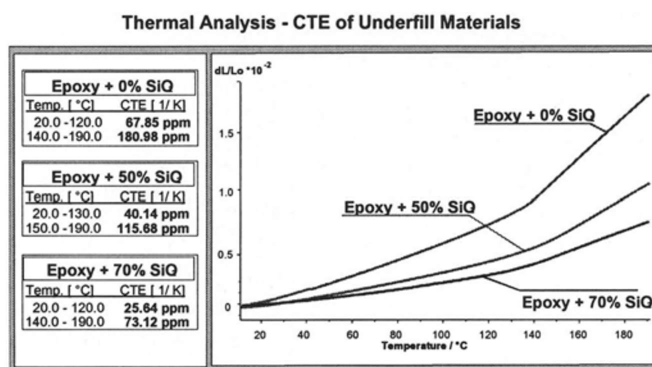


Fig. 1: Temperature dependent CTEs in dependence on filler content (matrix material unchanged)

Generally, the elastic stiffness' show a strong dependence on temperature. The static Young's modulus as well as the storage modulus decreases with increasing temperature and drop down when the temperature comes close to the glass transition range. In most cases, comparison of both experimental techniques exhibits a good agreement between storage

المنارة للاستشارات
ww

modulus and the initial static Young's modulus for the temperature range well below T_g, where tan δ is close to zero, but the differences become more pronounced at a temperature level close to Tg. This trend is closely linked to the shift of constitutive properties from elastic to viscoelastic.

Two typical examples are considered in the following, one underfill (Ablebond 7737) and one encapsulant for glob-top application (Ablebond 7950). The material characteristics of these two materials measured by DMA, isothermal tensile tests and tensile relaxation tests, respectively, are given in Figs. 2 to 4.

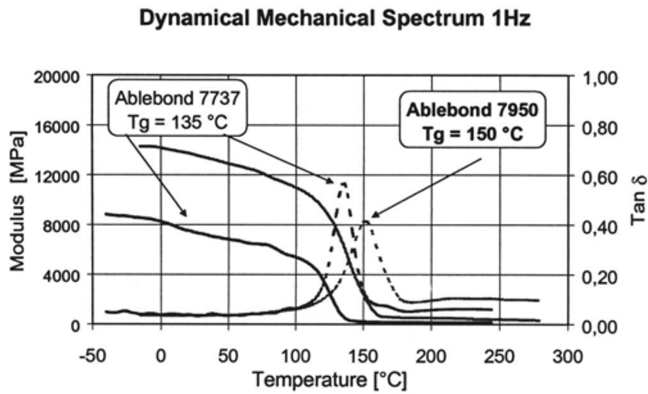


Fig: 2: Storage modulus and tan δ of the encapsulant vs. temperature for a typical underfill and encapsulant

It is evident from the stress-strain curves that the investigated specimens of filled materials exhibited a quasi-brittle behavior for temperatures well below Tg and fast loading conditions, e.g. in case of the Ablebond encapsulant for a temperature range up to 70 °C for this material with Tg=135 °C, Fig. 3. Note that the stress-strain curves are recorded up to the fracture strain. The specimen deformation is essentially elastic up to the fracture strain. In the Tg temperature range material behavior becomes usually non-linear for higher strains and strongly time dependent, see Figs. 3. However, within this temperature range strains are small in thermal stress situations with stress free temperature close to Tg.

المنارة للاستشارات

ww

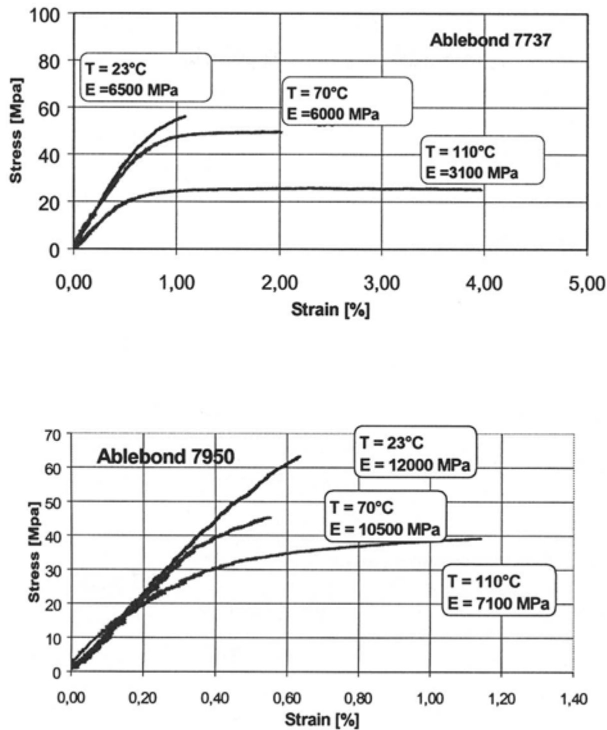


Fig: 3: Temperature dependent stress-strain curves, typical underfill and encapsulant (strain rate 1 % /min)

Despite the nonlinear character of the material properties shown, there are several reasons to use a simplifying constitutive description within a theoretical model of a polymeric material, e.g. a linear elastic material model. One is to keep the measuring effort as low as possible, another is the application of failure criteria adopted from fracture mechanics or interface fracture mechanics, which are only well defined within the framework of linear elasticity. In many situations the modeling error can be kept low despite this modeling simplification. Such a situation can arise for an isothermal mechanical or a thermal loaded specimen at a temperature range below or above Tg, where the material is either in its glassy or its rubbery state. Care has to be taken that the loading times are small compared to the characteristic material relaxation times for the applied temperatures if relaxation is not negligible at the specified temperature range.

المنارة للاستشارات

ww

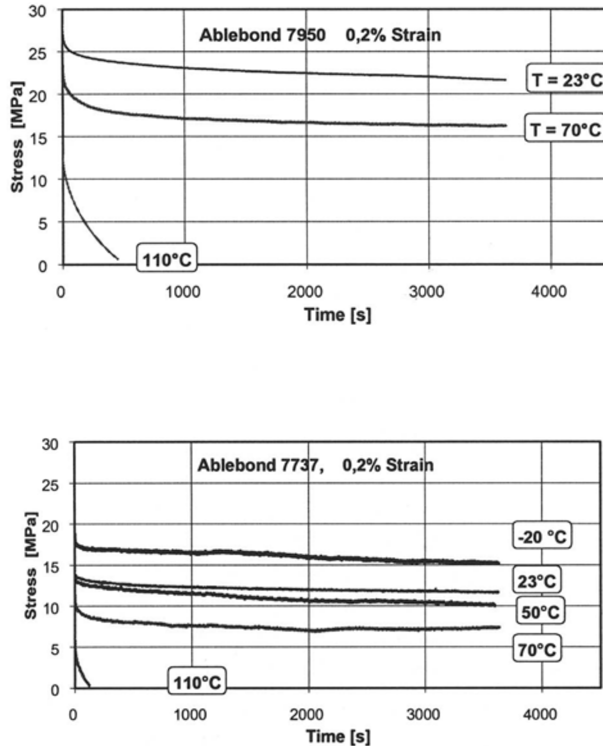


Fig: 4: Measured stress relaxation functions (0.2 % strain), , typical underfill and encapsulant

In many applications with small strains the temperature dependent linear viscoelastic theory is appropriate to model the constitutive behavior. Different linear viscoelastic constitutive models have been developed for finite element analysis.

3. FRACTURE INITIATION AT INTERFACES

Within almost all packaging applications, i.e. flip-chip, CSP, glob-top as well as molded packages, polymeric materials form interfaces to other materials needed in the assembly. Most frequently, these interfaces are between the polymer and the silicon die, which might be coated itself by a passivation layer, or a substrate material, which can be either a ceramic or an organic one. It has long been recognized from experimental experience that delamination or crack initiation at these interfaces are between the most

المنارة للاستشارات
ww

likely reasons for failure. Although various theoretical analyses have dealt with this subject, e.g. [9], [10], many related questions are still open.

3.1 Interface Edges

Interface fracture is usually initiated at sharp interface edges, which act as stress concentration locations. Because these interface edges can form different interface angles with some of them exhibiting a singular behavior of stresses at the edge points, the traditional failure analysis, which relates stresses in the material to appropriate strengths, will fail. The second applicable approach is based on fracture mechanics and consists of relating a fracture toughness measure of the interface, e.g. the energy release rate, to a critical energy release rate which has to be measured by appropriate tests. This approach is discussed later. It is important to note that the loading situation has to be chosen in accordance with the application situation, because the mode mixity at the crack tip will change in dependence on far field loading. Obviously, an initial interface crack is required in order to use fracture mechanics.

For linear elasticity as well as linear viscoelasticity the stresses at a bi-material notch tip with zero radius of curvature can have singular character. This property depends on several parameters like the relation of the shear moduli of the two joined materials, their Poisson's ratios and, most extensively, on the wedge angle of the materials at the notch tip. If the typical situation of a substrate covered with an encapsulant is examined, the local geometry at the interface edge can be described as that of two bonded dissimilar materials with interfacial edge, one of angle β=180° while the other angle α is of arbitrary size, see Fig. 5. An asymptotic solution of the problem has been given by Scherzer [11] for both elastic and elastic-plastic materials characteristics. At the interfacial edge the stress components can be formulated in the following set of equations for the stress components $t_{ij} = \sigma_{ij}/G$, G denoting the shear modulus and ξ the distance from the singular edge point in a polar coordinate system, related to a characteristic length l:

$$t_{ij}(\xi,\theta)=\sum_{i=1}^{\infty} C_i \xi^{\lambda_i}\left[f_{ij}(\lambda_i,\theta)\right]$$

When linear elasticity is assumed, the behavior of the asymptotic solution, i.e. the order of the highest stress singularity λi, for the interfacial edge follows one graph given in Fig. 5 (right), dependent on the relation of the elastic modulus of the two joined materials and their Poisson's ratio. For small angles α the solution is not singular, for intermediate angles it becomes singular with the highest real eigenvalues of $-0.5\leq \lambda_i<0$, but

المنارة للاستشارات
ww

complex stress eigenfunctions occur for angles α approaching the interface crack case α=180°.

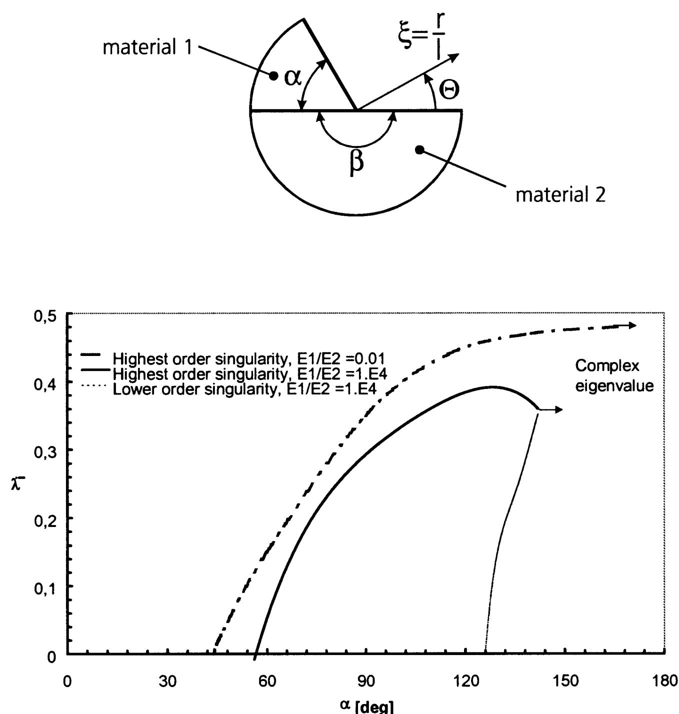


Fig. 5: Interface edge between dissimilar materials and schematic of the singular behavior in dependence on the wedge angle α (β =180 °) for different material combinations with Young's modulus relation E1/E2

This behavior should always be kept in mind when evaluating conventional FE-analysis results in the vicinity of interfacial edges, because it means that the FE-approximation can be mesh dependent at these locations. Thus, by conventional means, stresses at those edges can be compared only for equal mesh densities.

Common knowledge in mechanical engineering assumes that cracks starting at such edges can be taken into account in order to have a conservative evaluation of the fracture toughness of the several materials interfaces. The application of fracture mechanics concepts is the recommended procedure, but has not been examined just as well for mixed mode conditions at materials interfaces as it has for mode I situations in homogeneous media.

المنارة للاستشارات
ww

3.2 Interface Fracture Toughness

Coming from K-concept usage, a lot of work was done recently in order to explore mode mixity effects, e.g. [12], [13], [14]. But, in contrast to investigations into fracture phenomena in homogeneous materials, integral fracture concepts like the *J*- or the ΔT^*-integral

$$\Delta T_k^* = \int_\Gamma \left(\Delta W n_k - \left(\sigma_{ij} + \Delta\sigma_{ij} \right) n_j \Delta u_{i,k} \right) d\gamma + \iint_\Omega \left(\frac{1}{2} \Delta\sigma_{ij} \varepsilon_{ij,k} - \Delta\varepsilon_{ij} \left(\sigma_{ij,k} + \frac{1}{2} \Delta\sigma_{ij,k} \right) \right) d\omega$$

with

$$\Delta W = \left(\sigma_{ij} + \frac{1}{2} \Delta\sigma_{ij} \right) \Delta\varepsilon_{ij}$$

from Brust, Atluri et al. [15], [16], which have the potential to take into consideration the inelastic behavior of the related materials, remain almost unconsidered for use. One reason for this seems to be the subdivision of the region of interest into two or more parts and the mechanical behavior of the interface itself. Because integral fracture approaches base on energy release considerations, this methodology should be moved more into the center of interest, taking into account that some work has to be done in the future in order to make it utilizable. Therefore, the numerical investigations performed here are using the energy release rate and the phase angle between its components as parameters, which have been investigated in the literature for fracture toughness and mode mixity analyses.

4. INVESTIGATIONS ON INTERFACE STRENGHT OF POLYMERIC MATERIALS ON ORGANIC BOARDS

A specimen type suitable for interface adhesion investigations is a simple bi-material strip, which was produced for a number of material combinations to study their interface behavior.

المنارة للاستشارات
ww

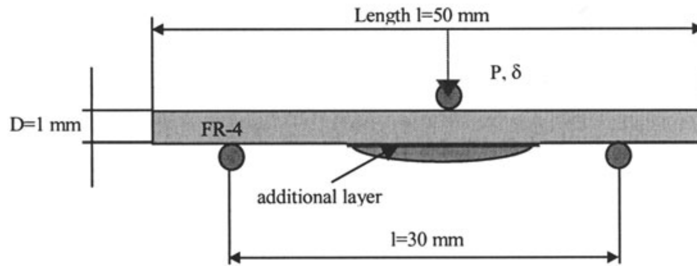


Fig. 6: Prismatic bi- or tri-material strip used for 3-point bending experiments

Fig. 6 shows the cross section of a prismatic bi- or tri-material strip – glob-top encapsulant on FR-4 board - customized for 3-point bending experiments. The measurement environment allows load and deflection control and the utilization of an in-situ local deformation analysis by the gray-scale correlation method, see e.g. [17]. As results, the damage propagation as well as the crack tip near deformation field can be observed during loading. This is very helpful for a better mechanical interface description and for the examination of the crack extension vs. time.

4.1 Strength of Materials Approach

The experiments were accompanied by FE-simulation to determine the stresses resulting from both the thermal and mechanical loadings. In a first step, different glob top geometries without initial flaws were studied. Generalized plane strain assumption was applied for the analyses, which were conducted by help of the ABAQUS [18] software package.

Two different geometries are considered, chosen from cross sections investigated microscopically. Besides the different edge geometries shown in Fig. 7, the glob-top height is lowered from 1 mm (geometry 1) to 0.5 mm (geometry 2).

The peel stress at the interface of both geometric versions after cooling from 150 °C to RT is shown in Fig. 8 for the underfill material Ablebond 7737 discussed above. Both specimen of different geometries are considered without solder mask. It can be observed from the figure that this stress component vanishes along the interface but exhibits a strong tensile peak at the interface edge. An asymptotic analysis yields the result for geometry 1 to be infinite with highest singularity order of λ_1 = -0.389. The changed glob-top geometry (geom. 2) with interface angle 17.5° results in a finite peel stress value with maximum value less than 10 % of that for geometry 1, achieved by the same FE-approximation.

المنارة للاستشارات

ww

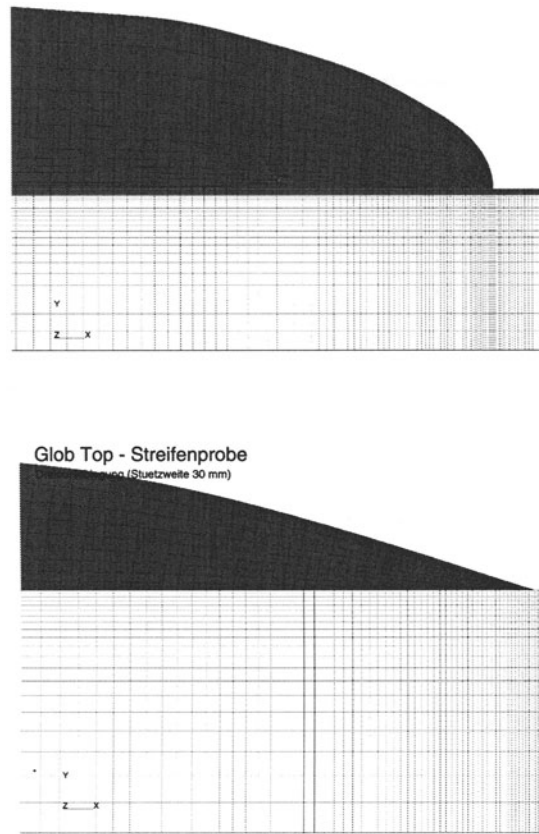


Fig. 7: Interface edge region of two bi-material strips, geometry 1 with solder mask layer (upper) and geometry 2 without solder mask layer (lower)

When the strips are subjected to additional bending, the stress distribution at the interface remains unchanged, but the stress amplitudes rise for both geometries. This rise is much stronger for the geometric version 1. Analogous tendencies are also obtained for the shear stress distribution along the interface, which is only antimetrically distributed.

The different theoretical approaches all more or less fail to describe the interface edge phenomenon of different wedge angles with a quantitative failure criterion: The classical strength of materials approach fails, on account of the mesh dependence of the stress/strain maxima, the asymptotic analyses on account of the changing kind of stress singularity, and the interface crack theory on account of the required assumption of a predefined interface crack, which is not relevant for applications with small wedge angles. The latter case occurs e.g. for glob-top materials with low viscosity

المنارة للاستشارات
ww

and good wetting characteristics. The bending experiments have shown that for such bimaterial strips cracks do not start at the interface edges.

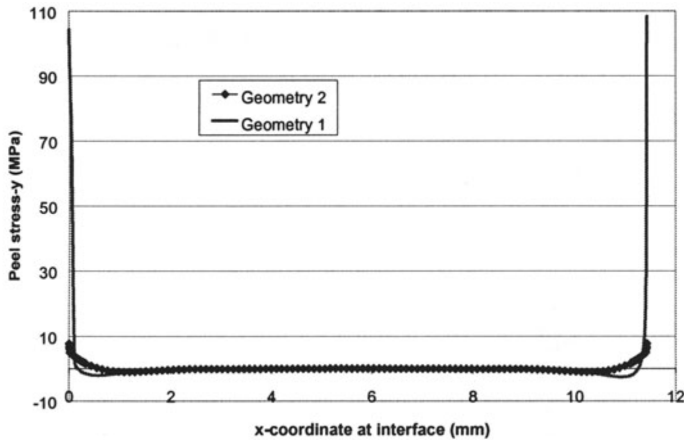


Fig. 8: Peel stress along the interface between polymer and FR-4 board (x=0 left interface edge), temperature drops from 150 °C to 20 °C

4.2 Interface Fracture Mechanics Approach

The samples considered here are made from an Alpha Metals underfill material on top of an FR-4 strip with or without a soldermask in-between, with geometries similar to Fig. 7 geometry 1.

The micrographs in Fig. 9 show a completely different interface behavior during delamination. While the interface between FR-4 and soldermask as the weakest link fails, the glob-top on board interface seems to be much tougher, so that the internal structure of FR-4 causes the “crack” propagation, destroying the surface of FR-4.

The load-deflection curves of specimens with and without soldermask depicted in Fig. 10 demonstrate their different failure behavior and interface toughness. The measurement of the local deformation fields in the vicinity of the propagating delamination by use of the gray scale correlation method lead to a more accurate analysis of the crack growth during loading, see Fig. 11.

المنارة للاستشارات

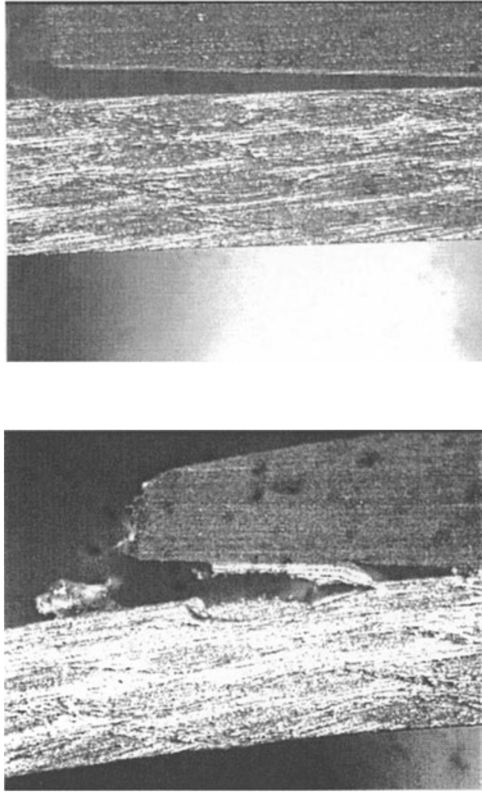


Fig. 9: Delamination between solder mask and FR-4 and failure in the case of no soldermask

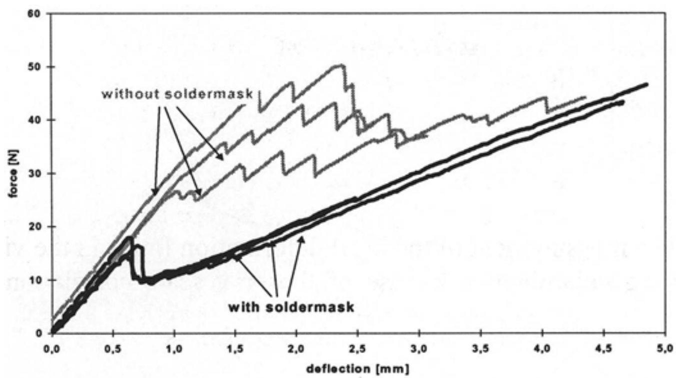


Fig. 10: Load-deflection curves of specimens with and without soldermask

المنارة للاستشارات

ww

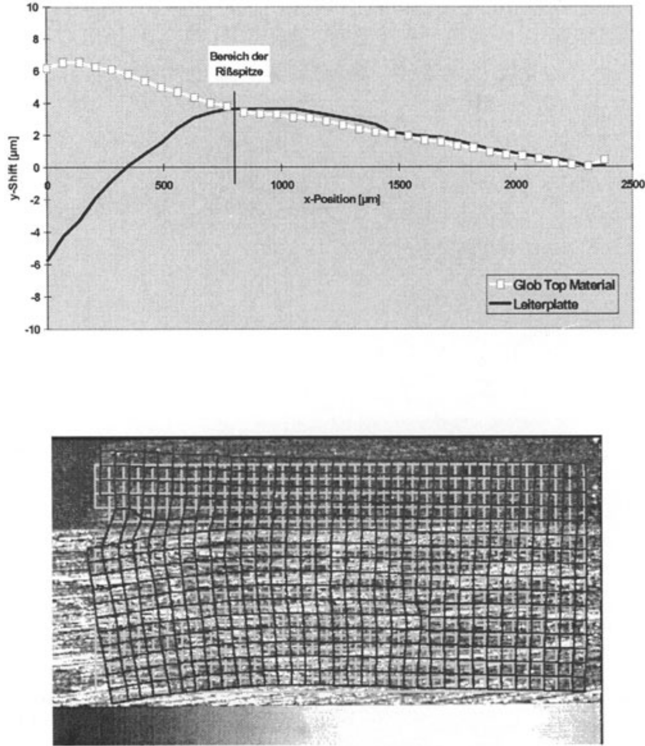


Fig. 11: Approximate crack tip localization utilizing the gray scale correlation method and correlation pattern

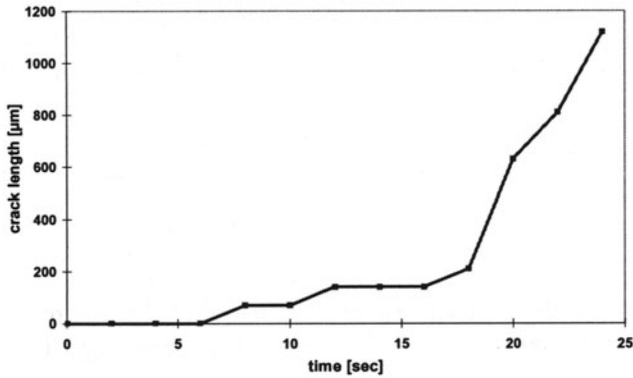


Fig. 12: Crack length vs. time in case of no soldermask

The observation of the estimated crack tip location at each loading step, Fig. 12, provides further input data for subsequent numerical investigations.

المنارة للاستشارات

4.2.1 Interfacial Crack Propagation

In order to evaluate the fracture toughness of the materials interfaces, the same sample was modeled assuming an initial interface crack. Again, generalized plane strain conditions were applied in the analyses. All boundary conditions including the load line deflection vs. time were taken from experimental data. While the mechanical behavior of FR-4 was modeled as linear elastic, the soldermask and the glob-top material were considered as viscoelastic. Residual stresses/strains from manufacturing process (acting as initial stresses/strains) were taken into consideration by a previous step, cooling down the entire sample from curing temperature to room temperature. Fig. 13 illustrates the deformation of the prismatic bi-material strip subjected to 3-point bending.

Besides the commercial code ABAQUS an in-house finite element code ALICE [19] (allowing programmed crack propagation simulations, supported by an adaptive meshing technique) has been used for the calculations.

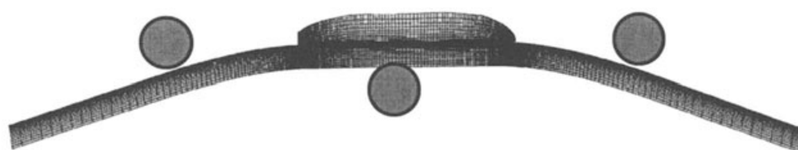


Fig. 13: Deformation of the prismatic bi-material strip subjected to 3-point bending

The goal of the finite element simulations is to relate the calculated energy release rate vs. time curves to experimental obtained results in order to identify a critical energy release rate describing the fracture toughness of the interface. Therefore, two different simulations were carried out:

1. The crack tip follows the experimental crack length vs. time behavior as shown in Fig. 12.
2. The crack propagation is controlled by G_C, where a sleeping crack tip starts when $G>G_{Cs}$, and a growing crack stops when $G< G_{Cr}$.

Simulating the crack extension as shown in Fig. 12, the resulting energy release rate G(t) is showing a stable time period between 11.45 and 19.6 sec., respectively between the crack length's 0.2 and 0.63 mm with a value of about G_c=14 [J/m^2] – see Fig. 14. Fig. 15 gives an impression of the stress/strain concentrations in the crack tip vicinity during the crack propagation.

المنارة للاستشارات

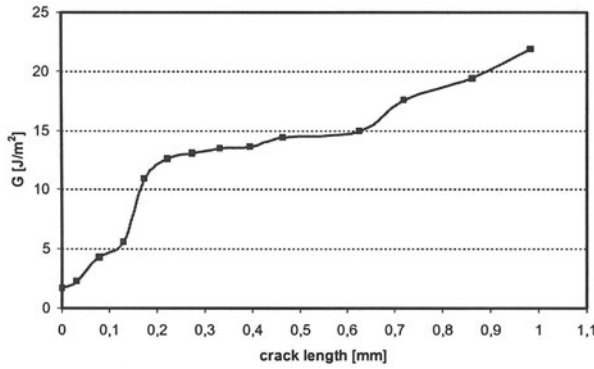


Fig. 14: Energy release rate vs. crack length in the case with soldermask

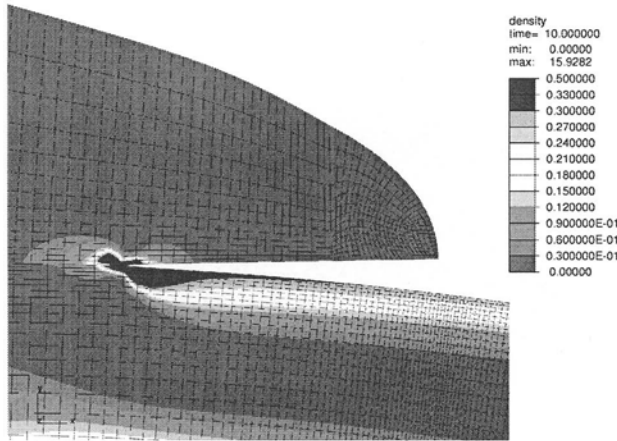


Fig. 15: Mechanical energy density at one time step of the G_c- controlled crack propagation simulation

Because of the impossibility to measure the crack propagation vs. time curve for the samples without soldermask, it was impossible to evaluate G_c in that way. A remedy might be the assumption of critical energy release rates starting or stopping the crack propagation (as described before) together with the comparison of both the load/deflection curves and deformation fields with those of the experiments.

Assuming a characteristic, G_c dependent crack velocity during propagation and the two values G_{Cs} and G_{Cr} as described above, a load/deflection curve like that shown in Fig. 16 (right) is a result. If the load/deflection curves from measurement and simulation are matching each other, G_c can be assumed to be

المنارة للاستشارات
ww

the critical energy release rate, that was sought, which characterizes the fracture toughness of such an interface.

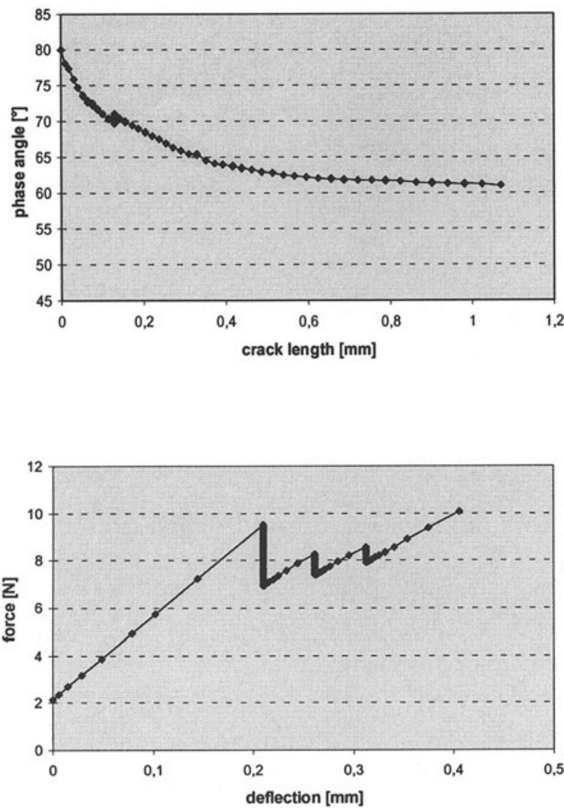


Fig. 16: Phase angle vs. crack length and load-deflection curve as result from G_c-controlled crack propagation simulation

5. SUMMARY

FE-analyses on the thermo-mechanical reliability of micro-components require special knowledge on characteristics of electronic materials as well as appropriate failure criteria to achieve reliable failure predictions. Mechanical properties of selected electronic polymers have been shown to be strongly temperature dependent. They depend also on the loading velocity. This viscoelasic effects are most pronounced in the glass transition temperature range. Additional effects like curing induced residual stresses, nonlinear

المنارة للاستشارات

ww

behavior, and moisture controlled property changes will be treated during subsequent analyses.
The limits of the strength of materials approach at interface edges has been pointed out. Solutions like an interface K-concept or interface fracture mechanics approaches with initial delamination are discussed in detail. Simulations on bending experiments of prismatic bi-material strips have shown the expected stress concentrations at the interface edges, which correlate with the non-singular or singular behavior, respectively, resulting from asymptotic analyses in dependence on the interface wedge angle. With the integral-concepts of fracture mechanics, applied to the same kind of bi-material strips, critical energy release rates Gc have been determined. Simulation of interface crack propagation by means of a special purpose FE-code have shown reasonable correspondence to experimental observations when a modified Gc criterion is applied. For the specimen underfill on FR-4 board an existing solder mask layer lowers the interface strength remarkably.

6. ACKNOWLEDGEMENT

Parts of this work are founded by the BMBF and SMWA, which is gratefully appreciated. The authors would like to express thanks to H.Walter, IZM Berlin, and E. Kieselstein, CWM GmbH Chemnitz, for their valuable measuring results. They thank also W. Schneider, MPD GmbH Dresden, for his support and the manufacturing of polymer specimens. Finally, they appreciate the assistance of R. Döring, CWM GmbH Chemnitz, in performing the FE modeling.

7. REFERENCES

[1] R. Dudek, M. Nylen, A. Schubert, B. Michel, and H. Reichl: An Efficient Approach to Predict Solder Fatigue Life and its Application to SM- and Area Array Components. *Proc. of ECTC 47*, San Jose, May 1997, pp. 462-471

[2] R. Dudek; D. Vogel; B. Michel: Mechanical Failure in COB-Technology Using Glob-Top Encapsulation. *IEEE Trans. Comp., Packag., Manufact. Technol.-Part C*, Vol. 19, no. 4, Oct. 1996, pp. 232-240

[3] B. D. Harper, L. Lu, V. H. Kenner: Effects of Temperature and Moisture Upon the Mechanical Behavior of an Epoxy Molding Compound. *EEP-Vol. 19-1, Advances in Electronic Packaging*, Vol. 1, ASME 1997

[4] S. Chew: Thermal And Viscoelastic Characterization Of Transfer-Molded Epoxy Encapsulants During Simulated Post-Mold Cure. *Proc. of ECTC 46*, Orlando, May 1996, pp. 1032-1038

[5] R.S. Chambers, D.B. Adolf, J.M. Caruthers: Nonlinear Viscoelastic Analysis of Engineering Polymers. *EEP-Vol. 26-1, Advances in Electronic Packaging*, Vol. 1, ASME 1999, pp. 609-610

المنارة للاستشارات

ww

[6] Z. Qian, M. Lu, W. Ren, and Sheng Liu: Fatigue Life prediction of Flip-Chips in Terms of Nonlinear Behavior of Solder and Underfill. *Proc. of ECTC 49*, May 1999

[7] R. Dudek, M. Scherzer, A. Schubert, and B. Michel, FE-Simulation for Polymeric Packaging Materials. *IEEE Trans. Comp., Packag., Manufact. Technol.-Part A*, vol. 21, no. 2, June 1998, pp. 301-309

[8] H. Walter, W. Schneider, E. Kieselstein, R. Dudek, E. Auerswald, A. Schubert: Evaluation of glob top materials for chip on board (COB) applications. *Proceedings MicroMat 2000 conference (to appear)*, Berlin, April 2000

[9] R.A. Pearson, T.B. Lloyd, H.F. Nied: Fracture Behavior of Underfil Resins and their Application to Flip-Chip on Organic Assemblies *EEP-Vol. 26-2, Advances in Electronic Packaging*, Vol. 2, ASME 1999, pp. 1749-1754

[10] R.J. Harries, S.K. Sitaraman: Numerical Study of Copper-Encapsulant Interfacial Delamination Propagation in a Peripheral Array Package. *EEP-Vol. 26-2, Advances in Electronic Packaging*, Vol. 2, ASME 1999, pp. 1755-1762

[11] M. Scherzer: Non-Linear Deformed Interface Corner Stress Characterization by Effective Parallel Numerical Methods

in: B.L. Karihaloo, Y.W. Mai, M.I. Ripley, R, O. Ritchie (eds.): Advances in Fracture Research*, Proceedings of the Ninth International Conference on Fracture*, Vol. 4, April 1997, Sydney, Australia, Pergamon Press, Amsterdam, Oxford, New York, 1997, pp. 1959-1970

[12] Liu, Sh.; Mei, Y.; Wu, T. Y.: Bimaterial Interfacial Crack Growth as a Function of Mode-Mixity. *IEEE Transactions on Components, Packaging and Manufacturing Technology - Part A*. Vol. 18 No. 3, Sept. 1995, pp. 618-626

[13] Wang, J,; Lu, M.; Ren, W.; Zou, D.; Liu, Sh..: A Study of the Mixed-Mode Interfacial Fracture Toughness of Adhesive Joints Using a Multiaxial Fatigue Tester. *IEEE Transactions of Electronics Packaging Manufacturing*, Vol. 22 No. 2, April 1999, pp. 166-173

[14] Wang, J,; Lu, M.; Zou, D.; Liu, Sh..: Investigation of Interfacial Fracture Behavior of a Flip-Chip Package under a Constant Concentrated Load. *IEEE Transactions on Components, Packaging, and Manufacturing Technology - Part B*. Vol. 21 No. 1, Febr. 1998, pp. 79-86

[15] Brust, F. W., Atluri, S. N.: Studies on Creep Crack Growth using the T*- Integral. *Engng. Fract. Mech.*, 23(1986)3, pp. 551-574

[16] Brust, F. W., Nishioka, T., Atluri, S. N., Nakagaki, M.: Further Studies on Elastic-Plastic Stable Fracture Utilizing the T*-Integral. *Engng. Fract. Mech.*, 23(1986)3, pp. 551-574

[17] D. Vogel, J. Auersperg, A. Schubert, B. Michel, 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*, Feb. 9-13, 1997, Orlando, USA, in „Design & Reliability of Solders and Solder Interconnects“, ed. by R.K. Mahidhara, TMS Publication Cat. No. 96-80433, pp. 429-438

[18] *ABAQUS Theory Manual* (V. 5.7), Hibitt, Karlsson and Sorensen, Inc., Providence, RI, 1997

[19] Auersperg, J.; Dudek, R.; Michel, B: Investigation of the mechanical-thermal field coupling effect during quasi-static crack growth and application in micro system technology. *Proceedings of the German ABAQUS User Meeting*, Ulm 1995, pp. 138-152

المنارة للاستشارات

ww

Experimental validation of finite element modeling.

Dietmar Vogel, Chen Jian* and Ingrid De Wolf*
*Fraunhofer IZM, Berlin, Germany, * IMECvzw, Leuven, Belgium*

Key words: Validation of finite element modeling, strain field measurement, electronic packaging, micro-Raman spectroscopy, microDAC

Abstract: **In this paper, two techniques which offer very interesting features for experimental simulation model verification are presented: micro-Raman spectroscopy and microDAC, a micro speckle technique.**

1. INTRODUCTION

Finite element simulations are commonly used to study the thermo-mechanical characteristics of devices, chips and packages. These simulations need input parameters, such as for example the thermal expansion coefficient, the intrinsic strain, the Poisson ratio and the elastic modulus for the different materials used and this for different crystal directions. Often these material parameters are not known precisely, or they are assumed to be constant for the different crystal directions or temperatures. Mostly perfect adhesion and stress transfer between materials is assumed, etc. In addition the exact shape of the different layers of the sample under study has to be imported in the program and a good grid has to be defined. Especially at singularities (interfaces, corners) the FE programs can produce erroneous output values. It is clear that, although FE simulation is a very valuable tool, some experimental verification is required, especially when new materials and new products are analysed.

In this paper, we discuss two techniques which are very useful for the experimental verification of strain and stress predicted by FE methods: micro-Raman spectroscopy and microDAC. The first technique can directly

G. Q. Zhang et al. (eds.), Benefiting from Thermal and Mechanical Simulation in Micro-electronics, 113–133.
© 2000 *Kluwer Academic Publishers.*

provide information of elastic strain present in for example a silicon chip. Its application to packaging is rather new, but it certainly promises to be very useful in that domain. The latter probes non-elastic strain induced by external forces.

2. MICRO-RAMAN SPECTROSCOPY

2.1 Introduction

Raman spectroscopy is a non destructive optical technique, which makes use of the interaction of light (photons) with lattice (phonons) or molecular vibrations. Until a few years ago, RS (Raman Spectroscopy) was mostly applied in chemical studies as a complementary technique to infrared spectroscopy, giving information on the chemical composition and crystallinity of the sample. However, since the first reports of Anastassakis et al. [1] on the sensitivity of the Raman peak for mechanical stress, the technique has been applied more and more as a stress sensor [2]. The advantages of this technique, compared to other stress-probing techniques such as X-ray and wafer bending measurements are: a high spatial resolution, better than 1 μm; non-destructive; depth probing possible by changing the wavelength of the exciting laser or by measuring on a cleaved sample side; the technique is relatively simple. With powerful software, most work can be done automatically. A disadvantage, especially when applying the technique to measure stress in packages, is that not all materials have a Raman signal. For example, the stress in metal lines and solder bumps can not be studied by Raman spectroscopy. However, indirect information on the stress in such materials can be obtained by measuring the area around them [3].

In the following, we will first briefly discuss the experimental set-up of a Raman spectroscope, give the physical background, and illustrate the usefulness of the technique with some examples concerning the stress due to polishing and the stress in some packages.

2.2 Experimental set-up

Figure 1 shows a typical Raman spectroscopy instrument. Laser light is focused on the sample through a microscope. The scattered light is collected and its spectrum is analysed (pre-monochromator, spectrometer). This

المنارة للاستشارات
ww

spectrum contains information on the Raman signal of the sample. The sample is mounted on a computer controlled X-Y stage, which allows to scan it in steps of minimal 0.1 μm in X and Y direction. This allows to make automatically line scans, measuring the Raman spectrum along a pre-defined line with a pre-defined spacing between the measurement points, or 2-Dim scans in a rectangle.

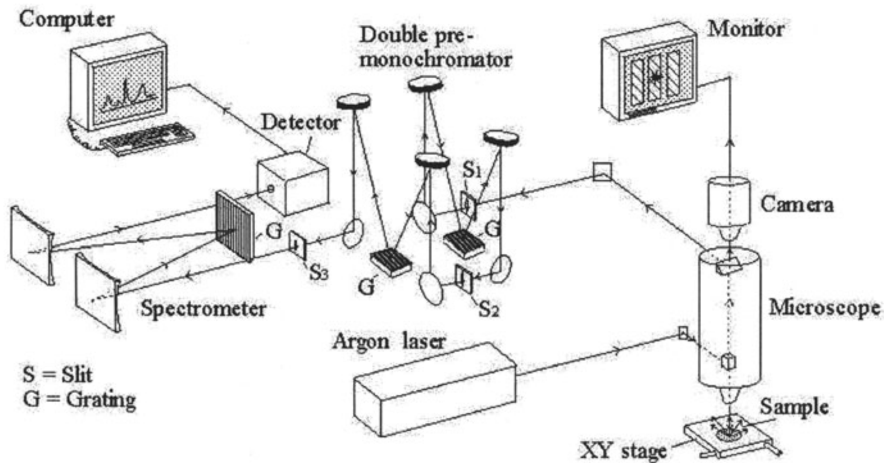


Figure 1. Set-up of Raman spectroscopy

A typical spectrum of crystalline Si is shown in Figure 2. The Raman peak of crystalline Si has a Lorentzian shape. In the spectrum, also two plasma lines from the laser are visible. They are used for calibration.

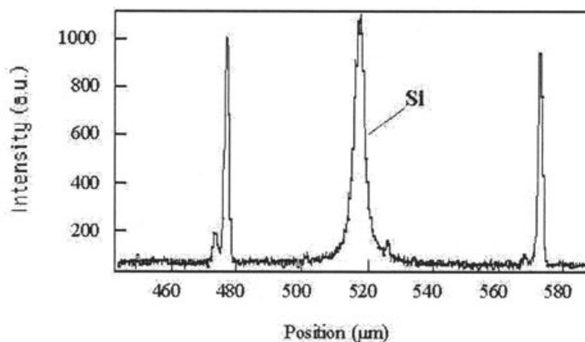


Figure 2. Typical Raman spectrum of crystalline Si. The two sharp lines at the left and right of the Si Raman peak are plasma lines from the laser

المنارة للاستشارات

2.3 Theory

The incident laser ray interacts with phonons in the sample. These phonons can take energy from the incident light (Stokes Raman scattering) or give energy to the light (anti-Stokes Raman scattering). So, the scattered light will have components with frequency equal to the one of the laser (ω_i, incident scattering), and with components $\omega_s = \omega_i - \omega_j$ (Stokes) and $\omega_s = \omega_i + \omega_j$ (anti-Stokes) where ω_j is the frequency of the phonon. At room temperature, the Stokes Raman signal has a higher intensity than the anti-Stokes Raman signal. For this reason, only the Stokes signal is considered for stress measurements using Raman spectroscopy.

When the material is under stress, it may cause a frequency shift in the Raman signal. In general, the relation between the frequency (ω) of the different vibration modes of a material and the strain is rather complex. The frequency is affected by all strain tensor components. For, silicon, in general only one Raman peak (one phonon mode) is observed. If there is no stress, this frequncy is equal to ω = 521 cm^{-1}. If one can assume uniaxial (σ) or biaxial ($\sigma_{xx} + \sigma_{yy}$) stress, the relation between the frequency of this mode and the stress is linear [1,2]:

$$\sigma\ (\text{MPa}) \approx -434 \times \Delta\omega\ (\text{cm}^{-1}) \quad \text{or} \quad \sigma_{xx} + \sigma_{yy}\ (\text{MPa}) \approx -434 \times \Delta\omega\ (\text{cm}^{-1})$$

From this follows that $\Delta\omega > 0$ indicates compressive stress, while $\Delta\omega < 0$ a tensile stress.

If more complex stress pictures are expected, such as for example at the edge of a film, or near a trench or LOCOS structure, the relation between $\Delta\omega$ and the stress tensor components is more complicated. All non-zero strain tensor components influence the position of the Raman peak. In order to obtain quantitative information on the stress in this complex case, some prior knowledge of the stress distribution in the sample is required. In other words, one has to presuppose a stress model. From this model, the expected Raman shift can be calculated and compared with the Raman data, and feed-back can be given to the model. Of course, some experimental parameters, such as the penetration depth of the laser light in the sample and the diameter of the focused laser beam on the sample have to be taken into account [3-5].

So, the frequency of the Si Raman peak depends on mechanical stress. In order to determine this frequency, the peak is fitted with a Lorentzian function. The plasma lines of the laser are used as reference. They are fitted using a Gauss function to determine their frequency. Variations in the position of the plasma line are due to instrumental variations (laser, room

المنارة للاستشارات

ww

temperature, detector temperature,...). The Raman spectra are corrected for these variations.

2.4 Experimental results

2.4.1 Stress due to nitride lines

As a first example, to demonstrate the principle of Raman spectroscopy measurements, we show results from stress measurements on very simple structures: lines on a silicon substrate.

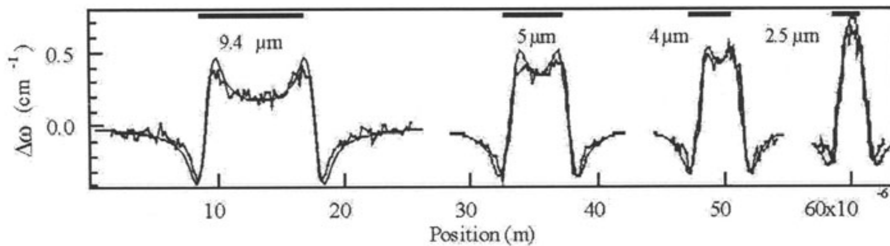


Figure 3. Δω (open symbols) measured on nitride lines with different widths. The 9.4 µm line is isolated, the others are part of an array of lines with width = spacing. The rectangles at the top indicate the position of the lines

An example of a µRS experiment on such a structure is shown in Fig. 3 [3]. This figure shows a map of the shift of the frequency of the Si Raman peak when scanning across Si_3N_4/poly-amorphous Si lines on a silicon substrate, with different widths. The position of the lines is indicated by the rectangles on top of the data. The laser is first focused far from the lines, where the stress can be assumed to be zero. Next the sample is moved, using an XY-stage, in steps of 0.1 µm and at each position a Si-Raman spectrum, as the one shown in Fig. 2, is recorded. A Lorentzian function is fitted to each of these Raman peaks in order to determine the frequency. The shift of this frequency from the stress free value, Δω, is plotted as a function of the position on the sample where the corresponding spectrum was measured. Fig. 3 shows that Δω becomes negative when approaching the lines, with maximal negative value just outside the line. When crossing the border, Δω changes sign very fast to reach a maximum positive value under the line, near the edge. Δω remains positive under the line, with some relaxation towards the centre. How can this be related to mechanical stress? In order to obtain an idea about the sign and the magnitude of the stress which corresponds to Δω, one can assume uniaxial stress, σ, along the width of the

المنارة للاستشارات
ww

line. This assumption is not too bad near the centre of the line, but it does not hold at the edges. Assuming uniaxial stress, we find that there is compressive stress under the lines and tensile stress near the edges. A shift of $\Delta\omega = 0.2$ cm^{-1}, as measured in the silicon under the centre of the wide nitride line, would then correspond with a compressive stress of -100 MPa. It is possible to obtain more detailed information on the different stress components by fitting a stress model to the Raman data [3]. This was done for these nitride lines using the so called analytical 'edge force model', where it is assumed that the stress imposed by the line in the silicon can be described by edge forces. A fit of this model to the µRS data, taking into account experimental parameters such as probing spot diameter and penetration depth, is shown in Fig. 3 (full lines). This procedure of fitting theoretical stress models to Raman data can be used for any model describing any device where Raman data can be measured. In this way, µRS can be used to experimentally verify stress models [3].

Another example is shown in Fig. 4 [4]. It concerns similar lines, but in this case the Raman data are compared to the stress predicted by a FE model. It is clear that the fit of the FE model to the data is very bad, indicating that some input parameters in the FE model were not correct.

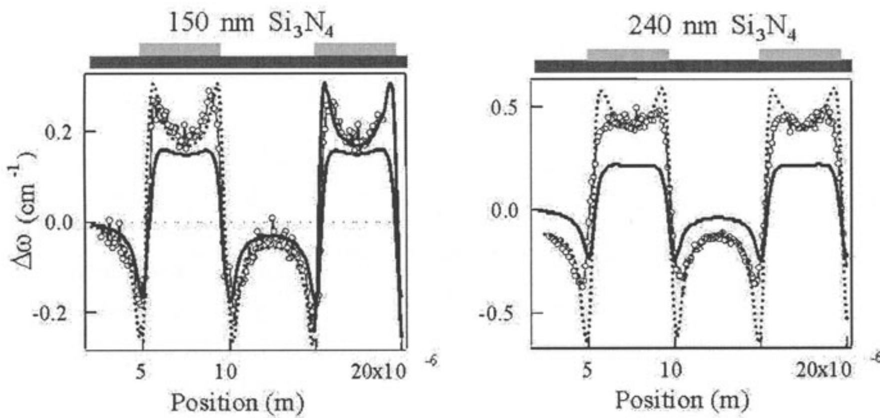


Figure 4. Comparison of the fit of an analytical model (dotted lines) and a FE model (full lines) to the Raman data (open circles)

2.4.2 Stress around solder bumps

FE simulations are often used to study packaging induced stress, and the effect of thermal cycling on this stress. It is for example often showed by FE that the stress near certain bumps in flip-chip packages may become very large. Raman spectroscopy cannot be used to study the stress in flip-chip

المنارة للاستشارات

bumps, because the PbSn material does not give a Raman signal. However, it can be used to study the stress in the silicon substrate near the bumps. An example of such an experiment is shown in Fig. 5. A package was cut and polished in cross-section. This allowed to probe the silicon chip next to the solder bumps, as indicated in the figure at the top of Fig. 5. The Raman data clearly indicate compressive stress in the silicon on top of the bump (C, positive Raman shift), and tensile stress near the edges. This experiment also shows that the stress may vary from bump to bump.

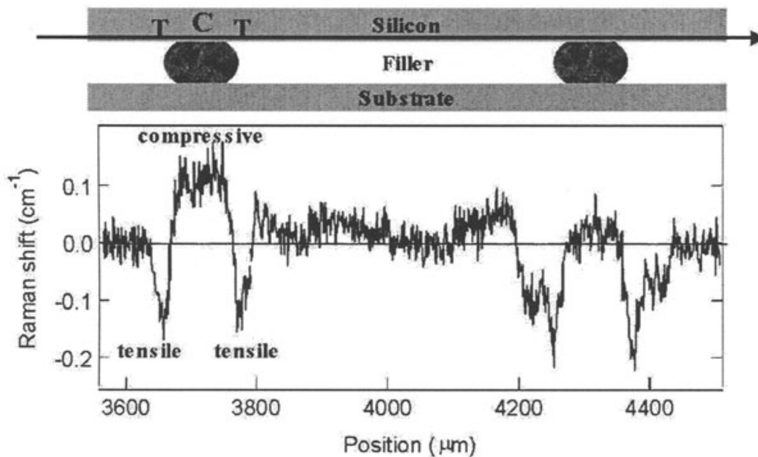


Figure 5. Stress induced by two solder bumps in the Si substrate

Experiments like the one shown in Fig. 5 are also possible in two-dimensions. A result of the stress measured near a solder bump is shown in Fig. 6. The top figure shows a schematic of the sample. Silicon substrate at the top, substrate at the bottom, and SnPb bump in between. The rectangular area in the Si near the bump, which is measured with micro-Raman spectroscopy, is indicated. At the bottom, the Raman shift measured in this area is shown. Notice a positive shift (white region: compressive stress) under the bump.

المنارة للاستشارات

ww

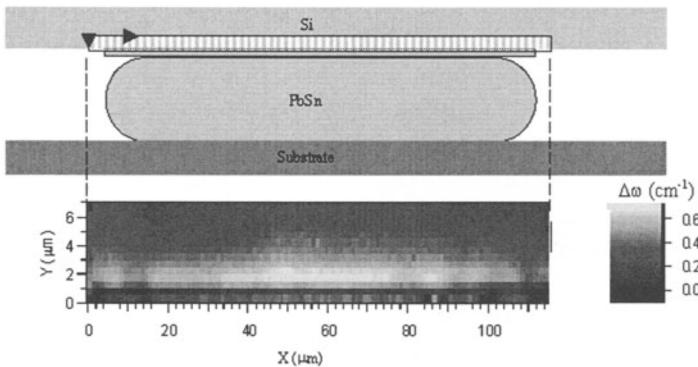


Figure 6. 2-Dimensional stress measurement in the Si near a solder bump

2.4.3 Stress in a packaged chip

As a last example, Fig. 7 shows the local stress near the edge of a glob-topped chip attached to a substrate. Two corners were investigated: the top right and the bottom right corner. Local high compressive stress (340 MPa) was measured in these corners. Similar pictures were obtained at the other edges of the chip. These stresses were not predicted by finite element simulations.

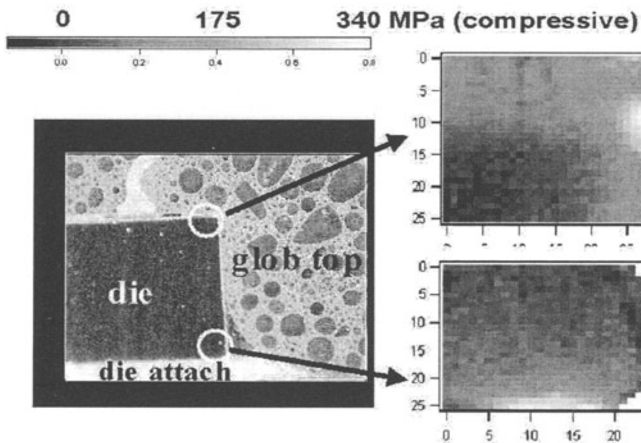


Figure 7. Stress near the edge of a packaged chip.

المنارة للاستشارات

2.5 Conclusions

We demonstrated in this paper that Raman spectroscopy can be applied for the study of mechanical stress in silicon devices and packaged chips. The applications of this technique to devices is well known [1], the application to packaging related problems started only the last two years. Until now, Raman spectroscopy was in the packaging field only applied to the study of the stress in the Si chip. However, as many polymers, plastics and glob-top may also provide a Raman signal, it is very likely that the technique will be applied in the near future also to measure the stress in these materials. This will make the Raman spectroscopy technique a very valuable tool for the experimental verification of FE models. Indeed, starting from the stresses provided by the FE models, it is possible to calculate the by the model predicted Raman shift, and to compare this shift with the measured one. This allows to give feed back to the FE models.

Since a lot of new materials and new structures are used in modern packages, stress simulations (e.g. FEM) are far from simple. They need a lot of assumptions which need to be verified. Raman Spectroscopy is certainly a very useful tool to measure stress. In the future, this technique will find more applications for FE model validation. Raman Spectroscopy is not new, but it is still very young.

المنارة للاستشارات

ww

3. MICRODAC® STRAIN MEASUREMENT FOR FINITE ELEMENT MODELING SUPPORT

3.1 microDAC® – a Universal Deformation Measurement Method – Its Basics

The basic idea of the microDAC® concept proceeds from the fact that microscopes of different kinds commonly allow to record tiny local and unique object patterns, besides of the more global object shape and structure. If they are temporary of stable position and appearance in the digitized images, they can function as a local image area marker. Commonly, these microstructures even remain stable during severe thermal and/or mechanical component load and can be recognized after load (see Fig. 8). A correlation based image processing algorithm is applied to determine a set of local pattern displacements between two object states, and finally whole in-plane displacement fields $u_x(x,y)$ and $u_y(x,y)$ are measured [6]. Numerical derivation of displacement fields results in in-plane strain components and the local rotation angle

$$\varepsilon_{xx} = \frac{\partial u_x}{\partial x}, \varepsilon_{yy} = \frac{\partial u_y}{\partial y}, \varepsilon_{xy} = \frac{1}{2}\left(\frac{\partial u_x}{\partial y} + \frac{\partial u_y}{\partial x}\right)$$

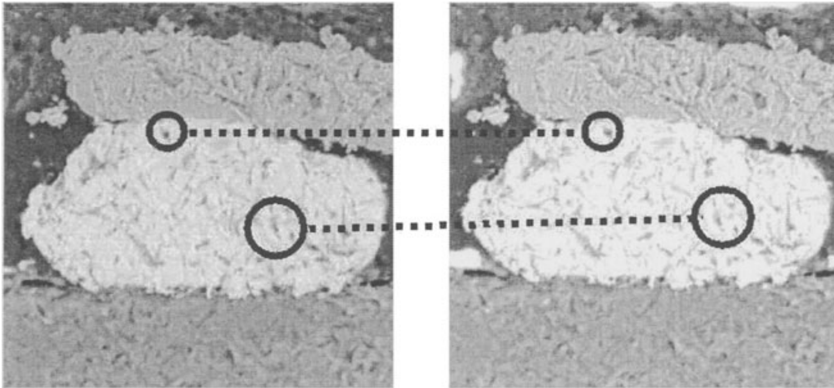


Figure 8. Appearance of local image structures (pattern) maintaining during specimen load, SEM images of a flip chip gold bump (left side: at room temperature, right side: at 125 °C), see pattern inside circles remaining recognisable during load

للاستشارات
ww

Within the past years different appropriate WINDOWS® based computer codes have been developed for efficient displacement and strain extraction from load state images. The codes are able to track local structures in digitized micrographs with a subpixel accuracy of approximately 0.1. Subpixel accuracy of 0.1 means that local displacements can be measured with a resolution of

$$\delta u = 0.1\, L / M$$

Here, L indicates the length of the field of view for the imaging equipment, M is the number of pixels along the image edge. The lateral image resolution of measurement points is a function of the local pattern size n. Non overlapping of neighboring searched pattern must be provided for independent displacement values, i.e. the lateral resolution equals

$$\delta l = L\, n / M$$

Where n is the pattern area pixel size. For the example of a typical low resolution SEM image of solder bumps the values $L = 100$ µm, $M = 1024$ and $n = 15$ are applied. Therefore, a measurement resolution of $\delta u \approx 10$ nm and a lateral structure resolution of $\delta l \approx 1.5$ µm can be achieved. For moderate displacement gradients over the image a measurement resolution in terms of strain of 1×10^{-3} can be achieved.

The microDAC software allows two alternative ways to extract displacement and strain data from load state images. In the first case deformation data is determined for a grid of equidistant measurement points. In the second case deformation values are found for the node points of a finite element mesh. The later approach implies that the mesh has been prepared for the object under investigation by the help of a FE preprocessor. Until now, meshes from ABAQUS and ALICE can be utilized. Their application aims at the direct comparison between results of finite element analysis and of corresponding measurements on real components. Fig. 9 shows a respective measurement example.

المنارة للاستشارات

ww

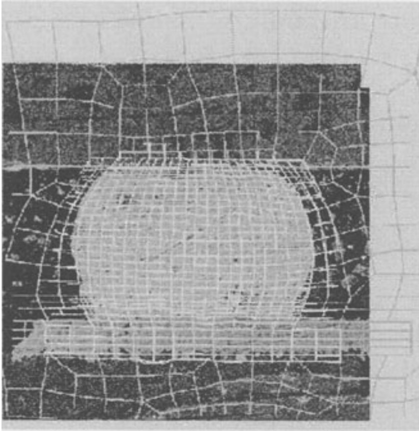
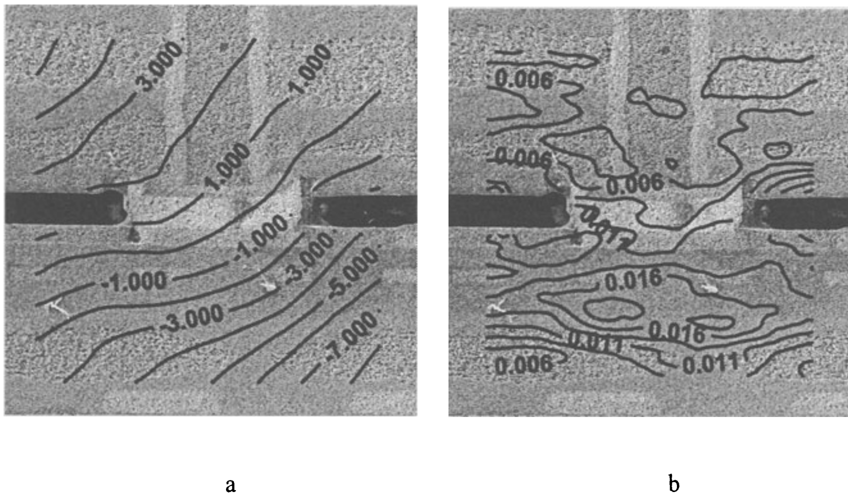


Figure 9: Solder bump deformation measurement utilizing an ABAQUS finite element mesh, gray mesh: part of the whole undistorted FE mesh, white mesh: deformed FE mesh obtained from a microDAC measurement (3 times enhanced displacement values), approx. micrograph field of view: 100 µm

The type of measurement data available without using predefined finite element meshes is illustrated on the following Fig. 10. The shown chip scale package (CSP) solder joint refers to a rigid carrier CSP, which had been mounted to an organic laminate. The whole device was heated up between room temperature and approximately 125 °C. Solder deformation did take place due to the thermal mismatch between different materials, namely between silicon and the organic laminates.



a b

Figure 10: Displacement and strain contour lines over the object image (solder joint of a chip scale package with rigid chip carrier), a – vertical displacement component, b – vertical strain component.

Displacement contourlines (Fig. 10a) superpose local object movements originating from rigid body displacements and rotations, strains and warpage

المنارة للاستشارات
ww

(change of local rotation angle with coordinates). Consequently, simple interpretation of displacement maps is not possible in many cases. In the opposite, strain data (Fig.10b) is free of rigid body movements. Commonly, it is used for comparison with FEA data.

3.2 Validation of FEA Results by Strain Field Comparison

3.2.1 Material Deformation as a Source of Solder Joint Failure

Solder fatigue due to thermal cycling is one of the essential concerns of mechanical reliability in flip chip technology. Accumulated solder creep strains during cycling finally result in the solder joint failure. The amount of acceptable strains has been determined in numerous papers [7]. E.g., the curve of Fig. 11 gives the mean cycles to failure for eutectic PbSn solder in dependence on the accumulated equivalent creep strain. Taking into consideration this data an accumulated creep strain of approx. 1 % is acceptable to reach 2000 thermal cycles to failure.

For the design of new devices finite element simulations are performed taking into consideration component geometry, desired load and material properties. Obtained stress / strain distributions are utilized in different kind of failure concepts, e.g. in fatigue models as mentioned above. In order to get realistic life time estimations from these concepts adequacy of mechanical modeling should be proven. One possible approach can base on the comparison of selected displacement or strain fields with those derived from measurements on real components.

المنارة للاستشارات

ww

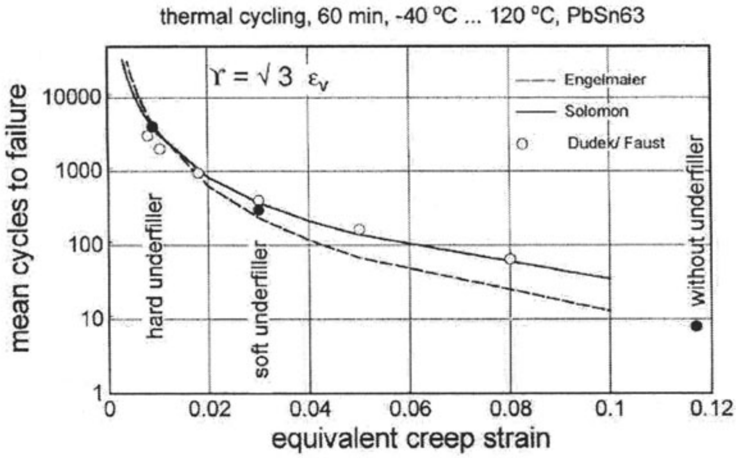


Figure 11: Dependence of "mean thermal cycles to failure" from accumulated equivalent creep strain per thermal cycle in eutectic PbSn solder.

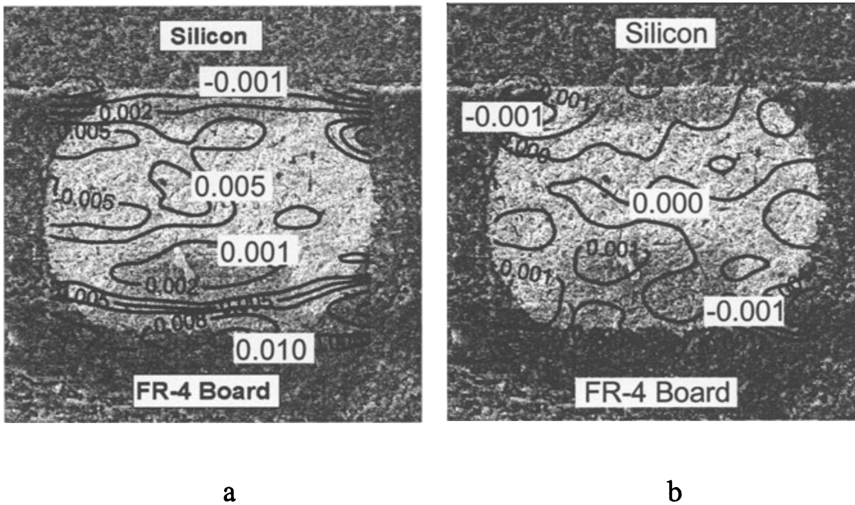


Figure 12: PbSn solder bump strain field for assembly heating from 25°C to 125°C, strain component ε_{yy} perpendicular to the die (a) and shear strain ε_{xy} (b), microDAC measurement inside scanning electron microscope, outermost bump

Fig. 12 illustrates how strain measurements can be performed by microDAC on cross sectioned flip chip samples. The strain fields of solder

المنارة للاستشارات

bumps reveal for thermal loading that strains perpendicular to the die plane are the major strain component. In the opposite, shear strains are suppressed by underfilling the die. They appear around places of very local material mismatch only, e.g. at the solder bump corners.

3.2.2 FEA Model Verification for Flip Chip Assemblies

Fig. 12 shows a typical behavior of PbSn solder bumps for thermally loaded flip chips on organic laminates like FR-4. The major ε_{yy} strain field exhibits a maximum value of less than 1 % in the bump middle. A similar second strain maximum occurs underneath the bump. Fig. 13a shows a respective simulated strain field, which has the same principal characteristics as the measured one.

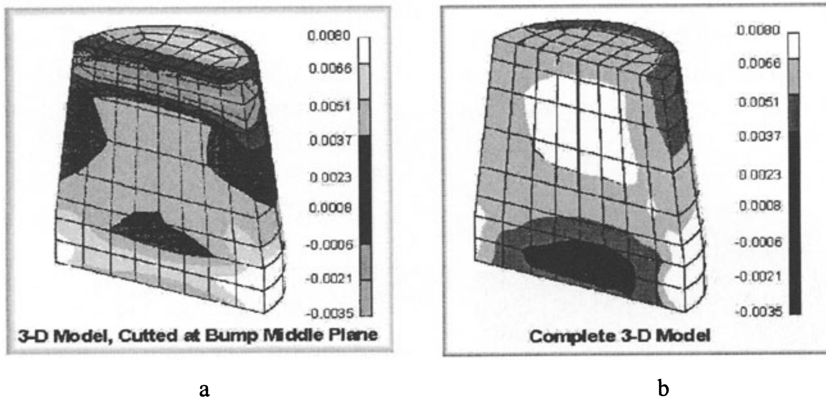
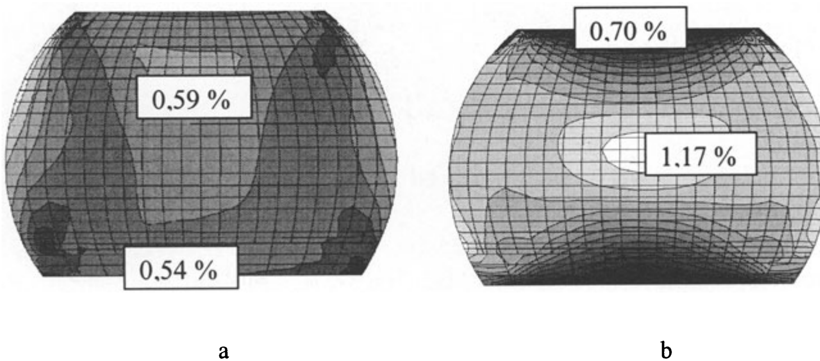


Figure 13: In-plane strain ε_{yy} (along bump axis) for thermal flip chip loading (heating from 25°C to 125°C),
a - simulation for a cut 3D model, b - simulation for a complete 3D model,



المنارة للاستشارات
ww

Figure 14: PbSn solder bump strain field for assembly heating from 31°C to 118°C, strain component perpendicular to board direction, FE simulation for *isotropic* (a) and *anisotropic* (b) underfill CTE's

The examples on Fig. 14 and 15 demonstrate the improvement of mechanical modeling by means of strain measurement. Fig. 14a and 15 show a simulated and measured strain field of an outermost bump, respectively. In both cases the maximum of strain in the direction perpendicular to the board is located in the inner part of the bumps. Nevertheless, the comparison of the theoretic and experimental data reveals a difference of a factor of 5 for the maximum strain values.

Measurements of the coefficients of thermal expansion (CTE) on thin separated underfill foils resulted in a distinguished anisotropy. CTE values perpendicular to the foil are 5.2 times higher than those along the foil plane (for CTE's below Tg, see also [8]). Taking into account the anisotropic CTE values a new improved simulation was carried out. Its results are shown on Fig. 14b. Maximum strain values have increased at least to a level of 1.2 %, i.e. the gap between measurement and theory has dropped down significantly.

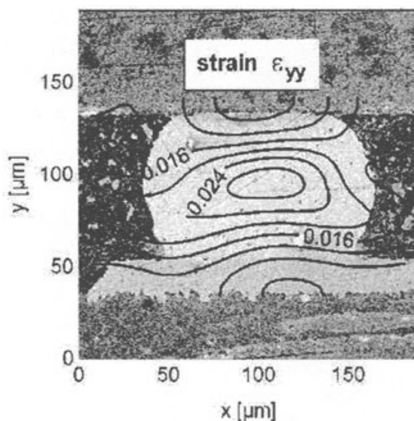


Figure 15: PbSn solder bump strain field for assembly heating from 31°C to 118°C, strain component perpendicular to board direction, microDAC measurement inside SEM

3.2.3 Adequacy between Simulated and Measured Strain Fields

Mechanical model improvement by comparison between FEA and measurement results assumes, that both, measurement and simulation, are accomplished for the same circumstances. Otherwise, uncertain benefit will be derived from the procedure. E.g., in the case of measurements on cross sectioned samples simulated strain fields principally have to be computed for

المنارة للاستشارات
ww

cut assemblies. Cross sectioning, e.g. of flip chip assemblies at the bump row plane can introduce changes in the deformation behavior, which generally may not be neglected. This effect is shown on Fig. 13 for a cone type bump. Obviously both local strain fields differ significantly from each other.

3.3 Prevention of Incorrect Simplifications on FEA Models

Restricted computer memory and run time resources as well as a missing data demand to keep mechanical models within a certain extend. Consequently, simplification for object geometry, material constitutive laws and loading conditions have to be made. Strain measurements can help to figure out undesired simplifications leading to wrong stress / strain distributions. At least they can be used to estimate possible error levels, if model refinement is impossible.

The following example refers to the influence of organic laminate structures to the deformation behavior of solder interconnects. Commonly, in FEA modeling their local structure is neglected. FR-4, one of the widely used board materials, is a laminate containing glass fabrics. The plain weave type causes distinct inhomogeneities in local strain/stress behavior, if thermal loading is applied. Commonly this fact is not considered analyzing reliability of mounted packages or dies. However, solder bump or ball size is less than the typical "wave length" of the yarn and the question for solder deformation analysis arises, whether this board inhomogeneity has to be taken into account or not. This was the objective to measure solder bump deformations for groups of adjacent bumps on a high pitch flip chip. So, the deformation differences caused by the varying distance to the neutral point of the die could be neglected for neighboring bumps. Consequently, possible changes in deformation of neighboring bumps should come from the varying epoxy height between bump pad and glass fiber.

المنارة للاستشارات

ww

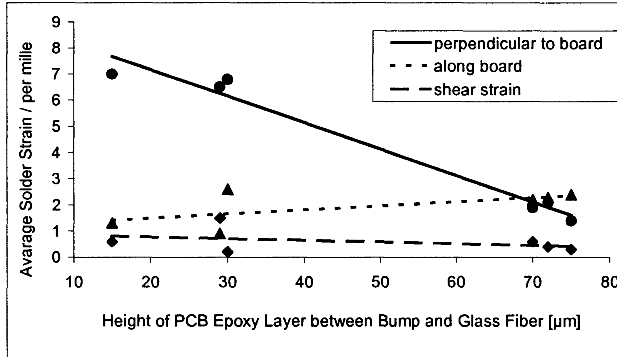


Figure 16: Dependence of average solder strain on bump position with regard to FR-4 glass fabrics, strain due to thermal heating between room temperature and approximately 125 °C, 10 mm x 10 mm die on FR-4

Fig. 16 demonstrates the results of the mentioned above strain measurements. The curves show the average in-plane strain values of eutectic PbSn solder bumps versus the local distance between bump pad and first fiber layer. The strain values of the plot are averages for different groups of bumps. As can be seen, for thermal flip chip loading the local epoxy thickness underneath the bump can influence dramatically the bump strain. Namely this refers to the main strain component ε_{yy} perpendicular to the board.

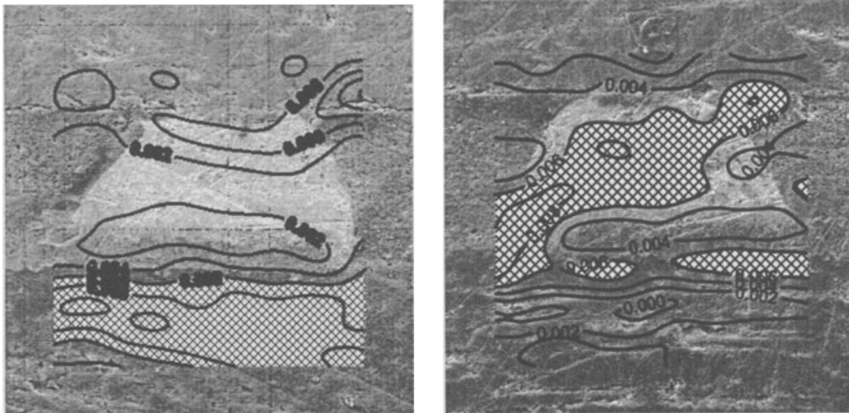
Fig. 17 allows to judge the whole ε_{yy} strain field in the bumps as well as in their vicinity. The flip chip of Fig. 17 has been loaded thermally, starting from room temperature and ending at approximately 125 °C. Regions of higher strain values are cross hatched. The right image refers to a bump with lower epoxy thickness. The left bump has been studied from the same specimen and is located nearby the other one. Its distance between bump and glass fibers is significant higher.

A remarkable change of strain pattern can be recognized, though not anticipated because of the similar shape and location of both bumps. A locally higher epoxy gap in between the bumps and the fabrics leads to a reduction of the bump strain. "Stress relaxation" now takes place inside the PCB epoxy, instead of the solder.

Experimental findings have been proven by modifying typical advanced flip chip models applied in thermo-mechanical analysis (see [9]). The somewhat simplified FE calculation presented in Fig. 18 assumes a glass fiber free area of 50 µm thickness exclusively beneath the second bump from the right die edge. Respectively, for that area poor epoxy material properties

المنارة للاستشارات
ww

have been introduced in the finite element model. The average equivalent creep strain values during one air-to-air temperature cycle have been found to be half a percent lower then in the neighboring solder bumps. I.e., the local board structure significantly influences the particular bump deformation.



thick epoxy layer underneath the bump thin epoxy layer underneath the bump

Figure 17: Strain redistribution due to PCB structure, sample heating approx. 25 °C ... 125 °C, strain component perpendicular to PCB, cross hatched areas: strain > 0,6 %

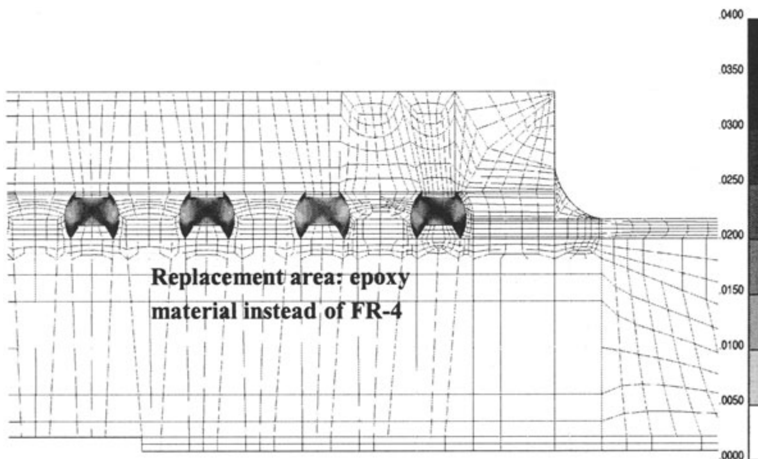


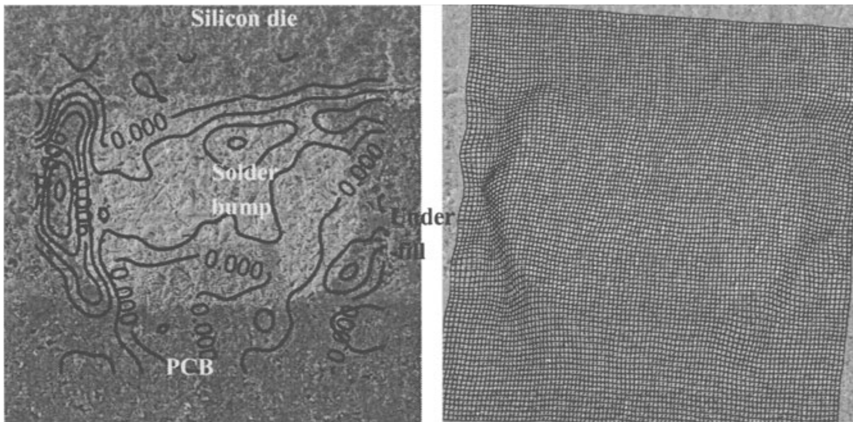
Figure 18: Equivalent creep strains in FC solder bumps after AATC with ΔT 150°C .. -40°C .. 150°C

المنارة للاستشارات

3.4 Defect Detection – Additional Data for FEA Modeling

The detection of failures and their inclusion in finite element models is essential for appropriate numerical simulation. microDAC allows to carry out such kind of failure analysis. The following example gives an idea about that type of measurement.

Measurement specimens for applications on flip chip have been cross-sectioned along a bump row and thermally loaded inside a scanning electron microscope. Measurements were done heating up the specimens from room temperature to levels above 100 °C inside a scanning electron microscope.



a- shear strain ε_{xy} b - deformed object grid

Fig. 19: PbSn solder bump strain for heating from room temperature over 100 °C, microDAC measurement inside SEM, delamination at bump/underfill interface

Fig. 19 shows defects occurring on solder bumps after thermal cycling. At the left bump interface to the underfill very high "pseudo" shear values appear (Fig. 19a). In fact, in Fig. 19b a distorted virtual object grid presentation at that place supports this result. The bump and underfill material slide against each other and a gap between bump and underfill closes, when the assembly is heated up.

4. REFERENCES

1. Anastassakis E., Pinczuk A., Burstein E., Pollak F. H. and Cardona M. 1970 Solid State Commun. 8 133-8
2. De Wolf I. Topical Review, Semicond. Sci. Technol. 11, 139-154, 1996.

المنارة للاستشارات
ww

3. De Wolf I., Maes H.E. and Jones S. K.. J. Appl. Phys. 79(9), pp. 7148-7156, 1996.
4. De Wolf I., G. Pozza, K. Pinardi, D.J. Howard, M. Ignat, S.C. Jain and H.E. Maes. Microelectron. Reliab., Vol. 36, No. 11/12, pp. 1751-1754, 1996. (Proceedings ESREF '96, Enschede, The Netherlands)
5. De Wolf I., M. Ignat, G. Pozza, L. Maniguet and H.E. Maes. J. Appl. Phys. 85(9), pp. 6477-6485, 1999
6. D. Vogel, A. Schubert, W. Faust, R. Dudek, B. Michel: MicroDAC - A novel approach to measure in situ deformation fields of microscopic scale, Microelectronics Reliability, Vol. 36, 1996, No. 11/12, pp. 1339-1342.
7. A. Schubert, R. Dudek, D. Vogel, B. Michel, H. Reichl: Material Mechanics and Mechanical Reliability for Flip Chip Assemblies on Organic Substrates, Advancing Microelectronics, July/Aug. 1997.
8. Uhlig, C., Kahle, O., Vogel, D., Bauer, M., 1997, The application of thermophysical methods in microsystems and microelectronics, Proc. of Micro Materials '97, Berlin, pp. 1097-1097.
9. Schubert, A., Dudek, R., Michel, B., Materials Mechanics and Thermo-Mechanical Reliability of Flip Chip Area Array Packages, Proc. of InterPack '99, Maui (USA), June 13-17, 1999, pp. 181-187.

المنارة للاستشارات

ww

Perspectives of Non-Linear Simulation

M.A. Crisfield and A.J. Burton
Dept. of Aeronautics, Imperial College, London, U.K. and FEA Ltd., Kingston-upon-Thames, Surrey, U.K.

Key words: finite elements, delamination, adaptivity

Abstract: This paper gives an overview of the current state of non-linear finite element computer programs in relation to the analysis of structures. In addition, it concentrates on two specific areas, namely the introduction of techniques for adaptive re-meshing and the simulation of the propagation of delamination.

1. NON-LINEAR FINITE ELEMENT PACKAGES

There are two main types, which we will here describe as 'implicit' and 'explicit'. The former is the natural follow-on from conventional linear analysis. Indeed, in its earliest implementation, the loading was cut into a series of increments and for each of these a 'linear analysis' was performed using the tangent stiffness matrix. It was soon realised that, unless the increments were very small, such a formulation could lead to an unacceptable drift from equilibrium and so the earlier 'incremental procedure' was supplemented with 'iterations' designed to restore full equilibrium at each step. For dynamics, the inertia terms are conventionally added at the end of each step, in conjunction with the Newmark trapezoidal rule [1] or the 'alpha modification' [2,3], which introduces 'numerical damping'. Again, a combined incremental/iterative procedure is adopted.

For high frequency excited dynamics problems, the required time-step can be very small. For this reason, 'explicit codes' were developed which dispensed with the equilibrium iterations and indeed also dispensed with the 'equation solving', because the explicit formulation leads naturally to an approach in which only the mass-matrix need be inverted. Once this

G. Q. Zhang et al. (eds.), Benefiting from Thermal and Mechanical Simulation in Micro-electronics, 135–149.
© 2000 *Kluwer Academic Publishers.*

ww

procedure is combined with a diagonalised, 'lumped mass' idealisation, the equation solving is no longer required. Explicit techniques [3,4] are 'conditionally stable' and require the critical time-step to be below a threshold value, which is usually computed from an element-level computation involving the 'smallest and stiffest element'. For low frequency excited problems, or indeed in the limit for pseudo-static problems, the required time-step may often become sufficiently small that implicit methods are more economic. The latter observation should really be re-phrased to read implicit methods should be more economic. In practice, for certain problems, implicit methods still have serious convergence problems (which will be discussed in more detail later in this paper), so that analysts may then use 'explicit techniques', even though the loading process may have to be speeded-up or the mass artificially increased. Most commercial explicit codes now introduce checks on the kinetic energy so that, for a pseudo-static problem, a user should not accept a solution unless this energy is below an acceptable threshold.

Returning attention to implicit codes, as larger and larger problems are tackled, it is clear that the 'equation solving' time can become a severe problem for implicit methods. Recently, a range of 'sparse-matrix' solvers [5] have been introduced into commercial packages and can reduce run-times by a factor up to the order of ten. In addition, there are 'iterative solvers' for which the main problem is the issue of 'robustness'. The required number of iterations is related to the 'condition number' of the stiffness matrix and many engineering problems are badly conditioned. For example, a solid 'engine block' may be well conditioned whereas a thin shell structure will not be. Iterative techniques will often have difficulties in achieving a sufficiently low 'convergence error norm' (here convergence refers to the linear system) for the latter problems, although improvements are continually being made to 'pre-conditioners'.

There is a close link between the 'iterative solvers' for implicit finite element methods and the use of explicit finite element methods for 'pseudo static problems'. Indeed, one may argue that where the latter may define an acceptable solution when the kinetic energy is small, the former might define an acceptable solution when 'the convergence error norm' is small but still significant. If one is only interested in the overall behaviour and not detailed local stresses, this may be true for a 'one-off' linear-static analysis. However, for a combined incremental-iterative non-linear analysis, the errors can accumulate and may lead to 'instabilities'.

For dynamic analysis, conventional wisdom argues that implicit methods are 'unconditionally stable'. However, this is not necessarily true for problems with geometric non-linearity. In particular, the commonly used Newmark trapeziodal rule is not unconditionally stable for the latter

المنارة للاستشارات

ww

problems [3,6,7]. However, there has recently been a growth of interest in new 'energy/momentum conserving procedures' [3,6,7], which overcome these problems. These techniques have some links with the explicit methods in that a form of 'mid-point dynamic equilibrium is used' [3].

Before leaving the topic of stability, it is worth mentioning another area in which issues of 'stability' can still pose a problem for both implicit and explicit finite element methods. This is the area of 'incompressibility' and 'near incompressibility'. Such issues are associated with 'rubber' and 'von-Mises plasticity'. Conventional lower-order displacement-based finite elements do not pass the LBB or 'inf-sup' condition [8] and can encounter stability problems. Many modified elements have been developed involving 'incompatible modes' or 'enhanced strains' or 'mixed formulations' [3]. These improve the performance but conceptually a difficulty remains. In the field of 'fluid dynamics', incompressibility is a big issue and analysts either use elements that satisfy the 'inf-sup' condition or by-pass it via forms of 'stabilisation' [9]. Both of these solutions lead to considerable complexities and in solids, analysts have not often directly confronted the issue. Frequently, there are no problems and this may well be because solids are not fully incompressible and in plasticity problems there are always regions of elasticity. Nonetheless, problems with instabilities do occur [10] with finite element formulations for solids involving near-incompressibility and work is still being undertaken to produce more effective element formulations.

A further area in which much work is required is that of contact, particularly with large deformations. For such simulations there is first the problem of keeping track of the continuously evolving contact area. Once this is achieved, there is the problem of matching the meshes on either side of the interface, where in general the nodes do not coincide. As a consequence, there are real difficulties with the 'shape functions' and integration procedures over the 'contact zone'. These difficulties have lead most analysts to avoid a direct integration procedure and instead to introduce a node on surface contact in conjunction with either a one-pass or two-pass procedure [3]. These techniques have usually been introduced with lower-order elements using 'slidelines' [11]. However, there are serious disadvantages. In particular, a curved surface is now effectively modeled by a series of facets and non-smooth responses can occur, as a node 'slides' from one element to another. This non-smoothness, or 'chatter', can wreak havoc with the iterative solution procedure.

As a consequence, many researchers are considering forms of 'curved contact', usually still with the aid of a node on surface approach and often using splines or Bezier interpolations [12]. The authors have considered an alternative approach, involving both low-order (linear) and higher-order

المنارة للاستشارات

idealisations in the contact region. The former are used for the weighting functions associated with the nodal forces, while the latter is used for the contact geometry [13].

While contact provides one form of non-smoothness, there is another that can cause similar serious difficulties with the non-linear solution procedures; this is the problem of softening which is associated with the real failure modeling of materials. Modern non-linear finite element codes have no problems with hardening plasticity or perfect-plasticity but encounter real difficulties once softening plasticity is considered. The same is true of cracking and fracturing materials. In a continuum environment, the problem can lead not only to convergence problems but also to solutions that are extremely mesh-sensitive [14]. Much research work has involved forms of 'non-local damage' [15], including 'gradient plasticity' [16] and in a recent paper, a formulation involving a complete gradient formulation (with the stress being related to not just the strain but the gradient of the strains) [17]. This kind of work, which does not fit easily into conventional finite element architectures, does not yet seem to have been introduced into commercial finite element packages.

Some of the difficulties associated with 'softening materials' are removed if one considers an interface surface rather than a continuum. The use of 'interface element' formulations [18,19], also called 'de-cohseive zone models' [20], with pre-defined options for the cracking delaminations will be addressed later in this paper. However, even for this restricted class of problem there is a need for fine meshes in the region of the crack front and this in turn, leads to a requirement for the adaptive refinement of the mesh. This is another topic that will be described in more detail later in the paper.

Finally, before completing this overview, it is worth noting that some of the requirements discussed above, for tracking and refining meshes in the regions of cracking fronts, have led some researchers to investigate a different form of idealisation to the conventional finite element method. In particular, there is currently much research work being devoted to various forms of 'meshless method' [21]. It is probably too early to assess whether or not these methods will have a big commercial impact.

2. ADAPTIVITY AND SHELLS

Instead of using a fixed mesh, significant computational advantages can be gained by refining the mesh in the regions where the errors are greatest. Most work in this area has been related to linear analysis, but here we will describe an application to non-linear analysis using shells. The

المنارة للاستشارات

ww

adopted shell element is a non-linear implementation [22] of the Morley faceted triangular shell element [23] which has three translations at each corner node and one rotation at each mid-side. Because of the triangular geometry, the formulation is particularly easy to re-mesh. The adopted error measure or error norm is:

$$\|e\| = \left(\int(\mathbf{N}_h - \mathbf{N}^*)^T(\boldsymbol{\varepsilon}_h - \boldsymbol{\varepsilon}^*)dA + \int(\mathbf{M}_h - \mathbf{M}^*)^T(\boldsymbol{\kappa}_h - \boldsymbol{\kappa}^*)dA\right)^{1/2} \quad (1)$$

This error measure can be considered as a special form of the Zienkiewicz-Zu error estimate [24] and follows on from earlier work in [25]. In equation (1), $\mathbf{N}$ and $\mathbf{M}$ denote the in-plane and bending resultants respectively, whilst $\boldsymbol{\varepsilon}$ and $\boldsymbol{\kappa}$ represent the in-plane strain and curvatures. The subscript h implies that the quantity is obtained directly from the basic finite element formulation, while the superscript * implies that the quantity is obtained from a smoothed post-processing analysis. To obtain the smoothed values, the stress-resultants from adjacent elements are averaged using a superconvergent patch recovery technique. However before this can be achieved, they must be transformed to a common tangential framework. By expressing the error norm in its current form, which uses both the 'strains' and the 'stresses', it can be applied to an analysis that includes material non-linearity.

At the end of each load increment during a non-linear analysis a global relative error for the entire mesh is computed. This is associated to the error measure defined in equation (1). When this global error exceeds a prescribed tolerance, the non-linear analysis is halted and re-meshing is automatically activated. Local normalised nodal error values that define the spacing for the new mesh are obtained from equation (1).

To continue the analysis with geometric non-linearity alone, only the 'displacements' need to be interpolated from the old mesh to the new mesh during the refinement procedure. With the exception of the mid-side rotation terms this is fairly straightforward. However, for the current non-linear implementation of the Morley triangle [22], the mid-side rotations take a rather special form and their interpolation is more complex. Details are given in [25].

The current work will consider both material and geometric non-linearity. Some analysts interpolate the stresses as well as the strains. However, we aim to transfer as little information as possible and only interpolate the plastic strains and equivalent plastic strains in addition to the 'displacements'. Following interpolation, the structure will no longer be in equilibrium and hence iterations are required to re-establish equilibrium before the analysis continues.

المنارة للاستشارات

ww

The following example was undertaken using the non-linear finite element system LUSAS [26] and relates to the pinched cylinder defined in Figure 1. Material properties are defined in Table 1 where Cp is the linear hardening stiffness for the adopted von-Mises yield criterion.

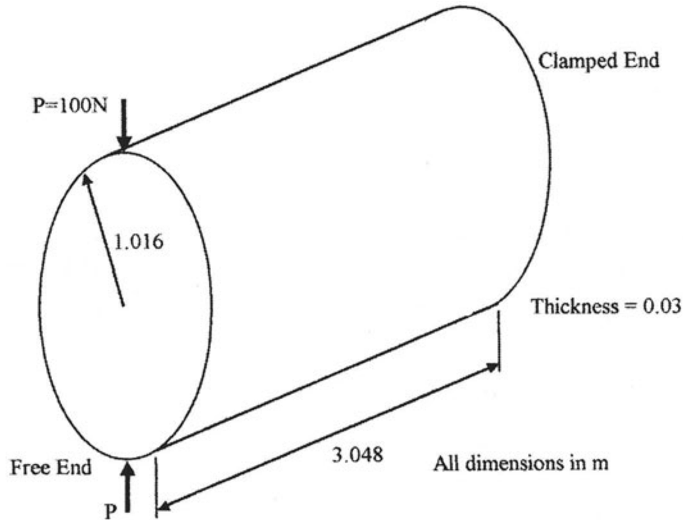


Figure 1. Pinched Cylinder: Geometry and boundary conditions

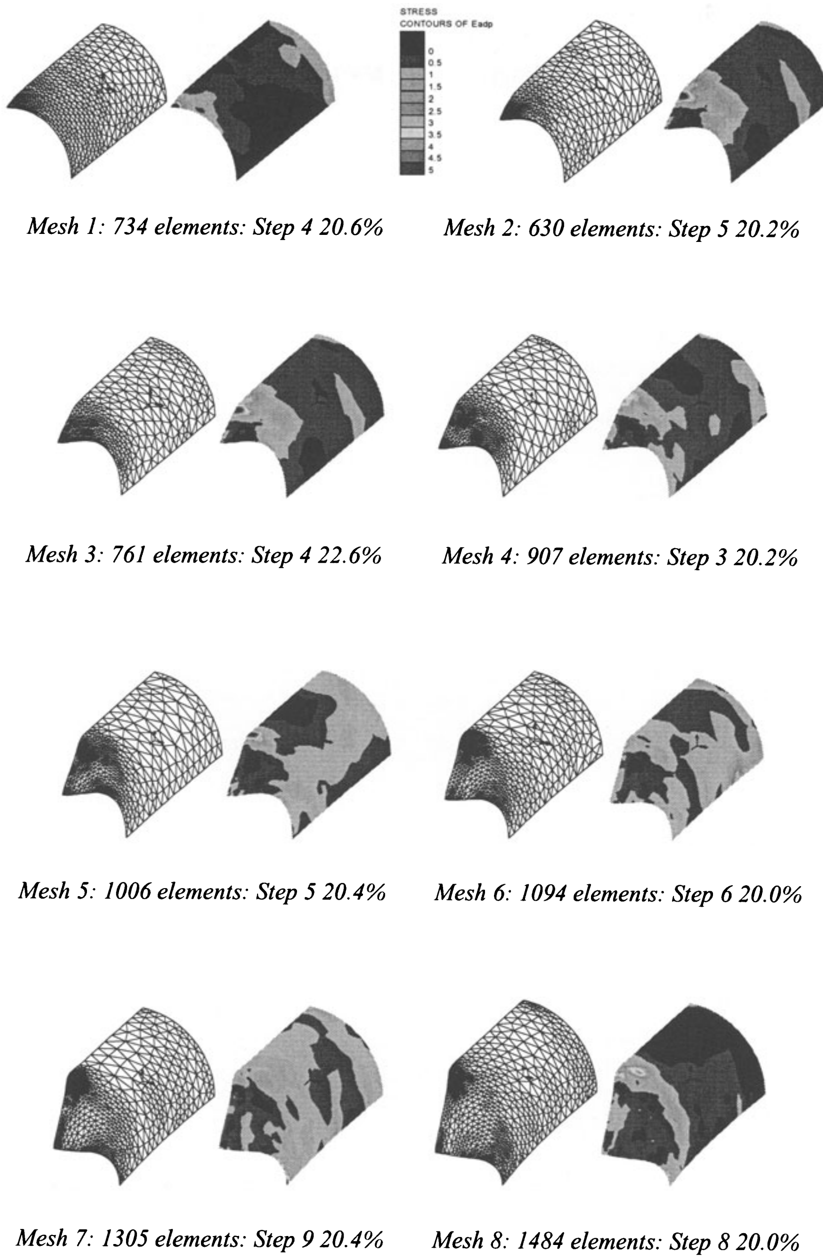
Table 1. Pinched Cylinder: Material Properties

E (N/m^2)	ν	σ_Y (N/m^2)	C_p (N/m^2)
2.0685×10^7	0.3	2.43×10^5	6.6192×10^5

The deformed shapes and contours of normalised nodal error are illustrated for meshes 1-8 in Figure 2. It can be seen that the adaptive procedure has led to a concentration of elements around the continuously evolving 'hinge lines'. (It is worth noting that the 'point load' of Figure 1, was in fact distributed over a finite patch, to avoid the concentration of elements around a singularity).

Figure 3 shows the computed load/deflection response for the complete analysis with the separate symbols relating to the different meshes. This compares extremely well to the solution shown in Figure 4, which was obtained from a standard analysis using an alternative element. The development of global relative error that drives the re-meshing procedure can be seen in Figure 5.

ww



Mesh 1: 734 elements: Step 4 20.6%

Mesh 2: 630 elements: Step 5 20.2%

Mesh 3: 761 elements: Step 4 22.6%

Mesh 4: 907 elements: Step 3 20.2%

Mesh 5: 1006 elements: Step 5 20.4%

Mesh 6: 1094 elements: Step 6 20.0%

Mesh 7: 1305 elements: Step 9 20.4%

Mesh 8: 1484 elements: Step 8 20.0%

Figure 2. Adaptive analysis: Deformed Configurations and contours of normalised nodal error

المنارة للاستشارات
ww

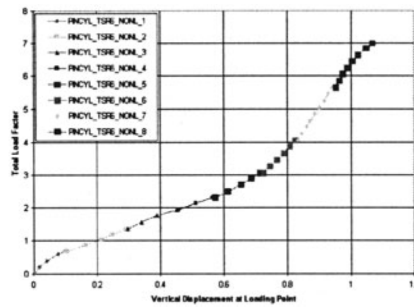


Figure 3. Adaptive analysis: Vertical displacement vs. total load factor at loading point

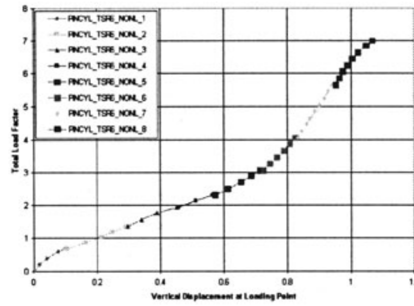


Figure 4. Standard analysis: Vertical displacement vs. total load factor at loading point

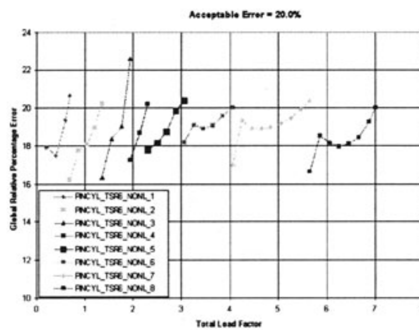


Figure 5. Development of global relative error during the adaptive analysis

المنارة للاستشارات
ww

3. INTERFACE ELEMENTS FOR DELAMINATION, INCLUDING ADAPTIVITY

Delamination is traditionally handled using a direct application of fracture mechanics [27]. For three-dimensional problems in particular, there can then be significant difficulties if we wish to trace the behaviour of the propagating crack front. There are sometimes also problems associated with the specification of a flaw size. In concept, some of these difficulties can be overcome by introducing an interface element formulation [18,19] or equivalent de-cohesive zone model [20]. Here fracture mechanics is indirectly introduced by providing a softening relationship between the traction and the opening displacement, see Figure 6. The area under this curve is set to equal the critical energy release rate while the initial stiffness is high and effectively provides a penalty, which also prohibits penetration in compression.

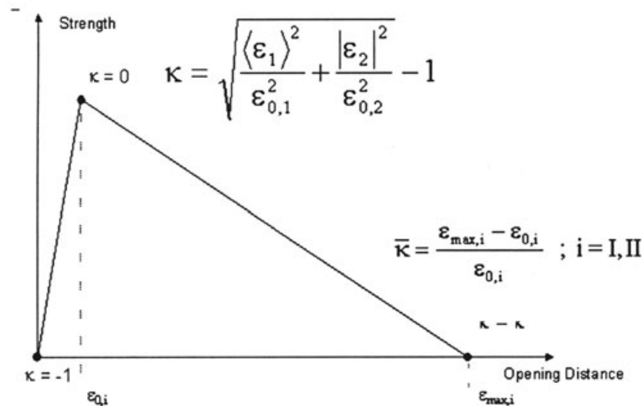


Figure 6. Softening relationship for the mixed mode interface model

Mixed mode fracture can also be handled and the tractions, $\boldsymbol{\sigma}^T = (\sigma_I, \sigma_{II})$ (considering here a line interface), are related to the relative displacements, $\boldsymbol{\varepsilon}^T = (\varepsilon_I, \varepsilon_{II})$, where the subscripts I and II relate to the opening and shearing modes respectively, by a form of damage relationship:

$$\boldsymbol{\sigma} = \left[\mathbf{I} - \frac{\kappa}{1+\kappa} \mathbf{F} \right] \mathbf{E}_0 \boldsymbol{\varepsilon} \tag{2}$$

The scalar κ in equation (2) allows for the interaction between mode I and II and for given opening displacements, ε_I and ε_{II} is given by:

المنارة للاستشارات
ww

$$\kappa = \left[\left(\frac{\varepsilon_{I}}{\varepsilon_{0I}} \right)^{\alpha} + \left(\frac{\varepsilon_{II}}{\varepsilon_{0II}} \right)^{\alpha} \right]^{1/\alpha} - 1 \tag{3}$$

where ε_{oI} and ε_{oII} are the initial cracking values of the relative displacements, see Figure 6. The scalar α is an input parameter that will be defined shortly. Furthemore, in equation (2), $\mathbf{E}_0$ is a diagonal matrix containing the initial stiffness in modes I and II respectively, so that:

$$\mathbf{E}_0 = \begin{bmatrix} \frac{\sigma_{tI}}{\varepsilon_{0I}} & \\ & \frac{\sigma_{tII}}{\varepsilon_{0II}} \end{bmatrix} \tag{4}$$

where σ_{tI} and σ_{tII} are the cracking strengths in the two modes. The matrix $\mathbf{F}$ in equation (2) is a diagonal matrix given by:

$$\mathbf{F} = \begin{bmatrix} \frac{\varepsilon_{max\,I}}{\varepsilon_{max\,I} - \varepsilon_{0I}} & \\ & \frac{\varepsilon_{max\,II}}{\varepsilon_{max\,II} - \varepsilon_{0II}} \end{bmatrix} \tag{5}$$

In equation (5), the terms ε_{maxI} and ε_{maxII} are the input openings, see Figure 6, at which the tractions have fallen to zero. These input opening displacements are directly computed from the critical energy release rates in the two modes. It can be shown that [18,19], if the previous relationships are used and a monotonic opening is applied with $\varepsilon_{II} = p\varepsilon_{I}$, where p is the fixed ratio, then the following interaction relationship is recovered:

$$\left(\frac{G_{I}}{G_{IC}} \right)^{\alpha/2} + \left(\frac{G_{II}}{G_{IIC}} \right)^{\alpha/2} = 1 \tag{6}$$

This interaction relationship is of the form often used in practice, that is, where α=2 [28] or where α=4 [29]. It is worth noting that with the mixed mode formulation, the variation of equation (2) leads to a non-symmetric tangent stiffness matrix [18,19] (as is the case with frictional contact [3]).

While analyses can be undertaken without adaptivity [18,19], it is found that to obtain a satisfactory convergence for the equilbrium iterations,

المنارة للاستشارات

ww

with a reasonably smooth load/deflection response, it is necessary to provide a fine mesh of elements around the crack front [19]. As a consequence we have introduced an adaptive refinement strategy along the interface elements. To this end, we have used the κ parameter of equation (3) to guide the refinement process.

In particular, see Figure 6, we determine the delamination front by locating nodes of the interface elements that are currently softening, that is, nodes whose value of κ lies between 0 and $\overline{\kappa}$. The mesh is refined in this area and fine elements are also placed ahead of and behind this 'softening zone' to enable the crack to propagate during subsequent load increments.

This strategy has been applied to the analysis of a simple Double Cantilever Beam Specimen (DCB), which is defined in Figure 7. Figure 8 shows a selection of the deformed meshes together with the nodes that are softening in mode I (denoted by the crosses) at various stages during the analysis. In addition, Figure 9 shows the computed load/deflection response at the open end of the beam and indicates the areas at which re-meshing has been activated. Figure 10 shows that the adaptive analysis gives a very similar result to an alternative analysis, with a uniform fine mesh, without adaptivity.

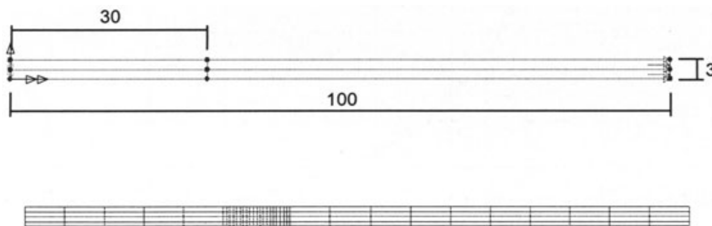


Figure 7. Double cantilever beam: Model and initial mesh configuration

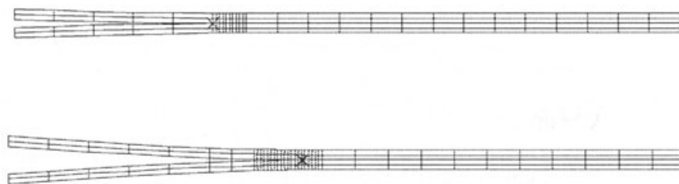


Figure 8a. Adaptive DCB test: Deformed configurations

المنارة للاستشارات
ww

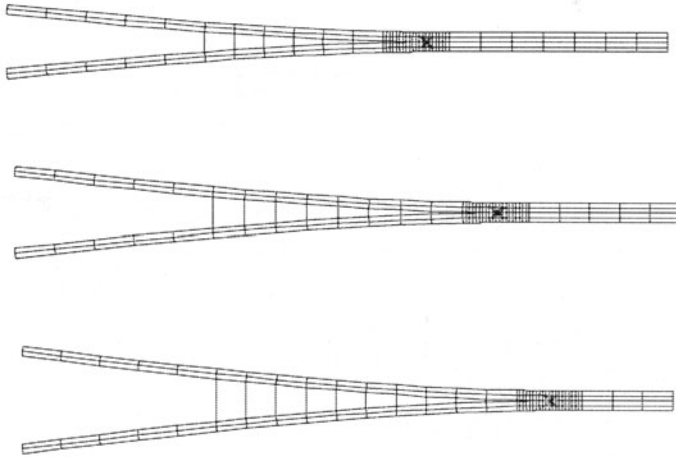


Figure 8b. Adaptive DCB test: Deformed configurations

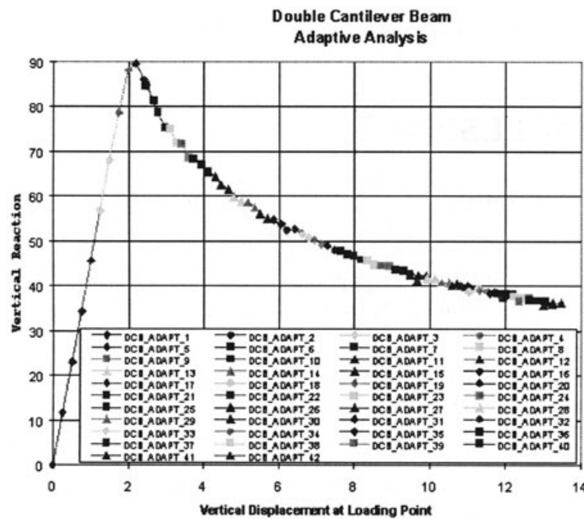


Figure 9. Adaptive DCB test: load deflection response

With adaptivity there is a considerable saving in computer time. The adaptive analysis took only 12.5% of the time taken for the standard analysis. The numerical results for this problem also show an excellent agreement with experimental results [18,19]. We have used our interface element formulation to tackle many more complex problems than the simple DCB described here [18,19] but the current work illustrates our first steps with adaptivity.

المنارة للاستشارات

ww

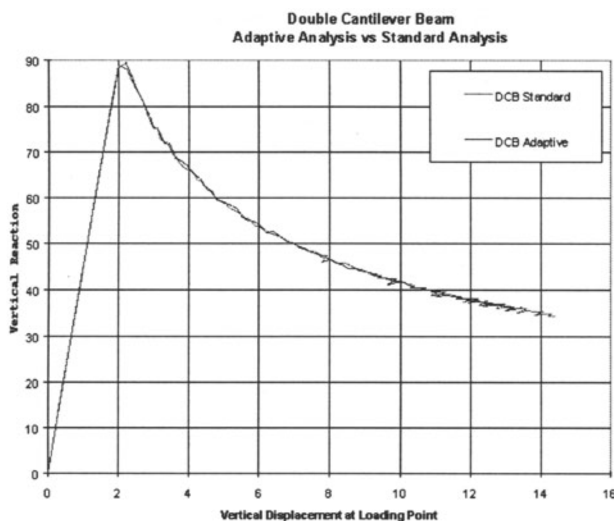


Figure 10. Comparison of adaptive and standard analyses

4. CONCLUSIONS

The paper has given an overview of the current state of non-linear finite element codes in relation to the analysis of structures. It has also given a more detailed description of two selected topics; adaptivity and the modelling of delamination in composite structures.

ACKNOWLEDGEMENTS

The examples presented in this paper were conducted as part of two Brite-Euram projects; BE-3580 and BE-97-4517.

REFERENCES

1. Newmark, N.M., A method of computation for structural dynamics, J. Eng. Mech. Div., ASCE, EM3, 85, 67-94,(1959)
2. Hilber, H.M., Hughes, T.J.R. & Taylor, R.L., Improved numerical dissipation for time integration algorithms, Earthquake Engineering and Structural Dynamics, 5, 283-292, (1977)

المنارة للاستشارات

ww

3. Crisfield, M.A., Non-linear finite element analysis of solids and structures, Vol. 2, Advanced Topics, J. Wiley, Chichester, 1997.
4. Belytchko, T., A survey of numerical methods and computer programs for dynamic structural analysis, Nucl. Engng. & Design, 37, 23-34, (1976)
5. SDR/LUSAS/169 Fast solver implementation, FEA Ltd., Kingston upon Thames, Surrey, U.K., (1999)
6. Simo, J.C.& Tarnow, N., The discrete energy-momentum method. Conserving algorithms for nonlinear elastodynamics, Z angew. Math. Und Phys., 43, 757-792, (1992)
7. Crisfield, M.A. & Shi, J., A co-rotational element/time integration strategy for non-linear dynamics, Int. J. for Num. Meth. In Engng., 37, 1897-1913 (1994)
8. Szabo, B. & Babuska, I., Finite Element Analysis, J. Wiley, 1991.
9. Silvester, D., Optimal low order finite element methods for incompressible flow, Comp. Meth. In Appl. Mech. & Engng., 111, 357-368, (1994).
10. Crisfield, M.A., Moita, G.F., Jelenic, G. & Lyons, L.P.R., Enhanced lower-order element formulations for large strains, in: Computational Plasticity, Fundamentals & Applications, Part 1, ed. D.R.J. Owen et al., Pineridge Press, Swansea, 293-320 (1995) (and Computational Mechanics, 17, 62-73, (1995))
11. Hallquist, J.O., Goudreau, G.L. & Benson, D., Sliding interfaces with contact-impact in large-scale Lagrangian computations, Comp. Meth. In Appl. Mech. & Engng., 51, 107-137, (1985).
12. Heegard, J.H., Curnier, A., Geometric properties of 2D and 3D unilateral large slip contact operator, Comp. Meth. In Appl. Mech. & Engng., 131, 263-286, (1996).
13. Crisfield, M.A., Re-visiting the contact patch test, Int. J. for Num. Meth. In Engng., 48, 435-449, (2000).
14. Bazant, Z.P., Instability, ductility and size-effect in strain softening concrete, ASCE, J. Eng. Mech., 102, 331-344, (1976).
15. Bazant, Z.P. & Cedolin, L., Stability of structures, Oxford University Press, 1991.
16. De Borst, R. & Muhlhaus, H.B., Gradient dependant plasticity: formulation and algorithmic aspects, Int. J. for Num. Meth. In Engng., 35, 521-539, 1992.
17. A. Zervos, P.Papanastasiou, & I. Vardoulakis, A finite element formulation for gradient plasticity, Int. J. for Num. Meth. in Engng., to be published.
18. Mi, Y., Crisfield, M.A., Davies, G.A.O. & Hellweg, H.B., Progressive delamination using interface elements, J. Composite Materials, 32(14), 1246-1272, (1998).
19. Alfano, G. & Crisfield, M.A., Finite element interface models for the delamination analysis of laminated composites: mechanical and computational issues, Int. J. Num. Meth. In Engng., to be published.
20. Needleman, A., An analysis of decohesion along an imperfect interface, Int. J. of Fracture, 42, 21-40, 1990.
21. Computational Mechanics, Special issue on meshfree particle methods, 25,2-3, 2000.
22. Peng, X & Crisfield, M.A., A consistent co-rotational formulation for shells using the constant stress/constant moment triangle.
23. Morley, L.S.D., The constant-moment plate bending element, J. Strain Analysis, 6(1), 20-24, (1971).
24. Zienkiewicz, O.C. & Zhu, J.Z., A simple error estimator and adaptive procedure for practical engineering analysis. Int. J. for Num. Meth. In Engng., 24,337-357, 1987.
25. Riccius, J. & Crisfield, M.A., Nonlinear adaptivity and shells using simple triangles, in: Computational Mechanics in UK, 6th ACME Conference, School of Engineering, University of Exeter, 159-162.

المنارة للاستشارات

ww

26. LUSAS User Manual, , FEA Ltd., Kingston upon Thames, Surrey, U.K., (1999)
27. Raju, I., Calculation of strain-energy release rates with higher-order and singular elements, Engineering Fracture Mechanics, 28, 251-274, (1987).
28. Wu, E.M. & Reuter Jr. C., Crack extension in fibreglass reinforced plastics, Report N 275, T & AM, University of Illinois, (1965)
29. Reeder, J.R., Evaluation of mixed-mode delamination failure criteria, Report NASA, TM 104210, (1992).

للاستشارات
ww

Simulation-based optimization in virtual thermo-mechanical prototyping of electronic packages

G.Q. Zhang
Center for Industrial Technology/Philips, P.O. Box 218, 5600 MD Eindhoven, The Netherlands, g.q.zhang@philips.com

H.P. Stehouwer
Centre for Quantitative Methods (CQM), 5600 AK Eindhoven, The Netherlands, stehouwer@cqm.nl

Abstract: *This paper presents some of our recent research and development results for virtual thermo-mechanical prototyping of electronic packages. Philips' simulation-based optimization strategy focusing on the development of reliable Response Surface Models for the underlying nonlinear responses is demonstrated. This strategy plays a vital role in the realization of virtual thermo-mechanical prototyping of electronic packages. The demonstrator presented in this paper shows that combining reliable nonlinear FEM-based simulation models with this optimization strategy is an effective and efficient method for virtual thermo-mechanical prototyping of electronic packages.*

Key words: virtual prototyping, electronic packaging, thermo-mechanical reliability, simulation-based optimization

1. INTRODUCTION

Due to the increasing demands for miniaturization, function integration and shorter time to market, thermo-mechanical reliability of electronic packages becomes one of the major concerns for electronic industry all over the world. Based on the root cause analyses from observed failures of electronic packages, it is found that most of the thermo-mechanical reliability problems originate from the product/process design phase. The traditional design approach based on trial-and-error (designing, building and testing of a multiplicity of physical prototypes) is no longer competitive, because it may not guarantee that the designed package satisfies performance specifications and reliability criteria, and at the same time it can be manufactured fast, economically and environmental friendly. Therefore, there is an urgent need

G. Q. Zhang et al. (eds.), Benefiting from Thermal and Mechanical Simulation in Micro-electronics, 151–164.
© 2000 *Kluwer Academic Publishers.*

المنارة للاستشارات

ww

to develop an innovative thermo-mechanical design and qualification method for electronic packages.

One of the most promising alternatives for traditional package design method is (simulation-based) *virtual prototyping,* which is beginning to draw more and more attention from both industries and academic world. Within Philips the following major building blocks for virtual thermo-mechanical prototyping of electronic packages are identified:

- developing and integrating multi-disciplinary input data and design spaces of packages, such as material properties, product geometry, process parameters, and environmental data, possible failure modes and ranking, damage models and criteria;
- building FEM-based parametric simulation models;
- validation of simulation models;
- carrying out simulation-based optimization (developing Response Surface Models (RSM) by integrating sequential Design of Experiments (DOE) with the reliable FEM simulation models; doing optimization)
- generating virtual thermo-mechanical prototyping rules for both design and qualification;
- implementation in engineering.

The results of virtual thermo-mechanical prototyping can be used to predict, evaluate and optimize the requirements prior to major design or manufacturing investment.

Among the building blocks for virtual thermo-mechanical prototyping of electronic packages, identified by Philips, "simulation-based optimization" is one of the key factors determining the success of virtual prototyping. It is worth to mention that using simulation alone is not virtual prototyping. It is only possible to carry out virtual prototyping if optimization methods, such as maximum/minimum, parameter sensitivity, robust design, are integrated with simulation.

This paper focuses on the description of methodology of simulation-based optimization for virtual thermo-mechanical prototyping of electronic packages. A demonstration case is given for optimizing package geometry with the objective of minimal die tension stress. MARC is used for nonlinear thermo-mechanical simulation, and integrated with COMPACT for design optimization. It can be concluded that by combining the strengths of reliable simulation models with advanced optimization techniques, virtual thermo-mechanical prototyping of electronic packages can be achieved with substantial contribution on the business profitability.

المنارة للاستشارات

ww

2. SIMULATION-BASED OPTIMIZATION

Simulation-based optimization is optimization based on and integrated with advanced simulation models that can predict the product/process behavior reliably and efficiently. Simulation-based optimization of packages involves finding settings for a number of package design parameters that are optimal with respect to several simulated response characteristics of the packages. These characteristics, called *response parameters*, may originate from different engineering disciplines. Since there are many possible design parameter settings and nonlinear FEM simulations are often time consuming, the crucial question becomes how to find the best possible parameter setting with a minimum number of simulations.

Many papers describe *direct optimization* procedures to obtain optimal parameter settings in such situations [6-9]. Disadvantage of these approaches is that little insight is obtained for the behavior of the responses in terms of the complete design space. Moreover, when the optimization problem changes (e.g., a stricter bound on a response parameter), the optimization procedure has to be restarted.

This paper presents a systematic non-direct optimization approach in which both optimal design parameters are found and insight is given by means of so-called *compact models* for each response parameter [1, 13]. This approach is fully implemented in the general-purpose design optimization tool COMPACT.

2.1 Compact model approach

The compact approach consists of four steps, see Figure 1. In the following, each of these steps will be described.

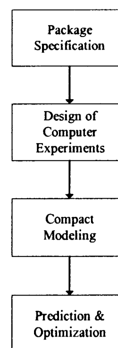


Figure 1. Compact model approach.

المنارة للاستشارات

2.2 Step 1: Package design specification

In the package design specification step, the design optimization problem is formulated as to find settings for some design parameters such that:

1. design parameters satisfy certain constraints,
2. response parameters satisfy certain constraints, and
3. a suitable objective function is defined.

Elements that need to be specified in Step 1 are design and response parameters and their ranges of variation, the constraints, and the objective function.

In the demonstration case, three geometric parameters of the package are specified as the design parameters, and the minimum tension stresses of the die the optimization objective.

2.3 Step 2: Design of computer experiments

The second step generates a set of suitably chosen package designs that satisfies all bounds on design parameters defined in Step 1. This is the set of feasible package designs to be simulated. Obviously, this set should be so chosen such that maximum amount of response information can be obtained while minimizing the number of experiments. *Design of Experiments (DOE)* techniques [10] can be used to obtain such a set. Examples are the full-factorial design or the Central Composite Design (CCD). However, those classical DOE methods mainly focus on physical experiments subjecting to noise. Those DOE schemes have the following drawbacks when used for computer simulations:

- In computer simulations, noise does not play a role, after the simulation models are validated. Hence, no extra information is gained from the repeated evaluations, which is often the case in classical DOE. In general, classical DOE needs relatively more experiments to be able to determine the influence of noise.
- Due to the noise inherit in physical experimentation it is often optimal to evaluation conditions that lie on the borders of the explicit feasible region. In computer experiments other parts of the design space are often equally interesting.
- Another drawback of most classical DOE methods is that they are only applicable for rectangular feasible regions.

For these reasons we do not suggest the use of classical DOE methods for compact modeling of computer simulation output. The suitable DOE methods for simulation-based optimization should be *space filling*, *non-collapsing and sequential.*

المنارة للاستشارات

ww

Space filling schemes – To obtain a compact model that predicts well for all feasible design parameter settings, i.e., for all feasible package designs, one has to choose the computer simulations such that as much information as possible is captured. Intuitively this is the case when the evaluation points are spread throughout the feasible region as evenly as possible, i.e., the DOE scheme is space filling.

Non-collapsing schemes – Initially it may be unknown which parameters are more important than others. A DOE scheme is called non-collapsing if, in case one or more of the design parameters appear to have non-significant influence, every computer simulation still gives information about the influence of the other design parameters on the response parameters. In this way none of the time consuming computer experiments may become useless.

Sequential – To minimize the number of computer simulation, the initial simulation set defined by the DOE should be as small as possible. Additional simulations may be carried out and the results should be able to be added to the known set of responses, if the initially developed RSM doesn't satisfy the specified criterion of accuracy. It is an iterative process.

COMPACT uses an approach for generating such schemes that searches for the most space filling evaluation scheme within the class of so-called Latin Hypercube Designs (LHD). It is an extension for the approach presented in [12].

2.4 Step 3: Compact modeling

The third step aims at obtaining satisfied response surface models described in terms of design parameters for each response parameter. These models are based on the outcomes of the simulated packages proposed in Step 2 and are used to predict the thermo-mechanical behavior of package without costly nonlinear FEM simulation runs. Depending on the underlying complexity of the response surface model, either low order polynomial regression models [10] or Kriging models [13] can be used.

2.4.1 Polynomial regression models

These are models in which the response can be expressed as a polynomial function of the design parameters. Least squares techniques are used to fit polynomial models for a given data set. To determine which terms are more important we apply model-pruning techniques. Since the number of

المنارة للاستشارات
ww

simulation runs has to as least as large as the number of terms in the model, higher order regression models are not always practical.

The polynomial regression models can be thought of as 'global' models in the sense that all of the evaluations are weighted equally in the fitting of the compact model. For the prediction *f(x)* of a compact model in a certain point *x*, nearby evaluation points contribute equally to *f(x)* as evaluation points that are further away. It may be argued that such a global model may not yield the best approximation in case the underlying function has for instance a high non-linear structure with multiple local minima. In such a case, 'local' modeling properties are more attractive, i.e., when *f(x)* is more strongly influenced by nearby evaluated points than those further away from *x*. Models with these characteristics are called *interpolation models*. Such models interpolate the evaluated response values at the evaluation points. Roughly speaking, interpolating models enable the simulated response data to 'choose' the interpolation model enabling the representation of more complexity than a least squares fitting with an imposed model.

2.4.2 Kriging models

One popular type of interpolating models is the *Kriging model*. Hereby it is assumed that the response is a realization of a Gaussian process. Although this assumption is not correct, since the underlying function is deterministic, it does provide a convenient framework from which to derive a class of functions that interpolate the responses. The Kriging model can be written in the form:

$$f(x) = \beta + Z(x),$$

where β is a constant and $Z(x)$ a random process with mean zero and covariance

$$V(x, y) = \sigma^2 \cdot R(x, y)$$

Here, σ^2 is the process variance and $R(x,y)$ is the correlation function defined by:

$$R(x, y) = \prod_{j=1,\ldots,D} e^{-\theta_j |x_j - y_j|^{p_j}} \quad ,$$

المنارة للاستشارات

ww

where θ_i denotes the correlation coefficient of dimension i. After calculating the Best Linear Unbiased Predictor (BLUP), it can be found that the Kriging model can be explicitly written as

$$f(x) = \beta + \sum_{d} c_d \cdot e^{\sum_i -\theta_i \left| x_i - x_{d,i} \right|^{p_i}}$$

where d runs over dimensions, i runs over experiments, $x_{d,i}$ denotes the ith dimension of experiment d and the constants c_d can be calculated from the BLUP expression.

2.4.3 Compact model validation and selection

Validation methods like for instance statistical significance testing loose their interpretation in the context of deterministic evaluations and may therefore be misleading. Perhaps the best way to assess model prediction capabilities is to evaluate he *Root Mean Squared Error* (RMSE) on an independent test set. However, taking into account the fact that the evaluations are typically very expensive, this is often not a serious option.

We have good experiences with the use of *(leave-one-out) cross-validation* to assess the prediction capabilities of a compact model. Given the evaluated data of n points this technique loops through all n points. In every cycle of the loop the current point is held apart and a model is constructed. The prediction capabilities of this model are then assessed on the point held out. Using the squared errors on the points held out the *cross-validation RMSE* is calculated given by

$$\text{cv - } RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(y_i - \tilde{y}_i \middle| \{D - i\} \right)^2} \, ,$$

where $y_i|\{D-i\}$ denotes the prediction of the ith simulation by the response surface model obtained by fitting all experiment data but leaving out the data of experiment i. Cross-validation is computationally rather demanding. However, compared with the time of a typical evaluation it is negligible.

It can be often seen from the literature that the Root Mean Squared Error (*RMSE*) is used to assess model prediction capability, which is given by

المنارة للاستشارات
ww

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \tilde{y}_i)^2} .$$

However, this statistic does not account for the number of degrees of freedom and therefore can be made arbitrary small by adding model complexity.

2.5 Step 4: Prediction and optimization

Steps 1-3 result in compact response surface model(s) representing all the specified thermo-mechanical responses of the package with respect to the specified design space. In Step 4 these compact models can be used for prediction, optimization, what-if, and sensitivity analyses. The resulting optimization problem can be formulated as a Non-Linear Programming (NLP) problem that can be solved using powerful mathematical optimization techniques like the Generalized Reduced Gradient (GRG) approach. This enables the optimization of complex design problems with thousands of response parameters.

3. DEMONSTRATOR

Due to the mismatch between the thermo-mechanical properties of different components in electronic packages and both the external and internal constraints, thermal stresses occur. Very often, these thermal stresses cause various types of thermo-mechanical failures during processing, testing, and use.

As the demonstrator, we consider an electronic package that consists of die, solder and heatsink. The vertical die crack, occurs in the cooling down phase of the die-bonding process, is taken as the critical failure mode. The selected electronic package is used to demonstrate the major procedures and principles of simulation-based optimisation in virtual thermo-mechanical prototyping.

3.1 FEM models

A parametric 2D FEM model with axisymmetric elements was developed using MARC. In this model, the solder material is modeled as temperature dependent visco-plastic, the heatsink as temperature dependent ideally plastic, and the die as temperature independent elastic. The maximum

المنارة للاستشارات
ww

tension stresses at the middles of the bottom and top of the die are used as the crack index. If the predicted maximum stress level is close to or higher than the allowable tension stress of the die, vertical die crack is assumed to occur. Experimental verification shows that this model can reliably predict the thermo-mechanical behavior of the package during cooling down from soldering to room temperature.

3.2 Simulation-based optimization of demonstrator

3.2.1 Package design specification

Three geometric parameters are chosen as the design parameters:
1. thickness of the die (Tdie),
2. thickness of the heatsink (Tsink), and
3. length of the die(Ldie).

All the other parameters (process, material and geometric parameters) are assumed to be constant. Table 1 shows the three design parameters and their ranges of variations.

design parameter	lower bound	upper bound
Tdie	0.1	0.6
Tsink	0.2	1.5
Ldie	1.5	3

Table 1. Design parameters and their ranges of variations

Using the verified nonlinear FEM model, the maximum tension stresses at the middle of both the bottom (StressBot) and the top (StressTop) of the die are calculated.

The design optimization problem is to choose the geometry parameters of the package such that the maximum of StressBot and StressTop is minimal, within the specified ranges of variations of design parameters. The objective function is defined as MaxStress = max{StressBot, StressTop}.

3.2.2 Design of computer experiments

A space-filling Latin-Hypercube-Design consisting of 20 design variations is first constructed by using COMPACT. FEM simulations are carried out for all the 20 designs, and the StressTop and StressBot are used as the response parameters.

المنارة للاستشارات

ww

Nr.	design parameters			response parameters		objective
	Tdie	Tsink	Ldie	StressTop	StressBot	MaxStress
1	0.389	1.295	2.053	375	380	380
2	0.179	0.337	1.974	551	538	551
3	0.363	0.405	2.368	444	402	444
4	0.205	0.200	2.605	558	539	558
5	0.100	0.747	2.526	735	730	735
6	0.232	0.816	2.211	525	520	525
7	0.337	1.432	2.842	448	456	456
8	0.416	0.953	2.447	411	402	411
9	0.258	1.500	1.579	406	413	413
10	0.600	0.884	2.132	327	306	327
11	0.153	1.226	2.684	642	640	642
12	0.284	0.542	2.921	537	514	537
13	0.547	1.158	3.000	363	364	364
14	0.126	1.089	1.816	604	603	604
15	0.495	1.021	1.500	299	269	299
16	0.311	0.611	1.658	399	387	399
17	0.442	0.268	1.737	361	283	361
18	0.468	0.679	1.895	353	318	353
19	0.574	1.363	2.289	312	318	318
20	0.521	0.474	2.763	402	336	402

Table 2. Latin hypercube scheme of 20 package designs.

3.2.3 Compact modeling

For both of StressTop and StressBot, quadratic models with interactions are used for RSM generation. Using COMPACT's automatic pruning procedure based on cross-validation, the unimportant model terms were deleted; see Figure 2. The regression statistics are listed in Table 3 indicating that the reliability and accuracy requirements are satisfied (error < 5%).

المنارة للاستشارات

ww

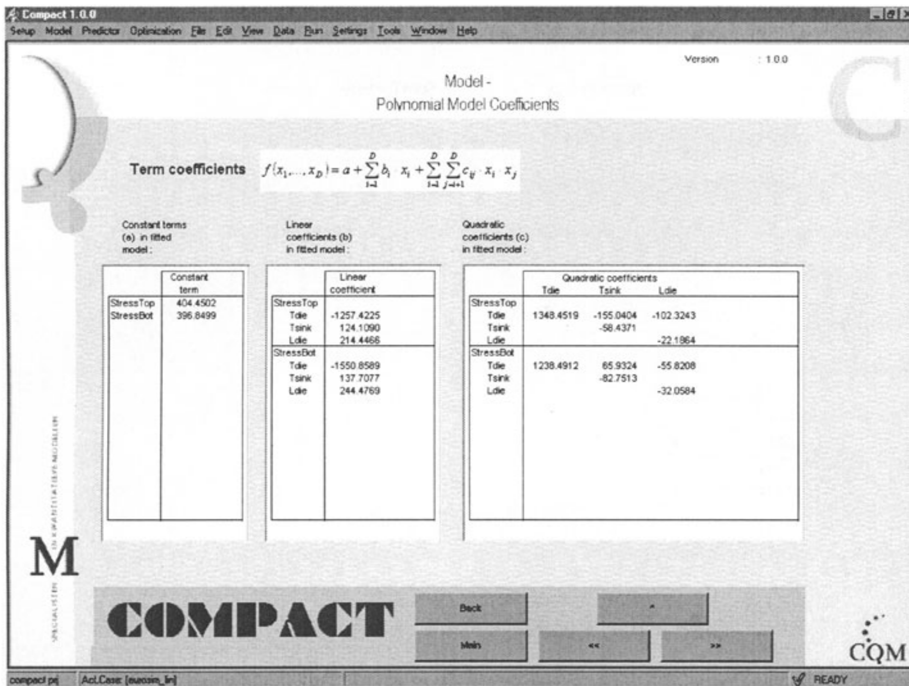


Figure 2. Pruned compact models StressBot and StressBot.

	RMSE	cv-RMSE
StressTop	8.50	16.03
StressBot	5.51	10.62

Table 3. Compact model validation.

Since these models are accurate enough, there is no need to do extra simulation runs or apply more flexible model types like Kriging models. We did apply Kriging, without significant improvement on the resulting cv-RMSE values (14.82 and 11.07). To really benefit from Kriging in this case, more simulation runs are required. As a rule of thumb it is suggested to take 10-15 times the number of design parameters for Kriging models. In Figure 3 the simulated StressBot versus the predicted ones is plotted, showing the accuracy of the stress prediction using the developed RSM model.

المنارة للاستشارات

ww

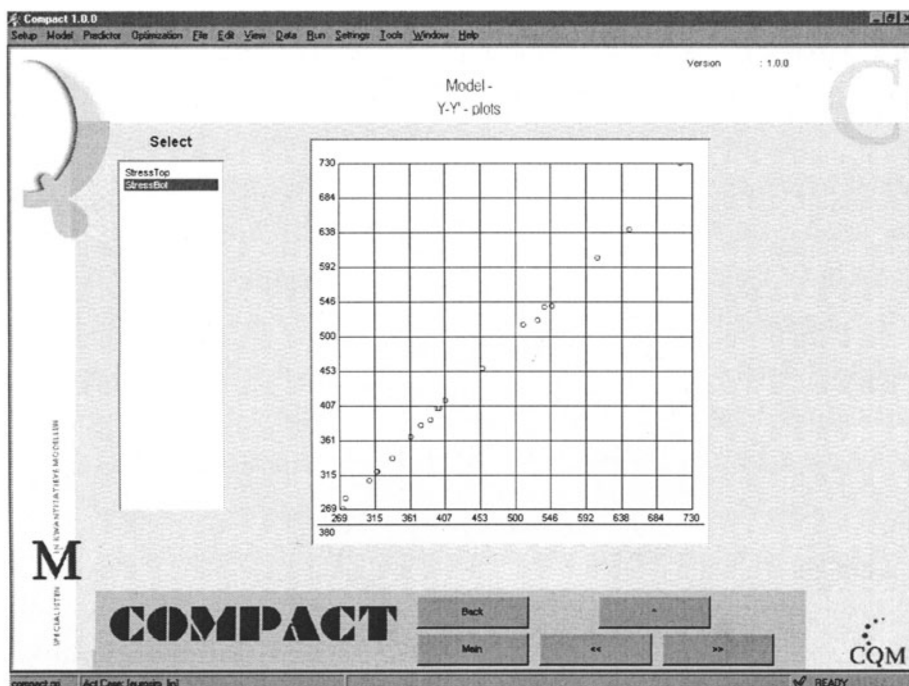


Figure 3. Simulated versus predicted stresses

3.2.4 Step 4: Prediction and optimization

The optimal design obtained by minimizing the MaxStress target is showed in Table 4. Tension stress reduction of more than 20% can be achieved, compared with the best-simulated package design obtained from Step 2. Figure 4 shows the 3D plot of StressTop as function of die and heatsink thickness.

design parameters			response parameters		objective
Tdie	**Tsink**	**Ldie**	**StressTop**	**StressBot**	**MaxStress**
0.6	1.5	1.5	230.24	236.25	236.25

Table 4. Optimal package design.

المنارة للاستشارات

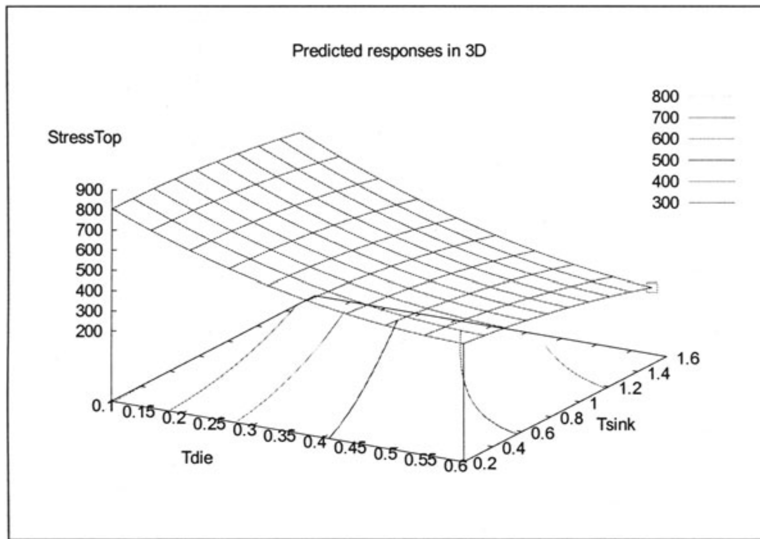


Figure 4. 3D plot compact model StressTop around optimum package design.

4. CONCLUSION

This paper presents part of our research and development results for virtual thermo-mechanical prototyping of electronic packages. The simulation-based optimization strategy focusing on the development of reliable RSM is demonstrated via a real package case. Using the suggested methodology, virtual thermo-mechanical prototyping of electronic packages can be achieved. This will make the first time right (die crack free design), shorter-time to market and optimized package design possible.

5. REFERENCES

[1] G.Q. Zhang, "The state-of-the-art of simulation-based optimization", Philips internal report, 1998.
[2] G.Q. Zhang, J. Janssen, L.J. Ernst, J. Bisschop, Z.N. Liang, F. Kuper and R. Schravendeel, "Virtual thermo-mechanical prototyping of electronic packaging using Philips' optimization strategy", IMAPS2000, USA, 2000

المنارة للاستشارات
ww

[3] V. Bouwman, G.Q. Zhang, J.W. ter Weeme, "Stochastic buckling strength predictions of orthotropic shallow shells" Proceeding of 41st AIAA/ASME/ASCE/AHE/ASC conference, USA, 2000

[4] V. Bouwman, G.Q. Zhang, J.W. ter Weeme, "Virtual stability prototyping of doubly curved orthotropic panels" Proceeding of 41st AIAA/ASME/ASCE/AHE/ASC conference, USA, 2000

[5] Dick den Hertog, Peter Stehouwer, Vincent Bouwman, Jan Willem ter Weeme, G.Q. Zhang, "Integral product optimization (in Dutch: Integrale produkt optimalisatie)", De Constructeur, February 1999

[6] Conn A.R., Toint, Ph.L. (1996), "An algorithm using quadratic interpolation for unconstrained derivative free optimization", in: Nonlinear Optimization and Applications, G. di Pillo and F. Giannes (eds.), Plenum Publishing, pp. 27-47.

[7] Powell, M.J.D. (1996), "A direct search optimization method that models the objective and constraint functions by linear interpolation", Presentation at the SIAM conference 1996, Virginia.

[8] Toropov, V.V. (1992), "Multipoint approximation method in optimization problems with expensive function values", in: Computational System Analysis 1992, Sydow, A. (ed.), Elsevier, pp. 207-212.

[9] Toropov, V.V., Filatov, A.A., and Polynkine, A.A. (1993), "Multiparameter structural optimization using FEM and multipoint explicit approximations", Structural Optimization, Vol. 6, pp. 7-14.

[10] Montgomery D.C., (1984), "Design and Analysis of Experiments (Second Edition)", John Wiley & Sons, New York.

[11] Morris, M.D. and Mitchell, T.J. (1995), "Exploratory designs for computer experiments", Journal of Statistical Planning and Inference, Vol. 43, pp. 381-402.

[12] Sachs, J., Welch, W.J., Mitchel, T.J., and Wynn, H.P. (1989), "Design and analysis of computer experiments", Statistical Science, Vol. 4, pp. 409-435.

[13] Stehouwer, H.P. and D. den Hertog (1999), Simulation-based design optimisation: methodology and applications (extended abstract), *"Proceedings of the First ASMO UK / ISSMO Conference on Engineering Design Optimization"*, Ilkley, UK.

المنارة للاستشارات

www

Thermal Fatigue Reliability Optimisation of Flip Chip Assemblies

Bart Vandevelde, Eric Beyne
IMEC
Kapeldreef 75, B-3001 Leuven, Belgium
Bart.Vandevelde@imec.be
Tel.: +32 16 281 513

Abstract

The thermal cycling life of flip chip assemblies is often limited by solder joint fatigue. The reliability can be increased by a factor 10 using an underfill material. However, the solder joint reliability is still very dependent on the choice of the underfill material. Using thermo-mechanical simulations, based on non linear finite element simulations, the induced inelastic strains in the solder joints are calculated, and give a value for the expected thermal fatigue life. Combing these FE simulations with an optimisation tool, the optimal underfill material is proposed. The optimisation is based on a parameter sensitivity analysis using DOE techniques. The optimal properties for the underfill are a high elastic modulus and a CTE closely matched to the CTE of the solder joint.

Thermal fatigue issues in flip chip assemblies

Flip chip technology is defined as mounting the chip to a substrate with any kind of interconnect materials and methods (e.g. fluxless solder bumps, tape-automated bonding, wire interconnects, conductive polymers, anisotropic conductive adhesives, metallurgy bumps, …), as long as the chip surface (active area) is facing to the substrate. The flip chip technology has many well-known benefits:

- *superior electrical properties*: the shortest possible leads result in lowest inductance and minimal propagation delay.
- *high number of I/Os*: with flip chip, an area array configuration of the solder joints is possible (wire bonding provides only connections at the perimeter).
- *the highest miniaturisation*: the package has almost the size of the chip itself.

However, the flip chip technology using eutectic solder as connection means, has one major drawback:

- *poor thermo-mechanical behaviour of the solder joints*. The chip with a low CTE is mounted to a substrate (e.g. FR4) with a much higher CTE, and if this assembly is subjected to temperature cycling, the deformation mismatch between the chip and the substrate causes high mechanical loads (shear force and moment) on the solder joints. A solution for this is to fill the gap between the chip and the substrate with an underfill material. This material is based on epoxy with a filler material to increase the elastic modulus, and also to decrease the CTE.

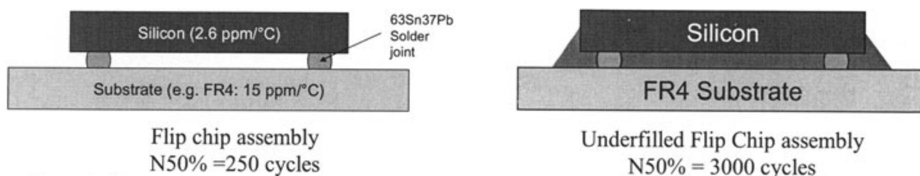


Figure 1: Schematic drawing of a flip chip assembly. Due to the high CTE mismatch, high mechanical loads affect the solder joints resulting in a limited life time. Using an underfill, the solder joint reliability increases with a factor 10.

G. Q. Zhang et al. (eds.), Benefiting from Thermal and Mechanical Simulation in Micro-electronics, 165–177.
© 2000 *Kluwer Academic Publishers.*

للاستشارات
ww

Temperature cycling may cause different types of failures in flip chip assemblies. The most well known failure is thermal fatigue of solder due to repeated inelastic (plastic and creep) deformation. Another failure mechanism occurring in flip chipped assemblies in underfill delamination. Often, adding an underfill between the chip and the substrate relaxes the loads on the solder joints, but replaces the failure problems to underfill delamination, or even die cracking due to the underfill. However, in this paper, only thermal fatigue of the solder joints is considered.

The loads on the solder joints cause repeated strong deformation of the solder joints. The deformations of the solder joint exceed the elastic region and causes plastic deformation, but also creep deformation at high temperature (above 100°C, the solder joints are very close to their melting point of 183°C). After a number of temperature cycles, a thermal fatigue induced crack initiates and propagates. Finally, a complete rupture of the solder joint occurs, resulting in an electrical disconnection. This is called as a 'thermal fatigue failure'. This solder joint fatigue failure mode may be classified in the group of 'low cycle fatigue', with a number of cycles to failure below 10000 cycles. The reason for the relatively low number of cycles before failure is due to the inelastic deformation in each cycle.

Electronic components are subjected to temperature changes by environmental cycles (day/night, heat under the hood of an automobile, …) and power cycles (on/off cycles of a computer). However, an experimental thermal cycling test in real life conditions would take to much time. Therefore, accelerated thermal cycling is applied to test samples, by increasing the temperature range and the frequency of cycling (Figure 2 and 3).

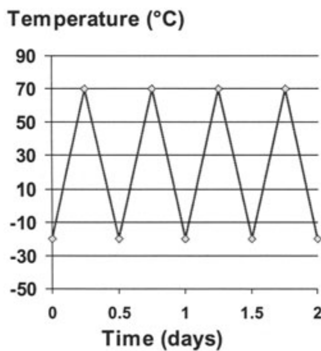


Figure 2: Normal temperature cycling of an electronic system

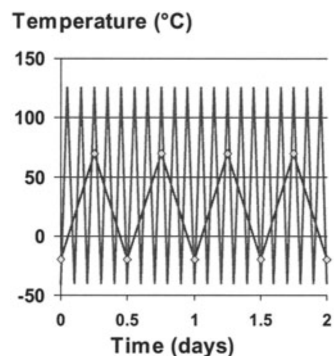


Figure 3: Accelerated temperature cycling to get earlier failures. The acceleration is done by increasing the temperature range and frequency of cycling

It is difficult to relate accelerated test results to real life conditions, because it may be possible that other failure mechanisms may be introduced or promoted due to the accelerated conditions (e.g. more creep effect due to higher temperatures). There is a model available in the literature:

$$Acceleration\ factor = \frac{N_{field}}{N_{test}} = \left[\frac{\Delta T_{test}}{\Delta T_{field}}\right]^{1.9} \cdot \left[\frac{f_{field}}{f_{test}}\right]^{1/3} \cdot e^{\left[1414\left(\frac{1}{T_{\max}^{field}} - \frac{1}{T_{\max}^{test}}\right)\right]}$$

المنارة للاستشارات

ww

with

$\Delta T_{test} / \Delta T_{field}$ = *temperature range at test/field condition.*
f_{test} / f_{field} = *cyclic frequencies at test/field condition.*
$T_{max}^{test} / T_{max}^{field}$ = *highest on-off temperature at test/field condition* ***in Kelvin****.*

A failure distribution curve is made from an acceleration thermal cycling test. From this curve, one may select Nxx% (number of temperature cycles to xx % failures). Most used values are N50% and N100ppm.

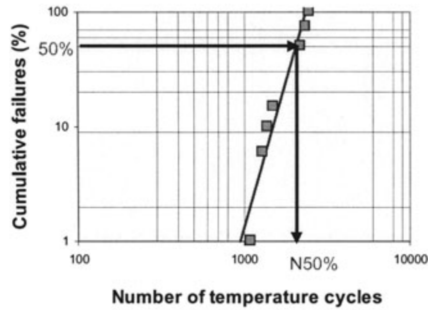


Figure 4: Lognormal curve for failure distribution.

المنارة للاستشارات
ww

Simulation of thermal fatigue reliability for flip chip assemblies

For non-underfilled flip chip assemblies, the main deformation mode of the solder joint is shear, due to the difference in horizontal displacement between the chip and the substrate. This includes that the induced plastic strain is linear dependent on the DNP (distance to neutral point) and on the CTE mismatch between the chip and the substrate, and inversely proportional to the stand-off height.

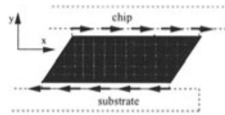


Figure 5: Main deformation mode in non-underfilled flip chip assemblies

For underfilled flip chip assemblies, it is much more complicated. Generally, three different inelastic strain modes are induced in the solder connections:

1. A shear strain (ε_{xy}^{inel}), caused by the horizontal displacement mismatch between the chip and the substrate.	
2. A horizontal normal strain (ε_{x}^{inel}), caused by the horizontal normal strains internal in the underfill material.	
3. A vertical normal strain (ε_{y}^{inel}), due to the vertical deformation mismatch with the underfill (thermal expansion and out-of-plane contraction).	

Figure 6: Deformation modes in solder joints of underfilled flip chip assemblies.

It is possible to do a rough analytical calculation of these three induced plastic strain modes. The underfilled flip chip assembly behaves like a tri-material structure. A uniform horizontal normal strain and a shear strain increasing toward a maximum approaching the die edge have been found. The existence of the relatively soft solder joints has almost no influence on the strain distribution in the underfill layer. In addition, these underfill strains will cause the plastic strains in the solder joints. These plastic strains can be calculated as follows:

1. The normal plastic strain ε_x^{pl} is caused by the horizontal normal strain $\varepsilon_x^{el}{}_{(underfill)} = \sigma_x^{el}{}_{(underfill)}/E_{underfill}$. This underfill strain is subjected to the solder joint and induces a deformation which cannot only be covered by an elastic strain due to the low yield stress of the solder material:

المنارة للاستشارات

ww

$$\varepsilon_{x\,(underfill)}^{el} = \varepsilon_{x\,(solder)}^{el} + \varepsilon_{x\,(solder)}^{pl}$$

2. The shear plastic deformation ε_{xy}^{pl} can be calculated in the same way:

$$\varepsilon_{xy\,(underfill)}^{el} = \varepsilon_{xy\,(solder)}^{el} + \varepsilon_{xy\,(solder)}^{pl}$$

with $\varepsilon_{xy\,(underfill)}^{el}$ the elastic shear strain in the underfill near the solder joint.

3. The vertical normal plastic strain in the solder joint (ε_y^{pl}) is caused by the vertical deformation mismatch with the underfill. The vertical deformation of the underfill is determined by its thermal expansion or shrinkage and by the out-of-plane contraction caused by the horizontal normal strains:

$$\varepsilon_{y\,(underfill)}^{el} = \Delta T.CTE_{underfill} - \varepsilon_{x\,(underfill)}^{el}\,\nu_{underfill}$$

This underfill deformation is applied to the relatively soft solder joint:

$$\varepsilon_{y\,(underfill)}^{el} = \Delta T.CTE_{(solder)} + \varepsilon_{y\,(solder)}^{el} + \varepsilon_{y\,(solder)}^{pl}$$

The maximum elastic strains in the solder joints (ε_x^{el}, ε_y^{el}, ε_{xy}^{el}) are determined by the yield stress ($\varepsilon_{Von\ Mises} \leq \varepsilon_{yield}$) but they are much lower than the plastic strains.
Table I depicts that these analytical calculations are a good estimation of the three plastic deformations. These results are compared to non linear finite element simulations. Three different underfill materials are considered:

Table I:
Maximum Value of Plastic Strains in the Solder Joints after the Assembly Process (ΔT = -160°C).
The Analytical Estimations are compared with FEM Simulations

Underfill material	Finite Element Simulations			Analytical estimations		
properties	ε_x^{pl} (%)	ε_y^{pl} (%)	ε_{xy}^{pl} (%)	ε_x^{pl} (%)	ε_y^{pl} (%)	ε_{xy}^{pl} (%)
E = 10 GPa CTE = 25 ppm/K	0.18	-0.09	0.32	0.21	-0.05	0.26
E = 5 GPa CTE = 50 ppm/K	0.36	-0.41	0.95	0.70	-0.51	0.91
E = 3.3 GPa CTE = 75 ppm/K	0.50	-0.79	1.70	1.11	-0.95	1.56

Estimating the inelastic strains by semi-analytical is too rough, but it gives a good indication and understanding what really happens . For more detailed analysis, we need the non-linear finite element technique. The non-linearity is related to the visco-plastic behaviour of solder, and also to the temperature dependent properties of several materials. Ideally, a 3D mesh for the full model should be made, but from efficiency point of view, 2D plain strain mesh (which is a cross-section of the 3D model) gives much faster results. The 2D plain strain is often a good approximation of the 3D model, but the consistency with the 3D model has always to be checked. Consistency means that the relative comparison of results

المنارة للاستشارات
ww

should remain the same for both techniques for different package configuration. In other words, if for example the 3D model depicts that an increase in chip thickness should result in an increase of solder joint strains, the same effect should be found for the 2D modelling analysis. Moreover, 2D plain strain analysis may be a large underestimation of strains in the corner joints. For flip chip assemblies, 2D plain strain analysis is an acceptable approximation.

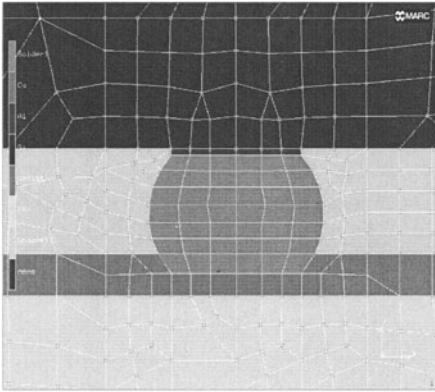


Figure 7: 2D plain strain FEM

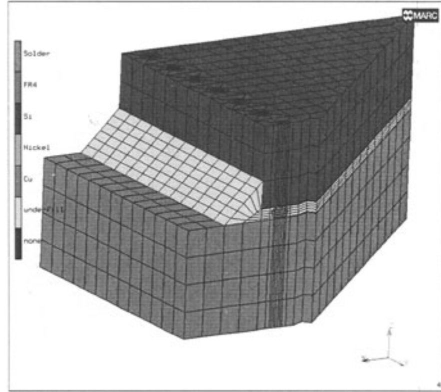


Figure 8: Full 3D FEM

The load applied to the finite element model is the accelerated temperature cycling (e.g. –55°C to +125°C, cycle time = 30 minutes). This temperature cycling causes inelastic strains in the solder joints. Figure 9 shows the shear stress as a function of the shear strain. This curve follows a hysteresis loop. The area in this loop is the induced inelastic strain energy per cycle, the width of the loop is the inelastic strain per cycle. Both values may be used as a damage parameter to indicate the expected thermal cycles to failure.
By comparing the induced inelastic strain with experimental results, an empirical model can be extracted which can be used in the future to estimate the thermal fatigue life. Important to know is that this empirical model is very dependent on the finite element model: the chosen material properties for solder, the simulation technique (2D, 3D) and area in the solder joints that is used to average the damage parameter.

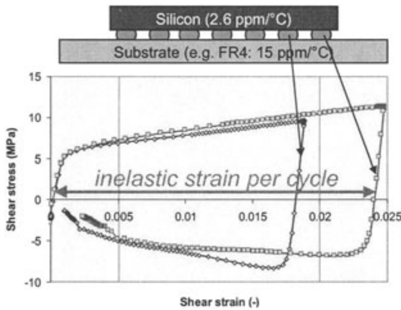


Figure 9: The stress strain curve follows a hysteresis loop during temperature cycling

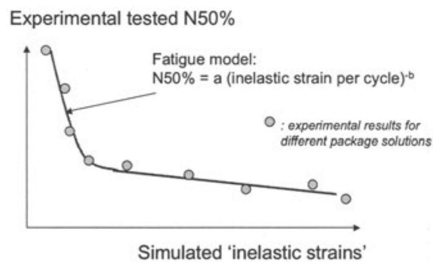


Figure 10: By comparing simulation with experimental results, an empirical model can be extracted

المنارة للاستشارات
ww

Parameter sensitivity analysis for thermo-mechanical simulation

The principal of parameter sensitivity analysis for thermo-mechanical simulation of flip chip assemblies is shown below. First, we need to define the material and geometry parameters to be analysed. Because the solder joint geometry after reflow is dependent on the chosen pad sizes and solder volume, we need to do first a simulation to estimate the solder profile.

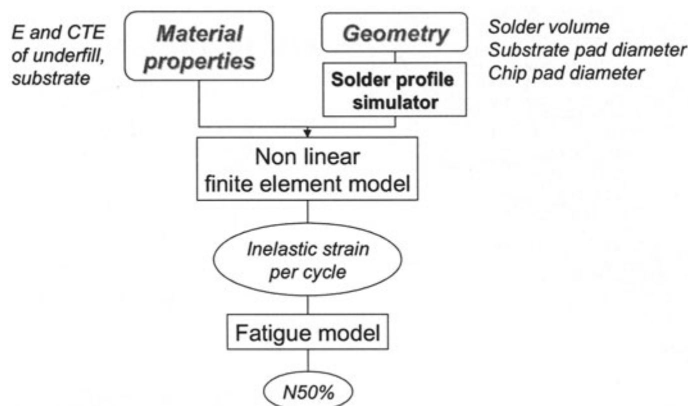


Figure 11: Schematic drawing how to do a full thermo-mechanical simulation analysis for flip chip assemblies, starting from defining the parameters, and ending with a thermal fatigue life.

The solder profile after reflow is mainly determined by minimisation of surface and gravitation energy (mainly by chip weight). Both analytical models *(Analytical derivation of the self-alignment motion of flip chip soldered components", N. van Veen; ASME Journal of Electronic Packaging, Volume 121, June 1999, pages 116-121*) as well as the simulation software Surface Evolver may be used for this:

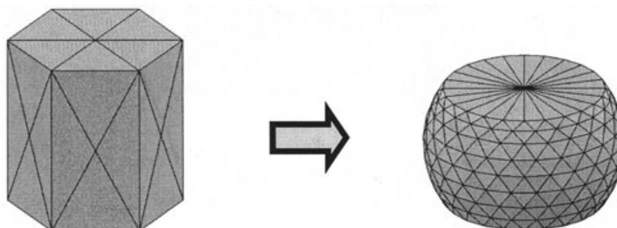


Figure 12: Using surface evolver to estimate the solder shape after reflow

To automate the parameter sensitivity analysis, the LMS OPTIMUS software is used to control the finite element simulations. The software changes for each simulations the parameters in the ASCII file (to be read by the finite element pre-processor), and extracts from the output file (generated by the FEM post-

المنارة للاستشارات

ww

processor) the result parameter. This result parameter is the maximum inelastic strain per cycle, transformed into N50% using the empirical model.

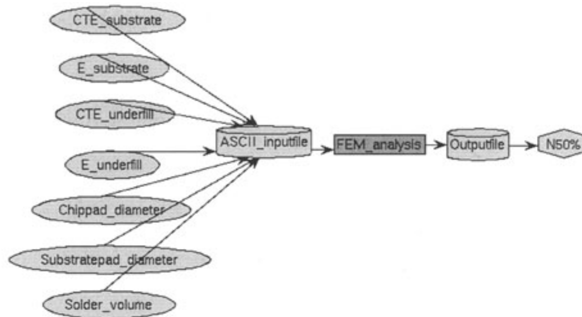


Figure 13: Parameterised finite element modelling with LMS Optimus.

The software also determines how to vary the parameters in the best way. This is based on Design Of Experiments (DOE). Instead of changing one parameter at a time, all parameters are varied at a time. Different DOE are possible, as shown at figure 14.

Type of DOE analysis:
- 2/3 level full factorial
- CCF
- Taguchi
- Latin Hypercube
- random
- ...

Type: Central Comp. Faced — Options ... — 43 experiments

Design Inputs

	On	Name	Nominal	Low	High
1	■	Solder_volume	1e-12	0.5e-12	1.2e-12
2	■	Substratepad_diameter	0.080	0.080	0.120
3	■	Chippad_diameter	0.080	0.080	0.120
4	■	E_underfill	10000	2500	10000
5	■	CTE_underfill	25e-6	20e-6	50e-6
6	□	E_substrate	19000		
7	□	CTE_substrate	15e-6		

Design Outputs

	On	Name	Evaluation
1	■	N50%	Analysis

Figure 14: Setting the parameter ranges + choosing the type of DOE

Using directly the output results from a DOE analysis is rather difficult because all parameters are varied at a time. Therefore, a Response Surface Model (RSM) is made which fits the results in quadratic model:

N50% = a0 + a1 * Parameter_1 + a2 * Parameter_2 + ...
+ a12 * Parameter_1 * Parameter_2 + ...
+ a11 * (Parameter_1)2 + …

The coefficients $\mathbf{a_i}$ are the sensitivities to N50% for each parameter **i**, the coefficients $\mathbf{a_{ij}}$ are the interaction effects between parameter **i** and **j**. This RSM can be also used as an estimator of N50% as a function of all parameters, but also to optimise the parameter to have the highest N50%.

المنارة للاستشارات
ww

Parameter sensitivity analysis for non-underfilled flip chip assemblies

As discussed earlier, the main deformation mode is **shear**. Consequently, the thermal fatigue life is dependent on chip size, the solder joint stand-off and the substrate.

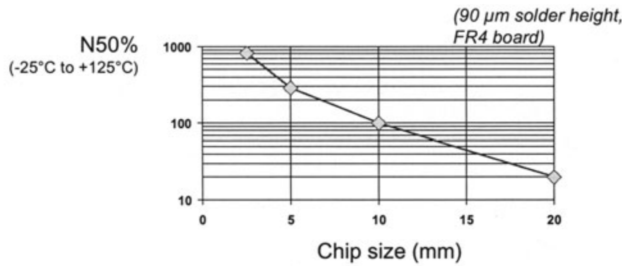


Figure 15: Effect of chip size on thermal fatigue reliability

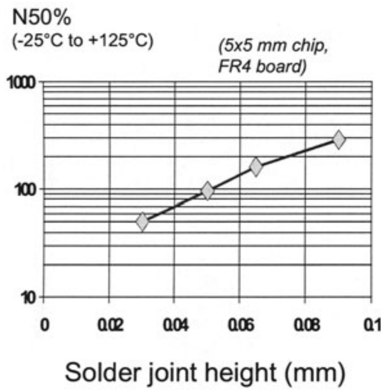


Figure 16: Effect of solder joint height on thermal fatigue reliability

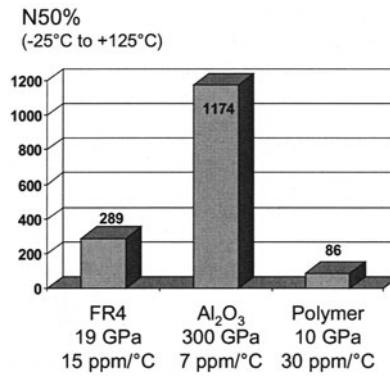


Figure 17: Effect of substrate material on thermal fatigue reliability

المنارة للاستشارات

ww

DOE analysis for underfilled flip chip analysis

To show the usefulness of DOE, a parameter sensitivity analysis has been done for underfilled flip chip assemblies. Five parameter have been defined:

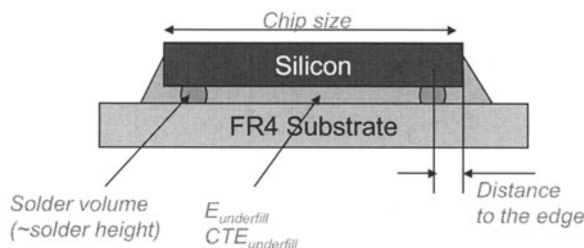


Figure 18: Defining five parameters for the DOE analysis

A 2D non-linear plain strain model is used to estimate the inelastic strains due to –25 to +125 °C thermal cycling. For the DOE, a CCF design has been chosen, and this requires 43 simulations for five input parameters. After simulation these 43 different cases (it takes about 2 days), a RSM is constructed. Figure 19 shows a good accuracy of the RSM by comparing the simulation values with the values determined by the RSM:

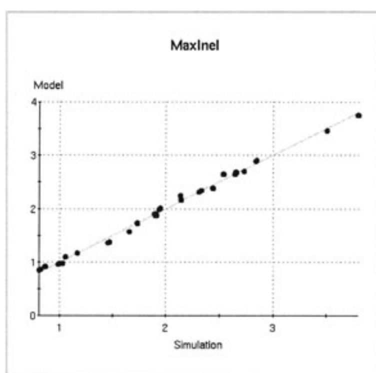


Figure 19: Accuracy of the RSM by comparing simulation values with RSM estimated values

First of all, we can look to the sensitivities of all parameters (these coefficients are for normalised ranges of the parameters). As shown at figure 20, the main parameter affecting the inelastic strain is the CTE of the underfill. Also the direction is given: the higher the CTE, the higher the strains. A stiffer elastic modulus of the underfill decreases the induced strains. The plot also shows that the strains are lower if the distance between the edge and the solder joint becomes larger. The chip size seems to be not affecting the reliability. The response surface model may also be used to optimise the parameters. With respect to the ranges and for a chip size of 10 mm, the optimal values for the parameters are $CTE_{underfill}$ = 25 ppm/°C, $E_{underfill}$ = 10GPa, V_{solder} = 1.1 e-12 m^3 (= max.) and Dist_edge = 0.5 mm (= max).

المنارة للاستشارات
ww

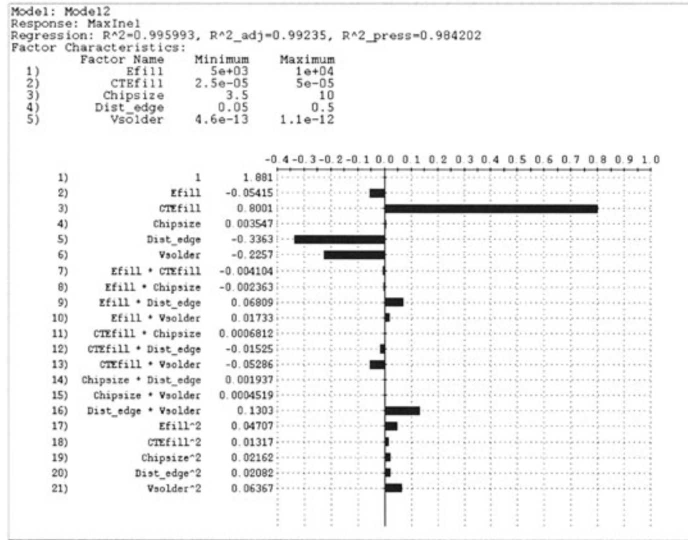


Figure 20: Sensitivity and interaction coefficients for all parameters

The optimal underfill material properties

The choice of a well matching underfill material is an important issue to achieve a good reliability for the flip chip assembly. To go more in detail in the effect underfill properties, a full factorial (6x4) parameter sensitivity analysis has been done (figure 21). The thermal fatigue life time is compared to a reference flip chip structure with a specific underfill material (E = 5 GPa; CTE = 50 ppm/K).

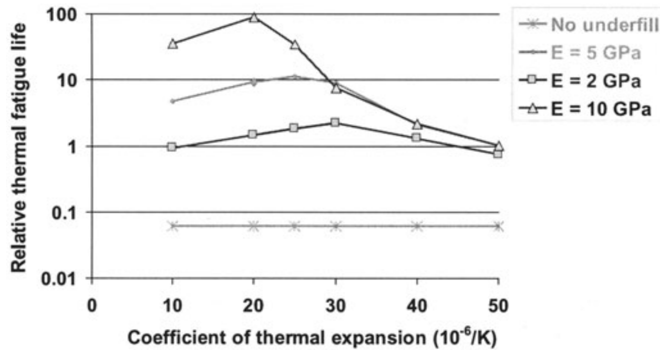


Figure 21: The relative thermal fatigue life time of the flip chip assembly as a function of the Coefficient of thermal expansion of the Underfill. The reference is a flip chip assembly using an underfill with properties E =5 GPa and CTE = 50 ppm/K.

المنارة للاستشارات
ww

As shows clearly, both material parameters have a high influence on the thermal fatigue reliability of the flip chip assembly (in figure 20, the sensitivity of E was not so high, but the range was much smaller, and in this small range, the effect of E is less, as shown in figure 21):

- The higher the stiffness of the underfill, the higher the thermal fatigue life. The reason for this is that the shear strains in the underfill become lower if the underfill has a higher elasticity modulus.
- Additionally, there is an optimal thermal expansion coefficient for the underfill, which depends on its elasticity modulus. For rigid underfills with E=10 GPa, the reliability is optimal for a CTE = 20 ppm/K, while for soft underfills (E = 2 GPa), the optimal CTE is 30 ppm/K. This significant dependence of the CTE on the reliability can be explained as follows. The horizontal normal and shear plastic strains in the solder joints increase with higher CTE. The vertical normal plastic strains become zero for a CTE of the underfill where the vertical deformation of the underfill is the same as the one of the solder joint. And for this vertical deformation, not only the thermal deformation of the underfill but also the out-of-plane Poisson contraction effect is important.

Previous figure has shown a full parameter sensitivity analysis for the underfill material properties. However, not all combinations of Young's modulus and CTE are possible to manufacture. Following figure depicts that there is a correlation between the Young's modulus and the CTE of underfill materials available on the market. This correlation means that the product of the Young's modulus with the CTE is constant value (figure 22).

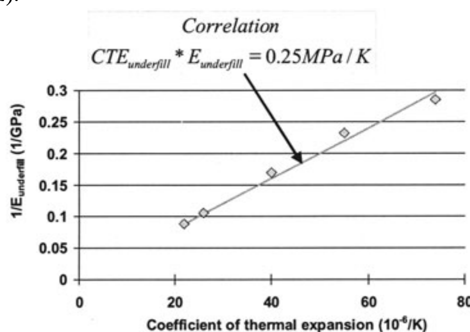


Figure 22: Correlation between Young's modulus and CTE of different underfill materials available on the market

Figure 23 shows the thermo-mechanical simulation results for flip chip assemblies for different underfill materials following the correlation curve of figure 22 (thermal cycling from –55°C to 125°C, chip size is 10 mm) . From these results, we can conclude that by introducing a relatively rigid underfill material, the thermal fatigue reliability can be increased. However, at the same time, restraining the movement of the chip relative to the substrate can produce significant high stresses in the underfill. Hence, the probability of a new failure mode, underfill delamination and rupture, increases. Delamination of the underfill may not only reduce the beneficial effect on the solder strain, but may also allow moisture to accumulate at the interface and lead to additional failure modes.

المنارة للاستشارات
ww

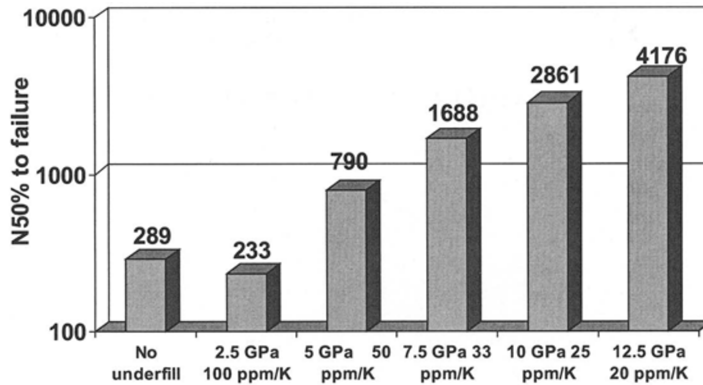


Figure 23: Thermal fatigue reliability results for underfill materials that are related to this 'E-CTE' correlation (figure 22).

Conclusions

Thermo-mechanical modelling based on non linear FEM, calculate the induced inelastic strains in the solder joints of flip chip assemblies. Using an empirical model, these inelastic strains are transformed into an estimation of the expected thermal fatigue life. Using optimisation tools based on DOE (design of experiments) and RSM (response surface modelling), multi-parameter sensitivity analysis and design optimisation can be done in an efficient way.
Because non-underfilled flip chip assemblies have a too low solder joint reliability, filling the gap between the chip and the substrate is required and can increase the reliability with a factor 10. Using thermo-mechanical simulation and optimisation tools, the optimal underfill properties are defined. With respect to solder joint fatigue resistance, the underfill should have a high elastic modulus and a CTE matching the CTE of solder.

المنارة للاستشارات

Product and Process Optimization with Simulation

Dick den Hertog and Peter Stehouwer
Tilburg University, Department of Econometrics, P.O. Box 90153, 5000 LE Tilburg, The Netherlands and Centre for Quantitative Methods, P.O. Box 414, 5600 AK Eindhoven, The Netherlands

Key words: optimization, simulation, product design, process design

Abstract: Simulation models are used frequently nowadays in the design phase of new products and processes. We will motivate that these simulation tools can be exploited more efficiently by using mathematical optimization methods. This enables the designer to optimize products and processes. We describe two ways of simulation-based optimization and discuss the advantages and disadvantages of both methods.

1. INTRODUCTION

The ever-increasing pressure on the development time of new products and processes has changed the design process over the years drastically. In the past, design merely consisted of experimentation and *physical prototyping*. In the last decade, physical computer simulation models such as finite element analysis models are widely used in engineering design and analysis. The current reliability and stability of these Computer Aided Engineering (CAE) tools has enabled the *virtual prototyping* of complex products and processes.

To stress the important consequences of the developments in simulation we also refer to Ken Wilson. He was the Nobel Prize winner in physics in 1982. He talked about three paradigms of science: the first is theory, the second is experimentation and the third (most recent) is computer simulation!

In this paper we will describe in more detail why simulation has become so popular and argue that even better use can be made of (often time-consuming) simulation by using mathematical optimization methods. Simulation-based optimization is in our view the next step. These optimization methods tell the

G. Q. Zhang et al. (eds.), Benefiting from Thermal and Mechanical Simulation in Micro-electronics, 179–192.
© 2000 *Kluwer Academic Publishers.*

ww

developer which simulations to carry out, such that with a minimal number of simulation runs still much insight into the design is obtained and finally the optimal design can be determined. This paper is a plea to use such optimization methods as an added value to simulation, and to implement efficient optimization methods in CAE tools.

The remainder of this part is organized as follows. Section 2 discusses why simulation has become so important. In Section 3 we argue that simulation-based optimization is the next logical step. We describe and compare two ways of optimization in more detail in Section 4. We end this paper with some conclusions.

2. SIMULATION AS IMPORTANT DESIGN TOOL

There are so many simulation models for simulating physical properties of products and processes and so many users that there is a specific name for it, namely Computer Aided Engineering (CAE)[1]. Simulation is not only widely used in the CAE field, but also in other fields, like:

- Chemical models that simulate chemical processes or chemical characteristics of products.
- Discrete event simulation that simulate, e.g., production, transportation, or inventory systems.
- Specific simulation models, e.g., for electricity generation and transportation, for maintenance optimization.

We see the following reasons why CAE models play such an increasingly important role in the design process[2]:

- Simulation models are becoming more and more accurate, realistic, and reliable, due to increasing availability of computer power (both workstations and PC's) and the improvement of simulation techniques. Therefore, simulation models are more and more trusted by designers.
- The ever-increasing pressure on the development time (time-to-market) of new products and processes has changed the design process over the years drastically. Physically prototyping and experimentation takes too much time and is too expensive. *Virtual prototyping* via computer simulations is much faster and cheaper.
- There is a strong trend towards diversification in products. This means that, for several subclasses of customers, different types are developed and manufactured. Consequently, much more design time is needed. So, there is much need for developing tools to shorten the time-to-market.

[1] Note that CAD (Computer Aided Design) is reserved for drafting, design version management, etc. , and not for calculations.

[2] Some of the reasons mentioned are also valid for other classes of computer simulation models.

المنارة للاستشارات

ww

- Many designers of the 'old generation', who are not accustomed to use CAE tools, are retiring. Moreover, due to job rotation the designers can not build up such a thorough experience as the 'old generation'. CAE tools can in some sense help to bridge this experience gap.

It is impressive to see the enormous market for this class of computer models[3]. Examples of CAE tools can be found in fields like:

- Aerodynamics;
- Computational Fluid Dynamics;
- Computational Electromagnetics;
- Mechanical Engineering;
- Electronic Circuit Engineering.

3. SIMULATION-BASED OPTIMIZATION

CAE tools enable designers to simulate the performance of their products and processes. However, still designers are confronted with the problem of finding settings for a, possibly large, number of *design parameters*. These parameters should be set optimal with respect to several simulated product or process characteristics; see Figure 1. These characteristics, called *response parameters*, may originate from different engineering disciplines. Besides that the design parameters have to satisfy certain constraints (e.g. simple bounds or joint constraints). Since there are still many possible design parameter settings and computer simulations are often time consuming, the crucial question becomes how to find the best possible setting with a minimum number of simulation runs.

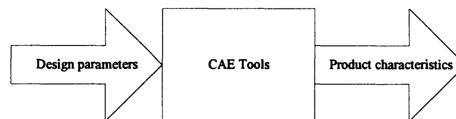


Figure 1. Design optimization problem.

Usually in such a situation, designers use their intuition and experience. They carry out a number of simulation runs and choose the design that gives the best results. This intuitive design approach can be considerably improved by using efficient mathematical optimization methods. Before describing such methods, we first model the design optimization problem as a so-called *high-cost nonlinear programming* problem.

[3] On the Silicon Graphics site (htttp:\\www.sgi.com) there is an overview of hundredth's of CAE tools. We also refer to Daratech for an overview of the CAE-software market.

المنارة للاستشارات

ww

The design problems can be formulated as so-called nonlinear programming (NLP) problems:

$$(HCNLP)\quad \begin{cases} \max f_0(y) \\ f_j(y) \le a_j, \quad j = 1,\dots,J \\ g_k(y) \le b_k, \quad k = 1,\dots,K \\ c_i \le y_i \le d_i, \quad i = 1,\dots,I. \end{cases}$$

In this problem y_i are the *design parameters*, and the unknown constraint and objective functions $f_j(y)$ are the *response functions*[4], and the a priori known design constraints are denoted by $g_k(y)$. The problem is now to optimize a certain chosen response function ($f_0(y)$) such that the other response function values are within certain bounds. The difference, however, with classical NLP problems is that in this case the objective and constraint functions are not explicitly given; only function evaluations by using simulation can be carried out. Since such simulations are often time-consuming we call our problem a *high-cost* non linear programming (HCNLP) problem. Methods for high-cost optimization methods can be divided into two classes: *iterative methods* and *compact model methods*.

4. TWO WAYS OF OPTIMIZATION

In this section we will describe the two possible ways of optimization for our HCNLP problem: iterative and compact model methods. In the sequel we will use the words *simulation point* or *design* for a design parameter setting. In both approaches the so-called *design space* places an important role. The design space is defined as the set of designs which satisfy the a priori known constraints on the design parameters:

$$\begin{cases} g_k(y) \le b_k, \quad k = 1,\dots,K \\ c_i \le y_i \le d_i, \quad i = 1,\dots,I \end{cases}$$

4.1 ITERATIVE APPROACH

An iterative method starts in a certain starting point in the design space, then proposes a new candidate point, simulates this candidate point, and proposes a new candidate point that is based on the new obtained information. Hence, the new candidate point depends on the outcome of the previous point(s).

[4] Sometimes also called *response parameters*, since they are output parameters of the computer simulation(s).

المنارة للاستشارات

ww

Most classical optimization methods are iterative. However, we will first argument that the classical NLP methods are not efficient for our HCNLP problem. We will then discuss other new iterative methods which are specially developed for HCNLP problems.

All iterative methods for general NLP problems can in principle be used for high-cost optimization problem. However, we will show that these methods need too many expensive simulations. We make the distinction between *first-*, *second-* and *zero-order methods*.

Most methods for NLP problems use first- or second-order information. For an overview see Gill, Murray and Wright (1981). Especially interior point methods have become popular in the last decade, because of their nice theoretical properties (den Hertog, 1994). Disadvantages of first- and second-order methods for high-cost nonlinear optimization problems are:

- Since the derivatives of the response functions can only be approximated numerically by, e.g., finite differencing, every iteration needs many expensive function evaluations (i.e., simulations). To calculate the gradient, at least as many function evaluations as there are variables are needed.
- Approximation, rounding, and truncation errors often affect the simulation outcomes. These errors are not stochastic – repeating the simulation will reproduce them – but their accumulation may introduce high-frequency, low-amplitude distortions of the underlying function. Since first and second-order information is obtained via finite differencing, accuracy problems with calculating derivatives may occur in both first- and second order methods.
- Information obtained from previous simulations is not optimally used. In many of these methods the next iterate only depends on the function values and gradients at the current iterate. Therefore these methods needs too many simulations.

Zero-order methods were popular in the sixties and seventies. Examples of zero-order methods are proposed by Hooke and Jeeves (1961), Nelder and Mead (1965), and Powell (1965). Recently a very elegant framework with nice convergence properties has been developed by Torczon (1991,1997) and Lewis and Torczon (1998). Disadvantages of zero-order methods for high-cost nonlinear programming problems are:

- For most zero-order methods there is no guarantee for convergence to local optimum. The exceptions are the pattern search methods of Torczon (1991,1997). But even for these methods there is no convergence guarantee in the case of nonlinear constraints.

- Generally speaking, zero-order methods can not handle NLP problems with nonlinear constraints.

- The third disadvantage mentioned for first- and zero-order methods also hold for zero-order methods. These methods are constructed for problems in which the function evaluations are relatively cheap. Hence the effort put into

المنارة للاستشارات

ww

generating a new iterate is very little, and the information obtained from previous iterates are not optimally used.

It is striking to see that most of the CAE-packages that have some optimization functionalities, use classical zero-, first-, or second-order methods.

To overcome the disadvantages summarized above, new iterative procedures have been developed. For examples we refer to the work of Conn & Toint (1996), Conn, Scheinberg & Toint (1997), den Hertog (1996), Powell (1994,1996), Toropov (1992), and Toropov, Filatov & Polynkine (1993). The common characteristics of all these methods are that:

- Not only the information for the current iterate is used, but also the information of (all) old iterates is used to generate the next iterate.
- The effort put in generating the next iterate is much more than for classical zero-, first-, or second-order methods. Since one simulation run is often time-consuming, spending more time on generating the next iterate is logical.

We will not describe these methods in detail here. We only mention that iterative simulation-based optimization methods is a rather new field that still needs a lot of further research.

4.2 COMPACT MODEL APPROACH

In the compact model approach the unknown objective and constraint functions are replaced by approximations. First, several different simulations are carried out to obtain information how the reponse functions behave. On the basis of this information, approximated (but explicit) functions for the response funtions are calculated. In this way the problem becomes:

$$(CMNLP) \quad \begin{cases} \max \hat{f}_0(y) \\ \hat{f}_j(y) \le a_j, & j = 1,\ldots,J \\ g_k(y) \le b_k, & k = 1,\ldots,K \\ c_i \le y_i \le d_i, & i = 1,\ldots,I, \end{cases}$$

in which $\hat{f}_j(y)$ are approximations for $f_j(y)$, which we will call *compact models*[5].

Below we will sketch a four steps compact model approach. For more details see den Hertog and Stehouwer (1999). In other papers on this subject (e.g., Barthelemy & Haftka (1993), Booker et al. (1999), Dennis & Torczon (1997), Montgomery (1984), Myers (1999), Sachs, Welch, Mitchel & Wynn (1989), Schoofs (1987), and

[5] There are many synonyms used in the literature: approximation models, compact models, meta models, surrogate models, substitute models, etc.

المنارة للاستشارات

Sobieszanski-Sobieski & Haftka (1997)) one or two steps of this four steps approach are treated.

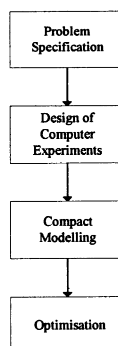


Figure 2. Compact model approach.

This compact model approach elaborates and extends on both Response Surface Modeling (RSM) and what is called Design and Analysis of Computer Experiments (DACE); see Myers (1999) and Sachs, Welch, Mitchel, and Wynn (1989). The approach consists of four steps and is supported by the software package COMPACT. Figure 2 gives these steps. We will now briefly discuss these steps. For a detailed description see den Hertog and Stehouwer (1999).

Step 1: Problem Specification

In the problem specification step, the high-cost NLP problem is formulated as to find settings for the design parameters such that the design and response parameters satisfy certain constraints and some optimality requirement is satisfied. This optimality requirement is expressed in terms of the *objective function* $f_0(y)$. Elements that need to be specified in Step 1 are:

- Design and response parameters
- Design and response parameter constraints
- Objective function

Since the importance of choosing suitable design and response parameters should not be underestimated, we consider the specification of (HCNLP) as the first step of our compact model approach. Some striking examples are given in den Hertog and Stehouwer (1999) that a good problem specification is essential.

Step 2: Design of Computer Experiments

The second step generates a set of suitably chosen simulation points that must lie within the explicit feasible region, i.e., the design space. One can argue the use the *Design of Experiments (DOE)* theory (Montgomery, 1984). However, classical DOE mainly focuses on physical experimentation in which experiments are subject to

المنارة للاستشارات
ww

noise. Classical DOE schemes have the following drawbacks when used for deterministic simulations:

- In deterministic simulations noise does not play a role. Hence, no information is gained from the repeated simulations of the same simulation point such is often done in classical DOE. The consequence is that, in general, DOE needs relatively many simulations to deal with noise.
- Also due to the presence of noise, in physical experimentation it is often optimal to have simulation points lie on the borders of the explicit feasible region. In computer experimentation other parts of this region are often equally interesting.
- A drawback of most classical experimental design methods is that they are only applicable for rectangular feasible regions.

For these reasons we do not propagate the use of classical simulation schemes for the high-cost NLP problem. In den Hertog and Stehouwer (1999) we showed that so-called *space filling* and *non-collapsing* schemes are better for HCNLP problems. For a detailed treatment we refer to den Hertog and Stehouwer (1999).

Step 3: Compact modeling

After Step 2 the proposed (time-consuming) simulations points are evaluated. The third step aims at obtaining good compact models for each of the response parameters. These models are based on the simulation outcomes.

These compact models are already useful at this stage for predicting the response values for specific points. Hence, the compact models can be used as a substitute for the expensive simulations. The software package COMPACT, developed by CQM, supports this predictor feature and also provides several graphical interfaces including 3D surface plots, to visualize the compact models.

As compact models we use either first- or second-order polynomial models (Montgomery, 1984) or Kriging models (Sachs, Welch, Mitchel, and Wynn, 1989). The latter models generally speaking yield the best approximations in case the underlying function has for instance a high nonlinear structure with multiple local minima.

To judge whether a first-order, a second-order, or a Kriging model is more appropriate different models can be compared. As model selection criterion we use cross-validation. See den Hertog and Stehouwer (1999). If, given the current simulation data, the most appropriate compact model does not fit the underlying relationship satisfactory, extra simulation points have to be generated and evaluated. To that end there are two options:

1. Generate and simulate an additional set of simulation points while keeping the feasible region fixed.
2. Shrink the feasible region and generate and simulate an additional set of simulation points for this smaller feasible region. Discard simulation points that lie out of the smaller region. Note that properly shrinking of the feasible

المنارة للاستشارات

ww

region can be done on the basis of the trends observed in the current compact model.

Both these options concern reprocessing of the Steps 2 and 3, i.e., generation and simulation of additional points and building a new compact model.

Step 4: Optimization

Steps 1-3 result in a compact model for each of the response parameters. In this way problem (CMNLP) is obtained. In Step 4 these compact models can be used for optimization, what-if, and sensitivity analysis. Note that performing an extra simulation to check the final solution often finishes Step 4. This section describes the optimization step in the compact model approach.

Optimization consists of finding an optimal solution for the approximated NLP problem (CMNLP). Note that the optimization problems can have thousands of response parameters. This leads to linear (LP) or nonlinear programming (NLP) problems with thousands of constraints. Hence, since we want to carry out the what-if optimization interactively, the (N)LP solvers should be fast. In the software package COMPACT the generalized reduced gradient code CONOPT[6] (Drud, 1992) is used for NLP problems and XA is used for LP problems. Note that the compact models may be non-convex, which means that (CMNLP) is perhaps a non-convex NLP problem, with multiple local minima. Therefore global optimization techniques are used.

It has often proven to be very useful to investigate the impact of the value of a particular lower or upper bound on the value of the objective function of the optimal solution. Such *bound sensitivity analysis* involves re-calculation of the optimal solution for a number of successive bound values. Bound sensitivity analysis is provided for all available types of bounds.

Compact models also enable you to calculate the impact of the value of one of the design parameters on the value of the objective function of the optimal solution. Such *parameter sensitivity analysis* involves re-calculation of the optimal solution for a number of successive parameter values. Note that we, of course, do the sensitivity analysis on the approximated problem (CMNLP), thus avoiding expensive simulations.

4.3 COMPARISON OF BENEFITS

In this section we will compare the two ways of simulation-based optimization by looking at the respective benefits.

Improved designs

By using either the iterative or the compact model approach, designers are able to find much better designs. In den Hertog and Stehouwer (1999) improvements of 10-20% and even 50% are reported for the compact model approach. Normally

[6] In fact the software package COMPACT is built in AIMMS, in which the solvers XA and CONOPT are available.

المنارة للاستشارات

ww

speaking the iterative method will obtain better designs than the compact model approach, since the iterative method really focus on obtaining an optimal solution for one specific optimization problem.

Reduction in developing time
There are three reasons why the compact model approach reduces developing time considerably:

1. Fewer simulations are needed by the compact model approach, since the proposed simulations are such that maximal information is obtained.
2. The simulations a priori proposed by the compact model approach can be carried out in a batch during nights and weekends. Using a trial-and-error method, the designer has to analyze the results of the first simulation(s), then start the next simulation(s), and so on, which costs much more time.
3. Adaptations and redesigns can be done quickly and effectively using the compact models already generated, without the need for extra, often complex and time consuming, computer simulations.

In den Hertog and Stehouwer (1999) a reduction in developing time of 50-60% is reported for the TV-tube design problem! For iterative methods only the first two of the above mentioned reductions in the developing time hold. Iterative methods even need generally speaking less simulations than the compact model approach, since the aim is not to obtain information on the whole design space as in the compact model case.

Multidisciplinary and integral optimization
For both the iterative and the compact model approach, the design and response parameters are not restricted to one simulation tool. This opens opportunities for multidisciplinary and integral optimization. Such a multidisciplinary approach is also reported in den Hertog and Stehouwer (1999) for the compact model variant. Such integral optimization of multiple disciplines avoids sub-optimizing and speeds up the design process because designers remain focussed on the original design objectives. We note that for the compact model approach it is easier to carry out such a multidisciplinary and integral optimization than for the iterative methods. The reason is that the compact model software tool can be implemented as a stand-alone tool, while the iterative methods need a closer interface with the simulation tools.

Non-simulation experts can use compact models
Normally speaking the simulation models that are used are rather difficult. Only experts can do the simulations. However, after a compact model session, these compact models can be used as substitutes for the simulation models. In this way, also non-simulation experts can use the compact models to visualize the responses, to do what-if analysis, to do optimization, and so on. In den Hertog and Stehouwer (1999) it was even reported that local developments in other countries were using compact models developed in a central development center. We observed, although this benefit is difficult to quantify, that this is one of the biggest advantages of the compact model approach; see also (Lasance, den Hertog & Stehouwer, 1999). This

المنارة للاستشارات

benefit of course only holds for the compact model approach and not for the iterative methods.

Certainty and cogency
Both the compact model approach and the iterative methods provide the designers with certainty with respect to the quality of the solution. With the normal trial-and-error method designers stop when there is no time left, or when they are satisfied. The consequence is that they are often uncertain whether the final design can be improved or not after putting more developing time in it. The iterative methods and the compact model approach gives more certainty. This approach also contributes to the cogency to for example management or suppliers.

Less designer dependency
It appeared to be an advantage that by adopting the compact model approach or the iterative method, the result becomes less designer dependent. The design process becomes more stable in the sense that repeating the work in roughly the same way will produce roughly the same (good) results.

Insight in parameter behavior
Besides getting optimal solutions, the designer also gets insight in the behavior of response parameters, the main important design parameters, and quantitatively how important they are. This benefit only holds for the compact model approach.

Insight in trade-off and sensitivities
In the previous benefit, we discussed the insight in the behavior of a response parameter in terms of its design parameters. With the compact model approach, also insight can be obtained in sensitivities in the optimization sense without performing new simulations. How does the optimal solution (or value) change when, e.g., a lower or upper bound on one or more response parameters changes? This benefit does not hold for the iterative variants, since new (often time-consuming) simulations have to be performed when the optimization problem changes.

Finally we want to mention that both the iterative and the compact model approach heavily rely on the reliability of the simulation tool. When the simulation results are unreliable, the outcomes of both the compact model and the iterative approach are also unreliable.

5. CONCLUSIONS

There is an exploding use of computer simulation models nowadays. These simulation models can be exploited more efficiently by using mathematical optimization methods. This leads to simulation-based optimization. There are roughly speaking two ways of simulation-based optimization: iterative optimization and compact model optimization.

المنارة للاستشارات
ww

With iterative optimization methods you can get an optimal solution with a minimal number of simulations. However, no insight is obtained in the behaviour in other regions of the design space. In some CAE packages iterative optimization methods are implemented. However, these methods are often the classical optimization methods, and assume e.g. that first- and second-order information is available. Recently developed optimization methods are much more efficient, since they exploit the special characteristics of simulation models. These new iterative methods should be implemented in CAE tools.

The compact model approach not only solves high-cost nonlinear programming problems in an efficient way, but also leads to a better understanding of the effect of design parameters on responses. Moreover, the advantage of this approach over iterative approaches is that once the compact models are built, slightly changed (HCNLP) can be solved quickly and effectively without the need for extra expensive simulations. A disadvantage of the compact model approach with respect to the iterative variant is that generally speaking less accurate solutions are obtained.

Our opinion is that these advantages are not fully recognized by computer simulation software developers and users. CAE-vendors can add much more value to their simulation tools by implementing these iterative or compact model functionality.

REFERENCES

Aarts, E.H.L. and J. Korst (1989), *Simulated Annealing and Boltzmann Machines,* John Wiley & Sons, New York.

Barthelemy, J.-F.M., and R.T. Haftka (1993), Approximation concepts for optimum structural design – a review, *Structural Optimization* **5**, 129-144.

Booker, A.J., J.E. Dennis, P.D. Frank., D.B. Serafini, V.J. Torczon, and M.W. Trosset (1999), A rigorous framework for optimization of expensive functions by surrogates, *Structural Optimization* **17**, 1-13.

Conn A.R. and Ph.L.Toint (1996), An algorithm using quadratic interpolation for unconstrained derivative free optimization, in: *Nonlinear Optimization and Applications*, G. di Pillo and F. Giannes (eds.), Plenum Publishing, 27-47.

Conn, A.R., K. Scheinberg, and Ph.L. Toint (1997), On the convergence of derivative-free methods for unconstrained optimization, in: *Approximation Theory and Optimization*, tributes to M.J.D. Powell, M.D. Buhmann and A. Iserles (eds.), Cambridge University Press, 83-108.

Dennis, J.E. and V.J. Torczon (1997), Managing approximation models in optimization, *Multidisciplinary Design Optimization*, 330-347.

Drud, A.S. (1992), *CONOPT – a GRG code for large-scale nonlinear optimization – reference manual*, ARKI Consulting and Development A/S, Bagsvaerd, Denmark.

المنارة للاستشارات

ww

Gill, P.E., W. Murray, and M.H. Wright (1981), *Practical Optimization*, Academic Press Inc.

Hertog, D., den (1994), *Interior Point Approach to Linear, Quadratic and Convex Programming: Algorithms and Complexity*, Kluwer Academic Publishers, Dordrecht.

Hertog, D., den (1996), Sequential Regression Programming, *Optix User Guide* (confidential).

Hertog, D. den, and H.P. Stehouwer (1999), High-Cost Nonlinear Programming, Part 1: General description, and Part II: Optimizing color picture tubes, submitted to "European Journal on Operational Research".

Hooke, R. and T.A. Jeeves (1961), Direct search solution of numerical and statistical problems, *Journal of the Association for Computing Machinery* (ACM) **8**, 212-229.

Lewis, R.M. and V.J. Torczon (1998), Pattern search methods for linearly constrained minimization, *ICASE Technical Report* **98-3**.

Morris, M.D. and T.J. Mitchell (1995), Exploratory designs for computer experiments, *Journal of Statistical Planning and Inference* **43**, 381-402.

Montgomery D.C. (1984), *Design and Analysis of Computer Experiments (Second Edition)*, John Wiley & Sons, New York.

Myers, R.H. (1999), Response surface methodology – current status and future directions, *Journal of Quality Technology* **31**, 30-74.

Nelder, J.A. and R. Mead (1965), A simplex method for function minimization, *The Computer Journal* **7**, 308-313.

Powell, M.J.D. (1965), An efficient method for finding the minimum of a function of several variables without calculating derivatives, *The Computer Journal* **7,** 308-313.

Powell, M.J.D. (1994), A direct search optimization method that models the objective and constraint functions by linear interpolation, *Advances in optimization and numerical analyses*, Proceedings of the sixth workshop on optimization and numerical analyses, Oaxaca, Mexico, Vol. 275, Kluwer Academic Publishers, Dordrecht, 51-67.

Powell, M.J.D. (1996), A direct search optimization method that models the objective and constraint functions by linear interpolation, Presentation at the SIAM conference 1996, Virginia.

Sachs, J., W.J. Welch, T.J. Mitchel, and H.P. Wynn (1989), Design and analysis of computer experiments, *Statistical Science* **4,** 409-435.

Schoofs, A.J.G. (1987), *Experimental design and structural optimization*, Ph.D. Thesis, Technical University of Eindhoven.

المنارة للاستشارات

ww

Sobieszanski-Sobieski, J. and R.T. Haftka (1997), Multidisciplinary aerospace design optimization: survey of recent developments, *Structural Optimization* **14**, 1-23.

Torczon, V.J. (1991), On the convergence of the multidirectional search algorithm, *SIAM Journal on Optimization* **1**, 123-145.

Torczon, V.J. (1997), On the convergence of pattern search algorithms, *SIAM Journal on Optimization* **7**, 1-25.

Toropov, V.V. (1992), Multipoint approximation method in optimization problems with expensive function values, in: *Computational System Analysis*, Sydow, A. (ed.), Elsevier, 207-212.

Toropov, V.V., A.A. Filatov, and A.A. Polynkine (1993), Multiparameter structural optimization using FEM and multipoint explicit approximations, *Structural Optimization* **6**, 7-14.

المنارة للاستشارات

ww