PROCESS DEVELOPMENT OF DOUBLE BUMP FLIP CHIP WITH ENHANCED RELIABILITY AND FINITE ELEMENT ANALYSIS

Except where reference is made to the work of others, the work described in this thesis is my own or was done in collaboration with my advisory committee. This dissertation does not include proprietary or classified information

Wei Yan

Certificate of Approval:

Stuart M. Wentworth
Associate Professor
Electrical Engineering

R. Wayne Johnson, Chairman
Ginn Professor
Electrical Engineering

Thomas A. Baginski
Professor
Electrical Engineering

Fa Foster Dai
Associate Professor
Electrical Engineering

Stephen L. McFarland
Dean
Graduate School

PROCESS DEVELOPMENT OF DOUBLE BUMP FLIP CHIP WITH ENHANCED RELIABILITY AND FINITE ELEMENT ANALYSIS

Wei Yan

A Dissertation

Submitted to

the Graduate Faculty of

Auburn University

in Partial Fulfillment of the

Requirements for the

Degree of

Doctor of Philosophy

Auburn, Alabama
December 16, 2005

PROCESS DEVELOPMENT OF DOUBLE BUMP FLIP CHIP WITH

ENHANCED RELIABILITY AND FINITE ELEMENT ANALYSIS

Wei Yan

Signature of Author

Date of Graduation

VITA

Wei Yan, daughter of Hong Yan and Dongdi Zhu, was born on February 14, 1977 in Shanghai, China. She entered the Shanghai Jiao Tong University in 1995 in the Automatic Control Department of the Electronic & Information College. After graduating with a Bachelor of Science Degree with honors in 1999, she was awarded the Graduate Presidential Fellowship from Auburn University to complete a Master of Science in Materials Engineering. In 2001 she has been inducted into Alpha Sigma Wu for her outstanding academic performance at Auburn University. After obtaining her M.S from the department of Materials Engineering, she entered the Doctoral program in the Electrical Engineering Department at Auburn University. She married Han Chen in 2003.

DISSERTATION ABSTRACT

PROCESS DEVELOPMENT OF DOUBLE BUMP FLIP CHIP WITH ENHANCED RELIABILITY AND FINITE ELEMENT ANALYSIS

Wei Yan

Doctor of Philosophy, Dec 16, 2005
(M.S., Auburn University, May, 2003)
(B.S., Shanghai Jiaotong University, P. R. China, 1999)

200 Typed pages

Directed by Dr. R.Wayne Johnson

Flip chip technology has drawn tremendous attention in electronic packaging in recent years due to the advantages it offers, such as better electrical performance, high I/O density and smaller size. However, as package sizes become smaller, with finer pitch and higher density, the standoff height of the solder joints becomes lower, which intensifies the effect of the coefficient of thermal expansion (CTE) mismatch between the die and the printed circuit board (PCB). Based on the Coffin-Manson relationship, higher standoff height provides a smaller shear strain in the solder joint and, therefore, a longer fatigue life.

In this study, an area array flip chip with a double bump solder joint structure was developed to improve the reliability of the package by increasing the standoff height. The new double bump flip chip on board package consisted of the original solder bumps protected by a wafer applied, low CTE underfill and a second layer solder bumps surrounded by no-flow fluxing underfill. This combination benefited from the fluxing underfill's advantage of faster throughput during assembly without sacrificing the reliability by having a low CTE underfill layer near the silicon die.

The bump structure and package geometry were optimized using both simulation and experimental data. Surface Evolver software was used to predict both the solder bump geometry and the standoff height of the solder joint after double bump formation and assembly on board. The computational aspects of solder joint prediction were considered, including the surface tension and gravitational effects. The impact of the second layer solder paste volume on the resulting double bump system was investigated and the results served as a guideline for stencil selection. The restoring force which provides self-centering was calculated for the different cases. The modeling results showed good agreement with experimentally measured results.

A three-dimensional nonlinear finite element analysis was also performed to further explore the solder joint fatigue in a flip chip package. The distributions of stress, strain and volumetric averaged accumulated plastic work in the package and their changes during thermal cycling were studied. Solder joint fatigue life was predicted using several different methodologies. The simulation results for a double bump flip chip model were compared to those of a regular single bump flip chip model and showed that the double bump flip chip experienced lower stress, smaller strain range during thermal cycling and lower accumulative plastic work per thermal cycle. This led to a fatigue life enhancement of about 40% compared with a single bump flip chip. These simulation results were validated experimentally with air-to-air thermal shock reliability tests.

ACKNOWLEDGMENTS

The author would like to thank Auburn University's Department of Electrical Engineering in the College of Engineering and Lord Electronic Materials Group for their assistance and support of this research. The author is deeply indebted to Dr. R. Wayne Johnson for his academic guidance in her pursuit of this degree. His wisdom, rigor, persistent pursuit of excellence in science and technology, and the care and attention he devotes to his students have made him a lifetime example for the author. The financial support provided for this work by Lord Corporation is acknowledged. The author would like to recognize the help and support from Kalyan Ghosh and Russell Stapleton. The author would also like to express sincere thanks to Dr. Stuart M. Wentworth, Dr. Thomas A. Baginski and Dr. Fa Foster Dai for their guidance and support in this research. Special thanks is also delivered to author's colleagues and friends, Tan Zhang, Zhenwei Hou, Yi Liu, Fei Ding, Guoyun Tian, Qing Wang, Mike Palmer, Dr. Shyam Gale, and William Baugh, for their help throughout the course of this research. Finally, the author would like to express deep gratitude towards her husband and parents, whose love, encouragement and understanding have allowed me to accomplish so much throughout my entire educational endeavors.

Style manual or Journal used: IEEE Transactions on Components, Packaging and Manufacturing Technology

Computer software used: Microsoft Office XP, Minitab v13, Ansys 7.0, Ansys 8.0, Surface Evolver, WinSmith Weibull, ImageLab Demo version, Endnote 8.0

TABLE OF CONTENTS

LIST OF TABLES xii

LIST OF FIGURES xv

CHAPTER 1 INTRODUCTION 1
- 1.1 Flip Chip 1
- 1.2 Solder Joint Fatigue 2
- 1.3 Underfill 4
- 1.4 Research Objective and Outlines 8

CHAPTER 2 LITERATURE REVIEW 11
- 2.1 Flip Chip Technology 11
  - 2.1.1 History 11
  - 2.1.2 Advantages 13
- 2.2 Underfill Solutions 14
  - 2.2.1 What is Underfill? 14
  - 2.2.2 Why Underfill? 14
  - 2.2.3 What Are the Desired Properties of Underfill Materials? 15
  - 2.2.4 Conventional Capillary Underfill 16
  - 2.2.5 No-flow Fluxing Underfill 18
  - 2.2.6 Wafer Applied Underfill 21
- 2.3 Reliability Issues 24
  - 2.3.1 Solder Joint Fatigue 24
  - 2.3.2 Underfill Delamination 26
- 2.4 Documented Solder Joint Design for Improving Reliability 26
  - 2.4.1 Stacked Solder Bump Interconnection with Separation Layer 27
  - 2.4.2 Double Bump CSP with Mechanical Spacer 28
  - 2.4.3 Solder Support Structure (S^3) Diepack 29
  - 2.4.4 Fab Integrated Packaging (FIP) and Encapsulated Double Bump WL-CSP 30
  - 2.4.5 Triple Stacked Solder Bumps with High Lead Solder 32
  - 2.4.6 Fine Pitch Pre-Encapsulated Flip Chip 34
  - 2.4.7 Micro-ball Pre-Encapsulated Flip Chip 35
  - 2.4.8 Summary 37

CHAPTER 3 SOLDER JOINT GEOMETRY PREDICTION BY SURFACE EVOLVER 38
- 3.1 Review of Solder Joint Prediction Methodology 38

3.1.1 Truncated Sphere Analytical Method 39
3.1.2 Force-balanced Analytical Model 41
3.1.3 Energy Based Method 42
3.2 Model Development 43
3.3 Stencil Designs 46
3.4 Shape Prediction of Standalone Double Bump Solder Ball After the Bumping Process 49
3.5 Shape Prediction of Double Bump Flip Chip on Board 54

3.6 Shape Prediction of the Restoring Force for a Double Bumped Flip Chip Mounted on a Board 60

CHAPTER 4 EXPERIMENTAL PROCEDURES AND RESULTS 63
4.1 Double Bump Flip Chip Die Fabrication 63
4.1.1 Test Vehicles 65
4.1.2 Coating Process 66
4.1.3 Polishing Process 67
4.1.4 Second Layer Solder Bump Printing Process 70
4.1.5 Reflow 75
4.1.6 Surface Cleaning after Reflow 78
4.1.7 Singulation Process 80
4.2 Assembly Process Development for No-flow Fluxing Underfill 81
4.2.1 Test Vehicle Description 83
4.2.2 Dehydration of Board and Die 84
4.2.3 Underfill Dispensing 87
4.2.4 Die Placement 89
4.2.5 Reflow Process 95
4.3 Reliability Tests 100
4.4 Failure Analysis 104

CHAPTER 5 FINITE ELEMENT ANALYSIS 110
5.1 Literature Review for Finite Element Modeling 110
5.1.1 Approaches to Establishing a Geometric Model 111
5.1.2 Different Material Constitutive Models For Eutectic Solder 115
5.1.3 Life Prediction Analysis Models 119
5.2 Implementation (Modeling Procedure) 126
5.2.1 Material Properties 128
5.2.2 Geometry and Mesh 133
5.2.3 Boundary Conditions 138
5.2.4 Thermal Loading 139
5.2.5 Numerical Simulation 142
5.2.6 Package Deformations 143
5.2.7 Stress Distribution Among the Solder Joints 145
5.2.8 Strain Distribution in the Solder Ball 150
5.2.9 Plastic Work Distribution in the Solder Ball 154

5.2.10 Thermal Fatigue Life Prediction of Solder Joints........................... 158

CHAPTER 6 CONCLUSIONS.. 167
6.1 Literature Review for Finite Element Modeling 167
6.2 Double Bump Flip Chip Fabrication 167
6.3 Double Bump Flip Chip Assembly with No-flow Underfill 168
6.4 Finite Element Analysis... 169
6.5 Future Work 169

BIBLIOGRAPHY 171

LIST OF TABLES

Table 2.1 Desirable features of underfill and corresponding benefits 15

Table 2.2 Comparison of material properties between conventional capillary underfill and no-flow fluxing underfill. 20

Table 2.3 Requirements for wafer applied underfill material during different process steps 23

Table 3.1 Parameters used to calculate flip chip solder joint height 43

Table 3.2 Stencil design guidelines for different stencil type 46

Table 3.3 Stencil aperture parameters and calculated deposited volume 48

Table 3.4 Prediction of double bumped solder ball shape after reflow for different stencil aperture designs 50

Table 3.5 Preferred stencil designs for three different bumping site opening diameters 60

Table 4.1 Properties of the underfill used for thewafer level underfill layer 67

Table 4.2 Preferred stencil designs for three different bumping site opening diameters 74

Table 4.3 Printer setup and printing parameters 75

Table 4.4 Reflow profile parameters 76

Table 4.5 Comparison of the different specifications for regular wafer dicing and coated wafer dicing 81

Table 4.6 Examples of the CSAM pictures under different prebake conditions for boards assembled with regular single bump flip chip dies 85

Table 4.7 CSAM picture for dies with prebake and without prebake for both regular FA10 dies and coated double bumped dies 86

Table 4.8 Material properties for Kester 9126 no-flow underfill 87

Table 4.9 Voiding study of different dispensing patterns 89

Table 4.10 The DOE design matrix for assembly process parameters 91

Table 4.11 Response table based on criteria: smaller is better for voids% response. 94

Table 4.12 Characteristic properties of the two reflow profiles 98

Table 4.13 Reflow oven temperature zone settings for the standard SMT and early-peak profiles (℃) 99

Table 4.14 Summary of AATS test results 104

Table 5.1 Wiese’s sinh law creep model 118

Table 5.2 Darveaux’s sinh law creep model 118

Table 5.3 Feustel’s double power law creep model 119

Table 5.4 Prediction parameters used for models with a 0.0127mm element layer for calculations 125

Table 5.5 Summary of material properties of model 128

Table 5.6 Strain vs. stress data points for Sn-Pb eutectic solder 130

Table 5.7 Creep equation coefficients (time-dependent plasticity) 131

Table 5.8 Anand’s model constants by Darveaux 132

Table 5.9 Model size parameters 134

Table 5.10 Elements and nodes statistics for two models 138

Table 5.11 Simulation results for single bump flip chip using elastic-plastic-creep constitutive model 160

Table 5.12 Simulation results for the double bump flip chip using elastic-plastic-creep constitutive model 160

Table 5.13 Simulation results for the single bump flip chip using Anand constitutive model 161

Table 5.14 Simulation results for double bump flip chip using Anand constitutive model. 161

Table 5.15 Summary of prediction models and parameters used in this study 162

Table 5.16 Summary of prediction using different models. 163

Table 5.17 Experimental results of single and double bump flip chip from AATS test 165

LIST OF FIGURES

Figure 1.1 Comparison of different surface mount device configurations.................2

Figure 1.2 Flip chip solder fatigue problem due to CTE mismatch during thermal loading...4

Figure 1.3 Coffin-Manson low cycle fatigue equation...4

Figure 1.4 A typical flip chip on board package with underfill 5

Figure 1.5 The conventional capillary underfill process 5

Figure 1.6 The no-flow fluxing underfill process 6

Figure 1.7 The wafer-level underfilling process 7

Figure 1.8 Illustration of the double ump flip chip die fabrication procedure..........10

Figure 2.1 IBM controlled-collapse-chip-connection (C4) structure 12

Figure 2.2 Current wafer-level underfill approaches 22

Figure 2.3 Thermal fatigue life of CCB (control collapse bonding) micro-solder joints with different shapes and standoff heights....................................24

Figure 2.4 Solder geometry influence on solder fatigue life....................................25

Figure 2.5 Schematic of stacked solder bumps with polyimide separation layers....28

Figure 2.6 Double bump CSP with mechanical spacer. 29

Figure 2.7 Schematic cross section of the S3-Diepack 29

Figure 2.8 Cross-section of staggered solder balls with SCL on RDL 30

Figure 2.9 Schematic of the double bump wafer-level packaging process flow.......31

Figure 2.10 3D finite element model for 64 bumps package (a) with 20% underfill coverage; (b) with 50% underfill coverage...32

Figure 2.11 High standoff solder joint structure: (a) Triple-stacked hourglass shape (a) Triple-stacked hourglass shape; (b) Triple-stacked barrel shape 33

Figure 2.12 Average fatigue life for different type of solder joint configurations 33

Figure 2.13 (a) Wafer-level encapsulation process diagram and (b) assembly process with pre-encapsulated die 34

Figure 2.14 Optimized pre-encapsulated, bumped die... 35

Figure 2.15 Process flow of micro-ball bumping.. 36

Figure 2.16 Cross-section of the solder joint after assembly on board........................ 36

Figure 3.1 Schematic for truncated sphere model.. 39

Figure 3.2 Schematic for truncated sphere model on board.................................... 40

Figure 3.3 Bumping site (a) top view of the polished wafer surface with exposed solder ball top as bumping site (b) schematic side view showing bumping site opening.. 44

Figure 3.4 The appearance of the predicted 2nd solder bump shape after the bumping process, displayed at an angle to show the 3D effect 45

Figure 3.5 The appearance of the predicted shape of the double bump flip chip on the board using Surface Evolver, displayed at an angle to show the 3D effect.. 45

Figure 3.6 Area ratio and aperture volume for different stencil designs.................. 49

Figure 3.7 Predicted solder ball shape for different stencil aperture types 51

Figure 3.8 Predicted value and measured value of the solder ball height and radius after bumping process for different stencil types (bumping site diameter=4.0mil)... 53

Figure 3.9 (a) The appearance of the solder bump shape predicted using Surface Evolver, (b) the experimentally obtained cross-sectional image of a flip chip die with double bump structure 54

Figure 3.10 The shrinkage problem for the original solder ball during the assembly process... 55

Figure 3.11 Solder joint shape and height prediction for different stencil aperture types with a bumping site opening diameter =4.0mil 57

Figure 3.12 Solder joint shape and height prediction for different stencil aperture types with a bumping site opening diameter =3.5mil 58

Figure 3.13 Solder joint shape and height prediction for different stencil aperture types with a bumping site opening diameter =3.0mil 59

Figure 3.14 Definition of offset percentage 61

Figure 3.15 Solder joint cross-sectional shape for different offset % with 4.5mil pad diameter and 5mil solder joint diameter.. 62

Figure 3.16 Restoring force vs. offset % for double bump flip chip.......................... 62

Figure 4.1 Illustration of the double bump flip chip die fabrication procedure 64

Figure 4.2 Test vehicles: (a) 200x200mil FA10 die (b) 800x800mil wafer square... 65

Figure 4.3 Wafer surface after encapsulation with the underfill layer 67

Figure 4.4 Schematic of the cross-section of the coated die showing the exposed area of the bumping site after polishing (all Diameter and H shows in unit: mil)... 69

Figure 4.5 Coated wafer surface after polishing with exposed solder ball top served as the bumping site for the 2nd layer solder ball bumping........... 69

Figure 4.6 Six stencil patterns on one stencil for double bumping process with apertures matching the FA10 die 72

Figure 4.7 Reflow profile used for bumping process... 76

Figure 4.8 Solder bridging after reflow.. 77

Figure 4.9 Shifted solder balls and aligned solder balls after second layer solder bumping 77

Figure 4.10 Cross-section of the double bumped solder ball after bumping with total bump height of 6.68mil... 78

Figure 4.11 Double bumped die surface before and after cleaning............................ 79

Figure 4.12 Illustration of the assembly process using no-flow fluxing underfill for double bump flip chips.. 82

Figure 4.13 (a) The test board; (b) closer picture of the individual bonding site........ 83

Figure 4.14 Voids detection and area calculation using software ImageLab 93

Figure 4.15 The main effect plot for three factors with voids% as the response 94

Figure 4.16 2-factor interaction plots for all combinations with the three evaluated factors: the underfill weight, placement force and dwell time................ 95

Figure 4.17 Reflow profile no.1 (“standard”) 97

Figure 4.18 Reflow profile no.2 (“early peak”) 97

Figure 4.19 The two different types of reflow profiles 98

Figure 4.20 Non-wetting solder joint of double bump flip chip 99

Figure 4.21 Cross-sectional pictures of single bump flip chip on board and double bump flip chip on board... 100

Figure 4.22 Air-to-air thermal shock chamber temperature profile 101

Figure 4.23 Weibull plot of the thermal shock test data for single bump flip chip and double bump flip chip assemblies 104

Figure 4.24 Cross-sectional picture of a failed single bump die with a crack at the interface between solder and die... 106

Figure 4.25 Cross-sectional picture of a failed double bumped die with a crack at the center neck of the solder joint 106

Figure 4.26 CSAM picture of the double bump flip chip after 1750 cycles showing underfill delamination... 107

Figure 4.27 TAMI scan pictures at different layers using the scanning acoustic microscope, showing the delamination occurs at the interface of the die and the 1st underfill layer... 108

Figure 4.28 Failures due to the underfill delamination at the interface of the die and the 1st underfill 109

Figure 5.1 Two dimensional finite element model with free mesh at the solder ball.. 112

Figure 5.2 Nonlinear slice model... 113

Figure 5.3 Finite element octant model with one critical solder joint in refined mesh 114

Figure 5.4 (a) Octant global model with coarse mesh (b) refined submodel with refined mesh 115

Figure 5.5 Flow chart outlining the steps in solder joint fatigue life prediction using the FEA method 127

Figure 5.6 Temperature-dependent material properties for silicon die 129

Figure 5.7 Temperature-dependent material properties for PCB 129

Figure 5.8 General elastic plastic Model 130

Figure 5.9 Top view of the flip chip package and the slice region represented by the 3D diagonal slice model is marked in the lower corner of the diagram 133

Figure 5.10 Three-dimensional diagonal slice model with mapped finite-element mesh for (a) regular single bump flip chip assembly; (b) double bump flip chip assembly 137

Figure 5.11 Boundary conditions applied to a typical diagonal slice model 139

Figure 5.12 Temperature loads for two cycles in air-to-air thermal shock test 140

Figure 5.13 The deflection of the package due to the CTE mismatch: (a) single bump flip chip model; (b) double bump flip chip model 144

Figure 5.14 Sliced layers for calculation 146

Figure 5.15 Von Mises stress (MPa) distribution in the solder joint for (a) single bump model and (b) double bump model 147

Figure 5.16 Von Mises stress changes during two thermal cycles 148

Figure 5.17 Maximum Von Mises stress value in nine solder balls of the (a) single bump model and (b) double bump model 149

Figure 5.18 Strain distribution in single bump and double bump models 151

Figure 5.19 Von Mises strain changes during the 2 thermal cycles 152

Figure 5.20 Maximum plastic strain values in nine solder balls of the single bump model and double bump model 153

Figure 5.21 Accumulated plastic work/ volume distribution in the single bump and double bump models after two thermal cycles 155

Figure 5.22 Accumulated Inelastic energy density for (a) top/bottom layer in single bump model; (b) top/mid/bottom layer in double bump model............. 156

Figure 5.23 Maximum plastic work/volume value for nine solder balls in (a) single bump model and (b) double bump model... 157

Figure 5.24 Air-to-air thermal shock chamber temperature profile 164

Figure 5.25 Weibull plot of the thermal shock test data for single bump flip chip and double bump flip chip assemblies. 165

CHAPTER 1

INTRODUCTION

The trend in semiconductor devices is to higher density and higher speed. This has forced advancements in device packaging technology. The development of flip chip technology is key to advanced device packaging.

1.1 Flip Chip

The trend in the electronics industry today is to make products more personal by making them smarter, lighter, smaller, and faster, while at the same time making them more functional, powerful, reliable and less expensive. Chip scale packages (CSP) and ball grid array packages (BGA) are the two most popular solutions for modern electronic packaging requirements as shown in Figure 1.1. BGA is a popular surface mount chip package that uses a grid of solder balls as its connection and is noted for its compact size, high lead count and low inductance. A BGA package is typically 35mm or larger with 760µm diameter solder balls. The BGA package has led to the development of the CSP where the package is not larger than 1.2x times the size of the silicon die. The two most common methods for assembly of the die in a BGA or CSP are wire bonding and flip chip. In wire bonding, individual wires are bonded between the I/O pads on the die and the pads on the substrate. The I/O pads on the die are restricted to the perimeter of the die, limiting to total I/O count. In flip chip assembly, the die is flipped to face down on the substrate, making direct electrical connection via solder balls. Solder balls may be distributed over the area of the die, dramatically increasing the number of I/O.

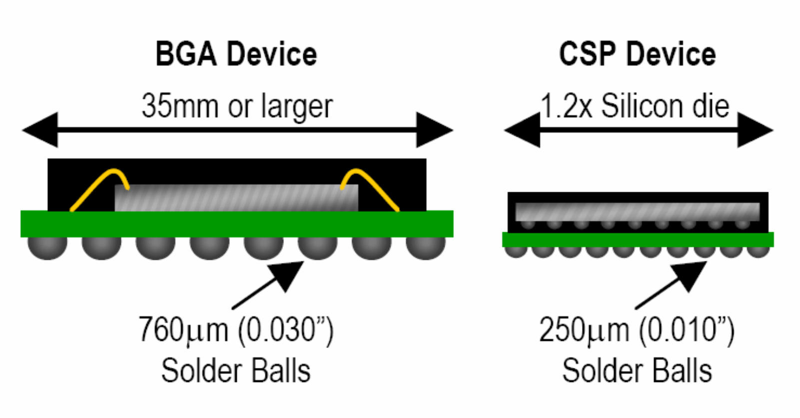


Figure 1.1 Comparison of different surface mount device configurations [1]

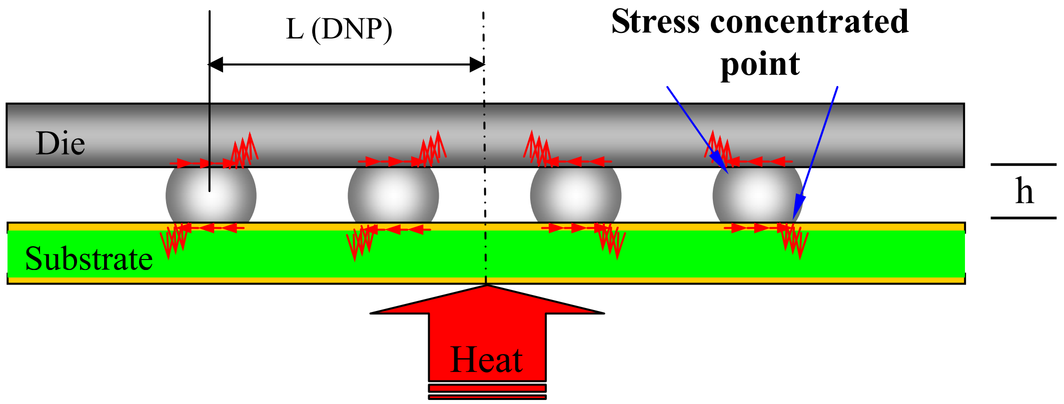
1.2 Solder Joint Fatigue

The major concern of flip chip technology is the thermal mechanical fatigue life of the solder joints. This thermal mechanical issue mainly arises as a result of the coefficient of thermal expansion (CTE) mismatch between the silicon chip (2.8 ppm/℃) and the substrate (typically 15-24 ppm/℃ for organic FR4 board), as shown in Figure 1.2. This CTE mismatch causes shear stress and strain at the solder joints, which makes the flip chip configuration vulnerable and often results in solder joint fatigue failure.

Figure 1.3 illustrates a first-order approximation of the fatigue life of a solder joint described by the Coffin-Manson equation. The solder joint fatigue life is proportional to the square of the solder joint height (h), and inversely proportional to the square of the distance from the neutral point to the center of a solder bump (L, DNP), the coefficient of thermal expansion difference (Δα), and the temperature change (ΔT) experience by the points. This means that there are four methods that can be used to increase reliability:

- ✓ Increase the solder joint standoff
- ✓ Use a smaller die
- ✓ More closely match the CTE between die and substrate
- ✓ Use a smaller temperature range of operation

However, the last three methods run counter to the trend of modern electronics devices, which is toward a larger die, fixing CTE difference by using a low-cost FR4 substrate, and a larger operation temperature range. Thus, increasing the solder joint standoff becomes a natural choice for improving reliability. When underfill is introduced in the package, the relationship between the shear strain and standoff height becomes even more complicated. The effect of the standoff height on reliability is lessened but the trend still remains the same, which is that higher standoff results in better reliability. Solder ball geometry studies and improvement approaches have been extensively reported in the literature and will be reviewed in Chapter 2.4.



$$ShearStrain = \frac{(CTE_{substrate} - CTE_{die}) \cdot \Delta T}{h} \bullet L_{DNP}$$

Figure 1.2 Flip chip solder fatigue problem due to CTE mismatch during thermal loading

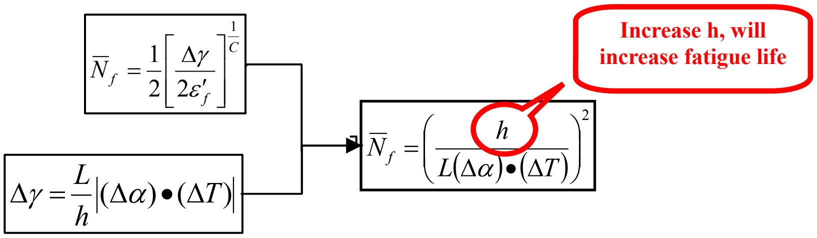


Figure 1.3 Coffin-Manson low cycle fatigue equation [2]

1.3 Underfill

As an effective solution to the CTE mismatch problem, an underfill is normally used to fill the gap between the die and substrate surrounding the solder balls. A typical flip chip on board package with underfill is illustrated in Figure 1.4. The underfill mechanically couples the chip and substrate, hence reducing the stress/strain imposed on the solder joint, and resulting in an improvement in reliability of an order of magnitude or more.

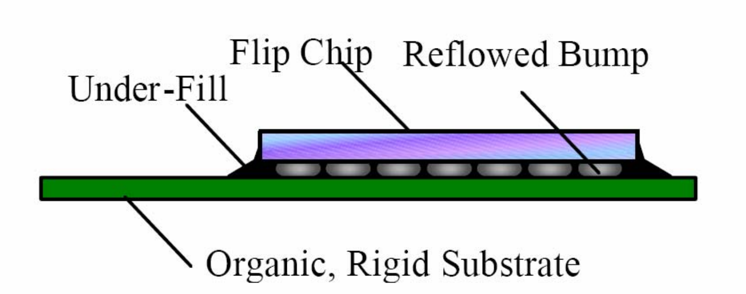


Figure 1.4 A typical flip chip on board package with underfill [3]

Underfill can be classified into three major categories, each of which offers its own advantages and limitations. The conventional underfill, namely capillary underfill, is the most widely used method for underfill [4-9]. This is a post-reflow process that is dispensed after the formation of the solder joints between the chip and substrate. The conventional capillary underfill process is illustrated in Figure 1.5. As its name indicates, capillary underfill uses the capillary forces between the fluid and the package/substrate components, which determines the speed of flow. The capillary flow underfill process is typically done one device at a time, which usually requires a relatively long time to fill the gap between the silicon chip and the substrate, especially for larger size, finer pitch dies. The full cure of capillary underfill is also a lengthy process, requiring about half an hour in most cases. As a result, it is always considered to be the "bottle neck" in the assembly process line.

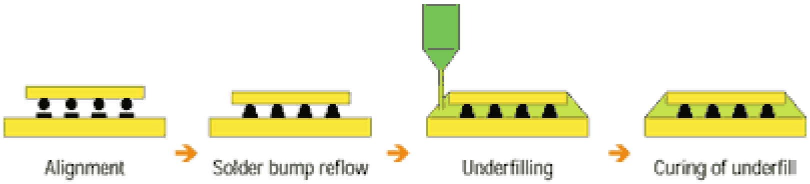


Figure 1.5 The conventional capillary underfill process [10]

No-flow fluxing underfill was developed in order to deal with these issues [11-17]. The no-flow fluxing underfill process dispenses the underfill material on the substrate prior to component placement, as shown in Figure 1.6. The underfill cure is then accomplished concurrently with solder joint formation in the reflow process. By incorporating a fluxing agent in the material, no-flow underfill eliminates the flux dispensing and cleaning steps. Furthermore, it also eliminates underfilling time and curing time. No-flow underfill thus improves the underfill process speed and reduces the number of ovens in the assembly line. However, it also suffers from its own inherent materials properties limitation. No-flow fluxing underfill cannot accommodate more than 20% filler particles because of concerns about filler particle trapped between the solder ball and the substrate pad resulting in low solder joint wetting yield. Consequently, the CTE of no-flow underfill, typically 70ppm/□, is much higher than that of conventional capillary flow underfill materials. Other drawbacks with the use of no-flow fluxing underfill include the need to minimize the voids in the underfill layer and the problem of die floating during reflow.

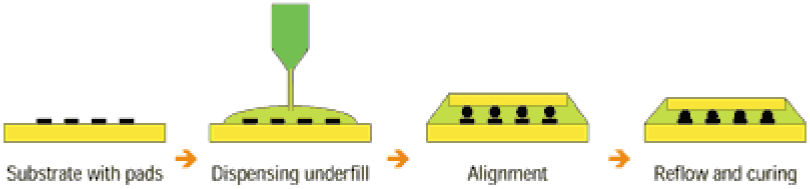


Figure 1.6 The no-flow fluxing underfill process [10]

A more revolutionary approach to the underfill process is the use of wafer applied underfill [18-22]. In this approach, the underfill is applied to the die at the wafer level, eliminating the individual underfill dispensing and hence reducing the process time and cost, as shown in Figure 1.7. The underfill is applied either onto a bumped wafer or a wafer without solder bumps using a proper method. The underfill is then B-staged and the wafer is diced into individual chips. In the case of an unbumped wafer, the wafer is coated before dicing, when the underfill can also serve as a device protective mask [23]. However, though the use of wafer-level underfill seems to offer the most promising answer to an efficient SMT process for flip chips, it still has a lot of material challenges since it needs to be B-stagable, stable at room temperature, and yet able to re-liquefy during the reflow process.

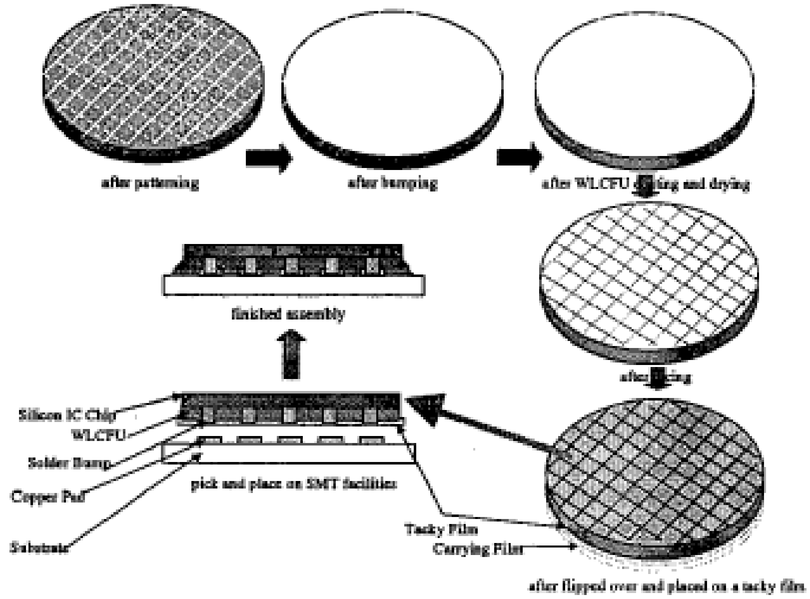


Figure 1.7 The wafer-level underfilling process [24]

1.4 Research Objective and Outlines

The objective of this study was to develop a new flip chip assembly process with improved reliability and faster process time. The flip chip package proposed here has a double bump solder joint geometry with increased standoff height and two underfill layers. The first layer of underfill was applied at the wafer level with a low CTE closely matched to that of the solder balls, and the second layer was a no-flow fluxing underfill applied during the assembly process. The standoff height was increased with improved reliability because of the double bump structure. The two layer underfill system separated the desired function of the underfill into two layers. The first layer underfill took advantage of the benefits of wafer-level underfill without facing its material challenges since it is fully cured rather than B-staged and thus does not need be re-liquefied during reflow, while the second layer benefited from the fast processing properties of no-flow underfill.

Figure 1.8 illustrates the steps for this process. An underfill layer with low CTE was applied at the wafer level and cured. The encapsulated wafer was then polished to expose the original solder ball top, which served as the bumping site for the second layer of solder balls. The second layer of solder balls were formed onto the first layer of solder balls using the stencil print technique, followed by reflow. Finally, the double bumped wafer was cleaned and diced into individual dies for later use. These double bump flip chips were then assembled onto the board using no-flow fluxing underfill and the reliability was compared with single bump flip chip assemblies. More details of the process are given in Chapter 4.

Chapter 2 reviews the different types of underfills and some of the solder joint designs with stacked solder bumps in the literature. Chapter 3 presents a solder joint shape prediction program developed using Surface Evolver software. The results obtained using this software provided guidance for the stencil design that was used in the bumping process. In Chapter 4, the double bumping process is discussed in detail, followed by a description of the development of the assembly process with no-flow underfill. Furthermore, a 3D finite element model was established for both single and double bump flip chip, as presented in Chapter 5. FEM results for both packages were compared in terms of stress, strain and plastic work, as well as solder fatigue life time predictions. Chapter 6 concludes the results and findings of this dissertation, and discusses the topics for future work.

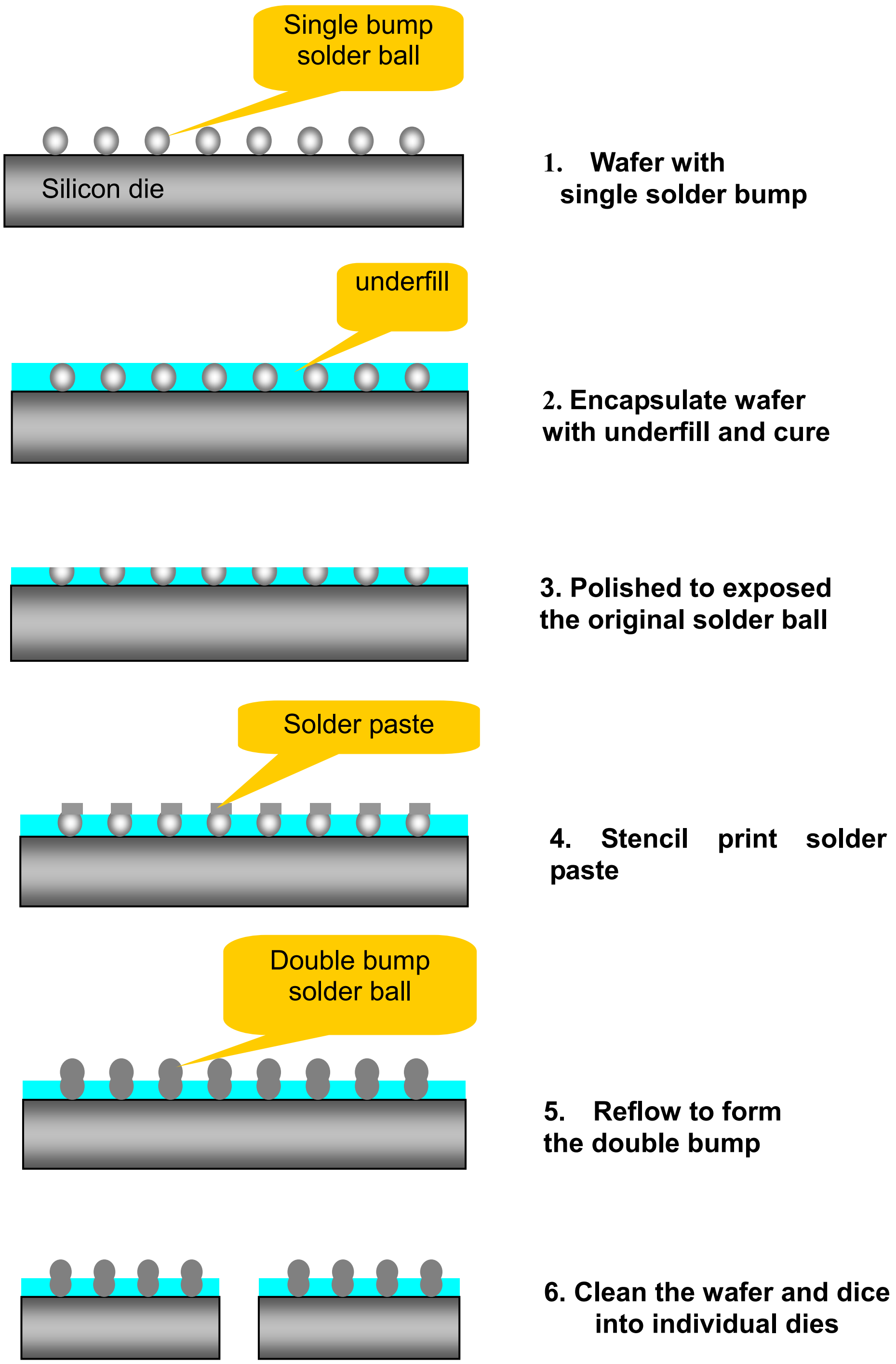


Figure 1.8 Illustration of the double bump flip chip die fabrication procedure

CHAPTER 2

LITERATURE REVIEW

2.1 Flip Chip Technology

In the four decades since the appearance of the first flip chip, an explosive growth and development in flip chip has occurred, with many major equipment makers and leading electronic companies now gearing up for this advanced assembly technology. It has already been used in the commercial products of many companies, including Intel, AMD, IBM, Delphi, Motorola and Hitachi.

2.1.1 History

The first flip chip was developed by Bell Labs in 1957, namely beam lead devices [24]. IBM later developed the Solid Logic Technology (SLT) around 1964, whose interconnection was by a copper micro-ball, not by solder. This was later followed by an innovative technology, the well know C4 (controlled-collapse-chip-connection), replacing the copper balls with solder balls [25]. A cross-section of the C4 structure is shown in Figure 2.1. C4 uses high-lead content (97Pb/3Sn) solder bumps, and C4 die are normally attached to ceramic substrates. C4 assembly offers many advantages compared to wirebond technology, such as self-alignment, high I/O count, design flexibility and low cost. This technology has been significantly improved upon over the decades. In the early stages, various forms of flip chip technology were developed, but were still limited to typically higher temperature solder balls (97Pb/Sn, 50In/Pb, etc.) that were attached to ceramic substrates.

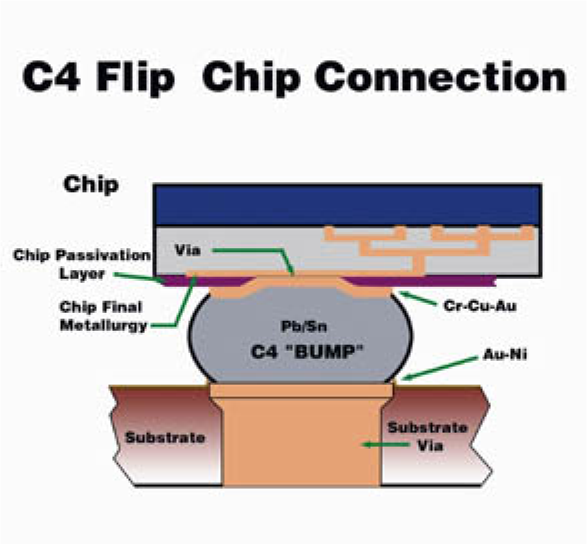


Figure 2.1 IBM controlled-collapse-chip-connection (C4) structure [26]

The transition of the C4 with ceramic or alumina substrates to low cost organic substrates, like FR4, can be described as the introduction of the second generation flip chip. Hitachi made the first flip chip that incorporated an organic chip carrier in 1987 [24]. The move to organic substrates lead to a transition of the interconnections from high lead solder bumps to eutectic Sn/Pb solder bumps, in order to lower the reflow temperature to be compatible with organic substrates. The other difficulty of this transition lies in the relatively high CTE of organic substrates. The solution to this problem is to use underfill to fill the gap, and the second generation of flip chips on organic substrates was not realistic until an underfill was added to the process [25].

Underfill was initially developed for a moisture resistance study, but was unexpectedly found to dramatically improve the life time of the flip chip [27]. The first underfill had a high viscosity, slow flow and needed 6 hours of curing time. The cure time was later reduced down to 30 minutes with a much faster flow rate in the early 1990s, and continued to lead the field by introducing a 15 min cure and finally a 5 min cure system in 1995 [27]. By the late 1990s, capillary flow underfill was a mature technology that was widely accepted by the flip chip industry.

Meanwhile, the concept of a no-flow fluxing underfill was born in Motorola, who received a patent on the combined flux/underfill adhesive concept in 1990. Alpha Metals, Kester and other underfill manufacturers were all involved in no-flow underfill material development. The flux research lab in Alpha/Cookson developed an epoxy-based flux to use with the epoxy-based underfill in order to solve the cleaning problem of flux residue after reflow soldering. Kester later introduced the first commercial product into the market. Substantial quantitive work on no-flow underfill was done at Auburn University [28, 29] and Georgia Tech [30].

Research has been in progress since the mid 1990s for wafer-level underfill technology. Motorola, Loctite, Aguila, Emerson & Cuming, National Semiconductor and Auburn University have been the major players in this underfill race. Approaches proposed include the use of single layer [31] and multilayers of wafer-level underfill [18, 32]. Wafer-level underfill, although still at the early development stage, will offer a broad field of potential applications. Much work remains to be done on both the materials and process aspects.

2.1.2 Advantages

Flip chips offer a number of advantages. First, the size of packages has been continually reduced over the past few decades. The use of flip chips essentially allows the package to be the same size as the die. The fact that flip chips allow area-array connection overcomes many of the limitations associated with wirebond pad pitches. Flip chips offer much lower capacitance and inductance values in a package. This allows the device to operate at higher speed and higher frequency.

2.2 Underfill Solutions

2.2.1 What is Underfill?

The underfill material used in flip chip assemblies is a liquid encapsulant, usually consisting of epoxy resins with or without filler particles (e.g., silica). It is applied in the gap between the chip and the substrate either before or after flip chip interconnection. The underfill materials must have a low viscosity when filling. After curing, the hardened underfill should have good thermal and chemical stability, high modulus and strength, a low coefficient of thermal expansion (CTE) matching that of the solder joint, low moisture absorption, and good adhesion to both the surface of the chip and the substrate.

2.2.2 Why Underfill?

Underfilling is a necessary process for most flip chip assemblies. The primary purpose of underfilling is to reduce the effects of the CTE mismatch between the silicon die and the substrate. Without underfill, a shear stress is introduced because of this CTE mismatch, which results in early life fatigue failure in the solder joints. By underfilling the flip chip, this stress on the solder joints is redistributed between the chip, underfill, substrate, and all the solder joints, rather than being concentrated on the peripheral joints. The PCB movement is constrained, hence the stress is reduced and the joint is preserved. Underfilling also provides moisture control, protecting the component from environmental substances by sealing the interconnect area.

2.2.3 What Are the Desired Properties of Underfill Materials?

An underfill with a high modulus and low CTE is known to significantly reduce the stress on the solder joints, increasing the thermal fatigue life of the package by an order of magnitude or more [33]. Table 2.1 summarizes the general desirable features for an underfill material and the corresponding benefits.

Table 2.1 Desirable features of underfill and corresponding benefits

Feature	Benefit
Low viscosity	Better flowability, complete fill
Optimum filler content	Lower the CTE of underfill to match that of the solder joint
Low CTE	Low stress and strain
Short or no filling time	High throughput
Fast cure	Reduced cost
Fine filler	For small gap fills

There are several different types of underfill technology, and these can be divided into three categories: capillary underfill, no-flow fluxing underfill and wafer applied underfill. These are reviewed in turn below.

2.2.4 Conventional Capillary Underfill

Capillary underfill is probably the most widely used process for underfill. These typically consist of epoxy resin highly filled with silica or other fillers.

As mentioned in the previous chapter (Figure 1.5), the underfill is dispensed along the periphery of one or two sides of the chip. Capillary action pulls the underfill inwards through the space between the chip and the board. In order to achieve a complete fill, the underfill material must have a low viscosity. Heating the board and component to 70-100□ is a common requirement that decreases the underfill viscosity, improves wetting and aids the flow. After filling is complete, the board is taken to an oven where it is cured. There are many different cure schedules, depending on the epoxy resin used. A typical cure schedule for underfill is between 15-30 minutes at 150-165℃ based on the properties of the materials. Some snap cure material can speed the cure process to as short as 5 minutes.

The essential features of a capillary flow of a Newtonian liquid between parallel plates are well described by the Washburn model [34]:

$$X_f^2 = \frac{\sigma h \cos\theta}{3\eta} t \qquad (2\text{-}1)$$

where σ is the surface tension, θ is the wetting angle, h is the plate spacing and η is the Newtonian liquid viscosity. The Washburn model assumes that the underfill is laminar, one dimensional and incompressible and that it can be treated as a Newtonian fluid. From this formula, the underfilling time can be estimated as

$$t_f = \frac{3\eta L^2}{\sigma h \cos\theta} \qquad (2\text{-}2)$$

where L is the length of the cavity to be underfilled. Experimental data have shown a good fit with the Washburn model prediction [34]. As the formula indicates, the underfilling time is proportional to the square of the flow distance and inversely proportional to the gap height. The problem with the slow underfilling time is aggravated further with an increase in chip dimensions and decrease in standoff gap height.

The flow properties are also temperature dependent, with higher temperatures increasing the flow rates. However, the temperature should not get to high or the material will start to cure while it is still being flowed. A typical board temperature of 70-100□ was reported to be suitable to aid the flow process [35, 36].

Voiding is one of the problems associated with low viscosity underfill. The sources of the voids are:

✓ Underfill material properties
✓ Board moisture level
✓ Fast flow at edges

The process variables affecting this issue include prebake temperature and time, board temperature, dispensing pattern and speed, and board surface topography.

Fillers are used to modify a number of key properties of the underfill, such as its rheology, strength, toughness, modulus, CTE and thermal conductivity. While some of these properties are improved due to the presence of the filler particles, other properties are compromised.

Filler settling is another problem associated with low viscosity underfill. This may result in deviations of the material properties from target values, resulting in a resin rich layer on the die side, and potentially causing reliability issues. Wang and Chen [5] investigated the effect of flow properties on filler settling in the underfill of a flip chip and found that the filler size in the underfill affects its rheological properties. An underfill with a small filler size, low viscosity at the dispensing temperature was found to have the least filler settling.

There has been a tremendous amount of research on process related underfill issues. Chia, et al. [37] developed a model estimating and optimizing the operating range of the dispensing volume of a defined flip chip assembly. Young [38] investigated the effect of solder bump pitch on underfill flow and estimated the flow resistance induced by the component, substrate and solder bumps, providing a useful tool for underfill flow front prediction. Further capillary underfill characterization literature has focused on topics such as void formation [39, 40].

Several novel capillary flow processes have been invented to address the problems [41]. Banerji [42] invented a capillary flow underfill process that utilizes a vacuum, which helps to produce a void-free underfill layer with shortened underfilling time. Pressure can also be used to assist in the capillary flow process, in a similar way to the vacuum-assist process [43]. Here, the underfill is injected under the chip with a sealed housing underfill pressure; the flow is primarily driven by the pressure gradient from the inlet rather than the surface tension at the wetting front.

2.2.5 No-flow Fluxing Underfill

Several innovative underfill methods and materials have been developed to deal with the issues identified with capillary underfills. No-flow fluxing underfill is a promising material method that offers many advantages. The material properties of the no-flow fluxing underfill are very different than those of the capillary underfill due to the process difference. Table 2.2 summarizes typical material properties for both capillary underfill and no-flow fluxing underfill for comparison. The major differences lie in the CTE, modulus and filler content.

No-flow fluxing underfill utilizes thermally polymerizable liquid resin systems that contain fluxing additives. The procedure for using no-flow materials is to dispense the desired volume of underfill material onto the substrate prior to the component placement. The component is then placed on top of the underfill and pressed through to establish contact with the substrate. For die of various sizes and different substrate layouts, many researchers have suggested that it is necessary to include a dwell time when pressing the component down into the underfill fluid; holding the component down momentarily allows the edge of the die to be wetted by the underfill, and therefore eliminates the floating of the component on the top of the underfill fluid [15, 16, 44].

By incorporated the fluxing agent into the underfill system, the no-flow fluxing underfill serves as the flux, removing possible oxidation on the solder joints during the reflow process. The underfill cure is accomplished later, concurrently with solder joint formation in the reflow process. In order to achieve a successful assembly, it is important that the distinct processes of oxide removal, solder melting and resin polymerization occur in a precise sequence. If for instance, the underfill gels prematurely, solder will be constrained in its original shape and not collapse to wet onto the pad. On the other hand, if the polymerization of the underfill is too slow during the reflow, it will not be completely cured, requiring a post-reflow cure. Hurley at Cookson and Johnson at

Auburn University [28] proposed a kinetic model for the no-flow underfill cure, by which solder melting and underfill volatilization can be related to the degree of underfill cure. This model provided a method for optimization of reflow profile with maximum assembly yields and minimum underfill voiding.

Table 2.2 Comparison of material properties between conventional capillary underfill and no-flow fluxing underfill

	Conventional Capillary Underfill	No-flow Fluxing Underfill
Curing temperature	< 150 °C	Compatible with solder reflow
Curing time	< 30 min	
Tg	> 125 °C	Usually lower
Working Life	> 16 hrs	Similar
CTE	22 ~ 27 ppm/ °C	60-80 ppm/ °C
Modulus	8-10 GPa	2-3 GPa
Fracture toughness	>1.3 MPa · $m^{1/2}$	Typically lower
Moisture absorption	<0.25 %	Typically higher
Filler Content	<70 wt%	None or very low

"Die float" is a potential problem in the no-flow underfill process. Here, the die is separated from the substrate during the reflow process due to the movement of the die resulting from the dynamics of resin expansion or mechanical movements induced during the reflow process. Rendse, et al. [45] from Nvidia proposed a "thermode assisted chip attach process" to solve the "die floating" problem in no-flow underfill. In this process, heat and force are used together to form a temporary mechanical/metallurgical bond between the bump and pad during the die placement. This process also successfully accommodate up to 50% filler in the underfill.

Voiding is another problem frequently associated with the no-flow underfill process. Unlike capillary underfills, where the flow front displaces air, the no-flow process has a tendency to trap air as the bumped chip is placed into the underfill. Air entrapment, the so called voiding problem, remains one of the main problems with the no-flow fluxing underfill method.

Zhang, et al. [46] at Auburn University developed a computational model for the no-flow fluxing underfill process using the Surface Evolver software. This model investigated the effect of volume and material properties on the no-flow process. Fillet geometry, solder geometry and forces were included in the model. This model is also suitable for the wafer applied underfill process, and will be used in the current study.

2.2.6 Wafer Applied Underfill

Wafer applied underfill technology has the potential to introduced a major shift in the current industrial structure, since it brings the underfilling task from the assembler to the backend wafer processing. Several strategies have been explored for wafer applied flip chip underfill since the mid-1990s. For example, the underfill can be either coated or printed onto the wafer as a liquid and then B-staged into its solid form, which means the material is partially cross-linked and can be re-liquefied by heating. These strategies can be divided into those used for single layer and two layer systems (Figure 2.2). The material requirements for wafer applied underfill are completely different from the other underfill types. Major requirements for wafer applied underfill during the three stages of the process are summarized in Table 2.3.

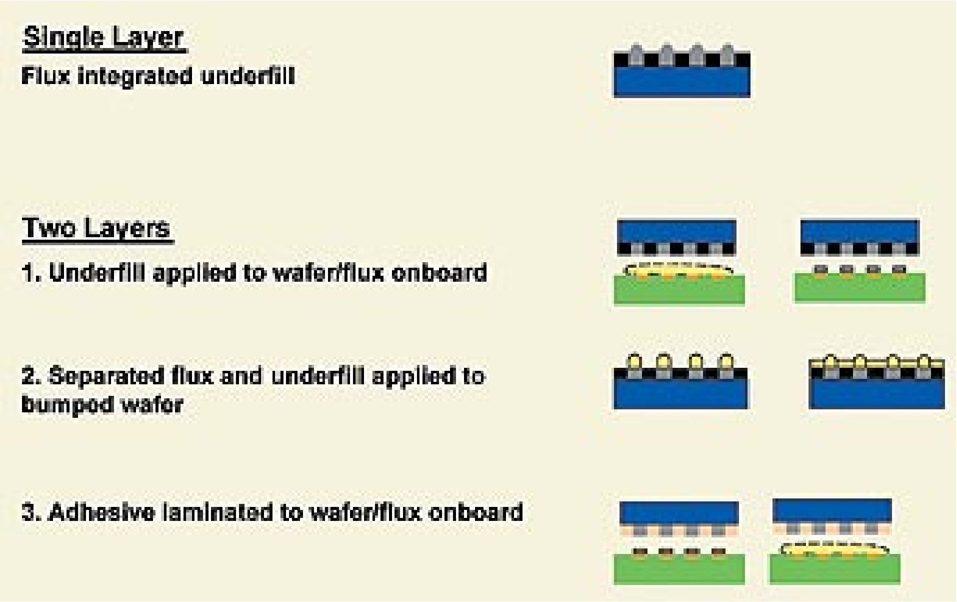


Figure 2.2 Current wafer-level underfill approaches [21]

A single material layer is generally the most desired approach. Here, the materials must have flux properties initially but then convert to an underfill [31]. This is possible from a chemistry standpoint, since flux agents such as carboxylic acids can also serve as hardeners. However, this limits shelf-life.

Two-layer systems consist of separate underfill and flux layers [18]. For obvious reasons, the development of two layer system has focused more on the method than on the materials themselves. No flux activity is required and filler can be used since the flux layer is applied separately. Alternatively, wafer-level underfill can also be applied before the bumping process [47].

Table 2.3 Requirements for wafer applied underfill material during different process steps

Process steps	Required underfill material properties
B-staged	T_g > Room temperature Tack-free for easy handling Stable at room temperature for ease of storage and longer shelf life
During wafer dicing process	Minimum hardness required to survive the dicing process Resistant to heat and moisture
Reflow process	Must re-liquefies during reflow Should not gel before solder melts and wets the substrate Fully cured after reflow

In both systems, the coating of the underfill on the wafer presents a great challenge and the coating height must be carefully controlled. Considering that the wafer-level underfill was originally intended for fine pitch flip chip, controlling the coating height is not an easy job, not to mention that uniformity across the wafer is always required. Most approaches coat the bumped wafer before dicing, which requires the B-staged material be able to withstand both heat and moisture. An alternative wafer applied underfill approach developed by Motorola, Loctite and Auburn University coated the wafer after dicing [18]. This technique eliminates any potential issues associated with exposing uncured epoxy to the water used in the cooling system of the saw blade. Instead of depositing a liquid underfill and then B-staging it, 3M and Delphi-Delco have used a solid film laminated onto the bumped wafer in a vacuum [32].

2.3 Reliability Issues

2.3.1 Solder Joint Fatigue

Solder joint fatigue is a serious concern in flip chip assemblies. Thermal strains and stresses caused by CTE mismatches are the main cause of failure in solder joints. Numerous factors affect solder joint fatigue performance, such as joint geometry, chip size, interface metallurgy, underfill and substrate materials. Of these factors, solder joint geometry has been extensively reported as playing an important role by many researchers [48-51]. Satoh, et al. [49] reported mathematical calculations that showed that hour glass-shaped solder joints would have the lowest plastic strain and stress during a temperature cycle, and thus the longest lifetime (Figure 2.3)

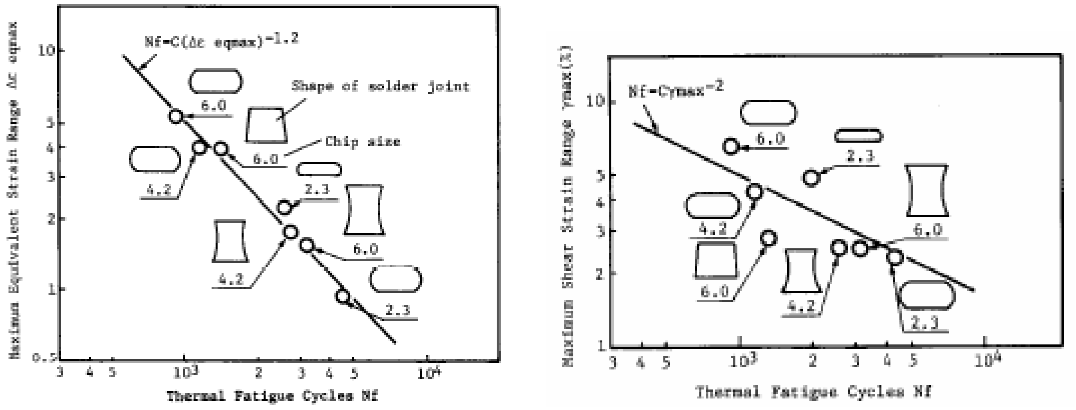


Figure 2.3 Thermal fatigue life of CCB (control collapse bonding) micro-solder joints with different shapes and standoff heights [49]

Liu [48] and Chiang and Lin [51] found that solder joint reliability is highly dependent on the solder joint geometry, such as the standoff height, lower/upper contact angles of the solder joints, and the solder joint profile. Popelar [50] employed a finite element solder fatigue model in his parametric study of the fatigue life of flip chips on organic substrate assemblies. His study investigated parameters including die size, die thickness, substrate thickness and underfill properties, as well as specific solder joint parameters such as conductor thickness ("C", as shown in Figure 2.4), total standoff ("s", as shown in Figure 2.4) and net solder joint height ("h", as shown in Figure 2.4). His results, shown in Figure 2.4, revealed that the CTE of the underfill and the standoff ratio (defined as solder joint height/standoff, h/s) are the two most significant parameters affecting solder joint fatigue life.

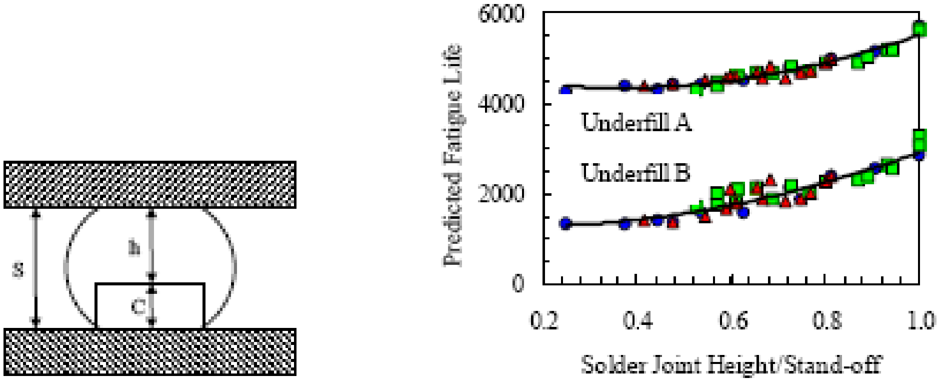


Figure 2.4 Solder geometry influence on solder fatigue life [50]

Liu and Lu [52] found that solder joint crack propagation time is mainly determined by the solder joint standoff height based on their experimental results; a tall compliant solder joint has enhanced mismatch absorption capability. Therefore, for solder joint reliability optimization, the most effective approach is to increase the solder joint standoff height to the maximum that the solder joint fabrication process allows. However, it is very difficult to increase solder joint height without sacrificing other parameters such as the solder joint pitch, I/O density and cost.

2.3.2 Underfill Delamination

Underfill delamination is another failure mode that affects solder joints, leading to the failure of the joint [53]. However, other researchers have demonstrated that no delamination is observed before solder joint failure in the case of a good adhesive underfill [54].

Several experimental measurements of underfill delamination have been reported in recent years. Gurumurthy, et al. [55] and Kook, et al. [56] calculated the strain energy release rates, G, near the crack tip under different conditions in their research and obtained relationships between G and da/dn, the crack growth rate. Wang, et al. [57] experimentally measured the interfacial fracture toughness of the chip/underfill interface using three-point bending and Wong, et al. [58] measured the apparent fracture toughness of the interfaces between epoxy-based underfill materials and various substrates. Lau, et al. [59] investigated the modes of delamination propagation using finite element simulations in flip chip packages. Cheng [60] measured the delamination propagation rates at the interface between the chip and underfill by using C-SAM inspection of flip chip assemblies under thermal cycle loading. A finite element model was also utilized in his work to calculate the strain energy release rates, G, and the phase angles, φ, near the delamination crack tip.

2.4 Documented Solder Joint Design for Improving Reliability

Based on the previous literature survey, the key factor that influences solder joint reliability is the geometry of the solder joint, especially the solder joint stand off height.

Due to the similarity between CSP and flip chip, where the major difference lies only in their size, several alternative solder joint designs in CSP packages are also reviewed in the following section. Applied Electronics Lab's stacked solder bump interconnection [61], AT&T Bell Lab's double bump CSP with spacer [62], Technical University of Berlin's S^3 diepack [63, 64] and Virginia Tech's triple stacked solder bumps [52] were all designed for CSP packages with a solder bump diameter range of 300μm to 900μm. Only Aguila's wafer-level pre-encapsulating device [65] and Nippon's micro-ball double bumping package [66] were designed for flip chips with much smaller solder balls and correspondingly finer pitches.

2.4.1 Stacked Solder Bump Interconnection with Separation Layer

Matsui, et al. [61] at Applied Electronics Laboratory proposed a new-type of solder ball interconnection technology using stacked solder bumps supported by polyimide films (Figure 2.5). Bump-limiting metal pads were used to separate the upper and lower solder bumps. The use of three metallurgy layers Cu-Ti-Cu was found to be the optimum bump-limiting metal pad structure, with titanium as the barrier layer. Otherwise the solder diffused into the lower bumps losing the desired standoff height and causing the interconnection to fail. Their thermal shock (-70℃ to +25℃) results indicated that the double stacked solder bumps with 300μm bump diameter joining 20x20mm silicon chips to alumina ceramic substrates provided a sixty times longer lifetime than the conventional un-stacked solder bumps. No underfill was utilized in this study.

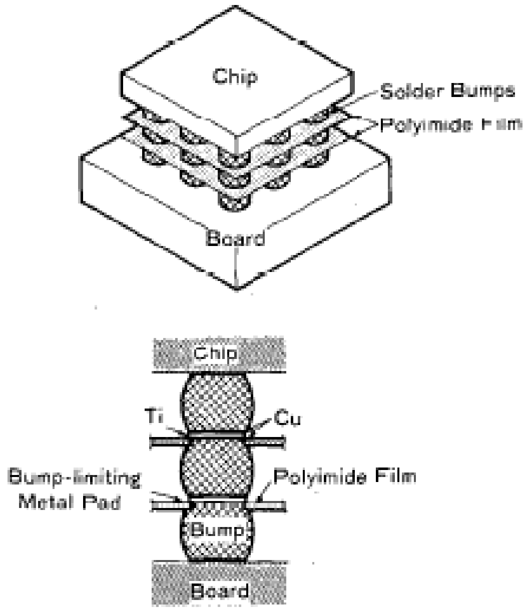


Figure 2.5 Schematic of stacked solder bumps with polyimide separation layers [61]

2.4.2 Double Bump CSP with Mechanical Spacer

Lovasco and Oien from AT&T Bell Laboratory [62] developed a double bump technology that entailed the controlled overlapping of two molten solder bumps on both package and substrate to form a nearly cylindrical joint as shown in Figure 2.6. The bumped package was placed on the bumped substrate and held laterally in place by a set of alignment pins attached to the corners of the package. Standoffs on the corner pins of the package were used as mechanical spacers to control the final separation between the package and the PCB, and hence the height of the resultant solder joints.

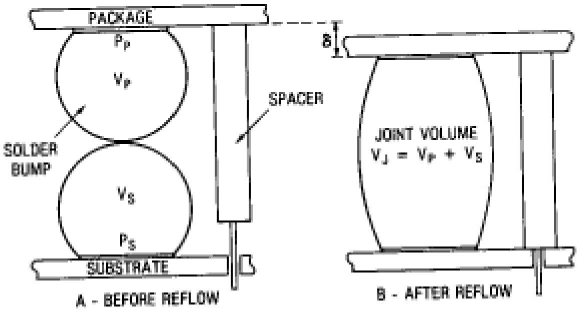


Figure 2.6 Double bump CSP with mechanical spacer.

(a) before reflow, (b) after reflow [62]

2.4.3 Solder Support Structure (S ^{3}) Diepack

Simon and Jasper at the Technical University of Berlin [63, 64] developed a CSP with a solder support structure (S^3) to improve board level reliability. Here, two layers of solder spheres were used. The first layer of solder spheres with a diameter of 300μm was attached to the redistribution layer with a 0.5mm pitch. These solder spheres were reinforced by a solder support structure using a filled epoxy to avoid collapsing of the stacked solder spheres during reflow.

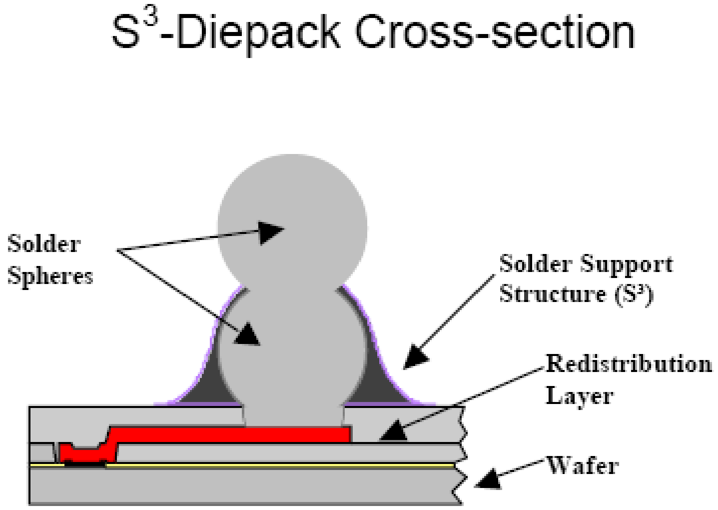


Figure 2.7 Schematic cross section of the S^3-Diepack [63]

2.4.4 Fab Integrated Packaging (FIP) and Encapsulated Double Bump WL-CSP

Motorola and Fraunhofer-IZM, Technical University Berlin started a program in 1996 to develop new concepts for high-reliability wafer level CSP's. Topper, et al. [67] at Fraunhofer-IZM proposed the initial idea of the so called Fab Integrated Packaging CSP (FIP-CSP), with the structure as shown in Figure 2.10. The concept they developed consisted of a stacked solder ball array with a stress compensation layer (SCL) surrounding the lower solder spheres. An SCL layer is used to embed the high lead solder balls before the second eutectic solder balls were stencil printed on top of the embedded balls.

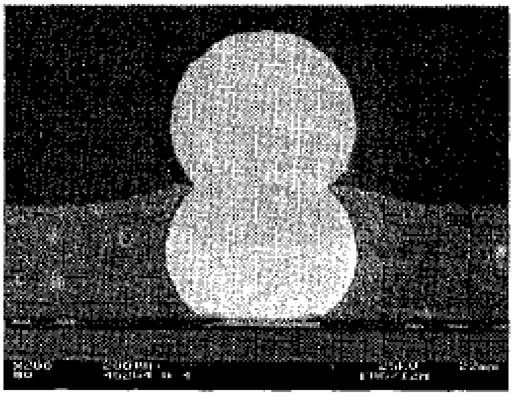


Figure 2.8 Cross-section of staggered solder balls with SCL on RDL in FIP-CSP packages [67]

Keser, et al. [68] at Motorola later improved the process by replacing the stencil printed bumps with direct placement of preformed balls in order to increase the bump height. The observed board-level reliability was an order of magnitude better than that of single ball analogous structures without underfill. Figure 2.9 shows the process flow of their package.

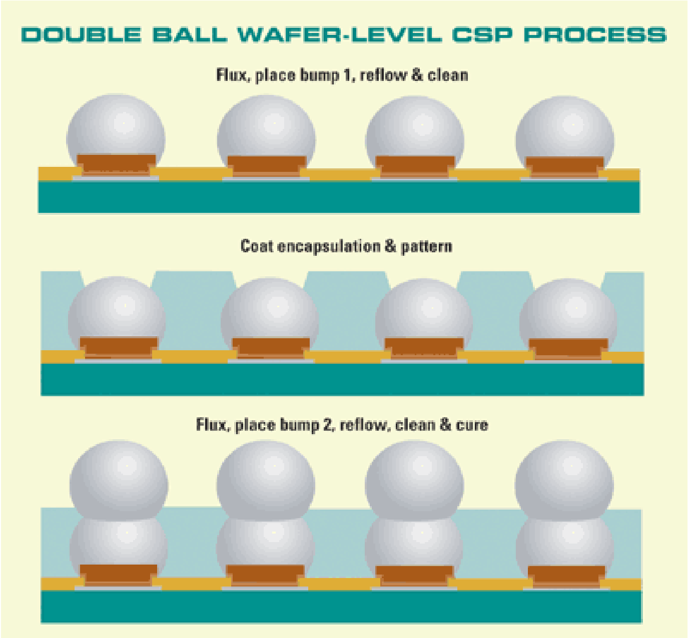


Figure 2.9 Schematic of the double bump wafer-level packaging process flow [68]

Other work on similar solder bump structures has been presented by Patwardhan, et al. at National Semiconductor Corporation [69]. This paper focused on the key advantage of this technology, the ability to control the thickness of the underfill layer as a percentage of the solder bump height. A FE model predicted a 3% increase and 14% decrease in fatigue life for packages with 20% and 50% underfill coverage, respectively as shown in Figure 2.10.

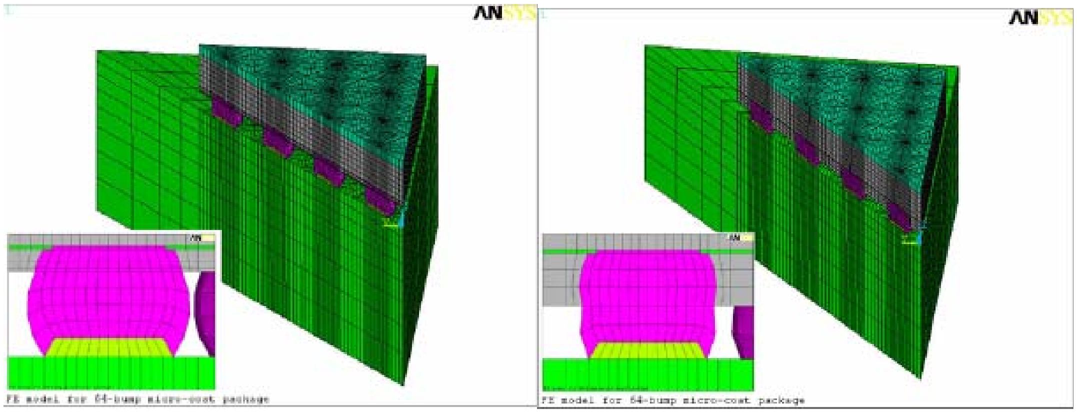


Figure 2.10 3D finite element model for 64 bumps package (a) with 20% underfill coverage; (b) with 50% underfill coverage [69]

2.4.5 Triple Stacked Solder Bumps with High Lead Solder

Liu and Lu [52] at Virginia Tech developed stacked solder bumping technology using triple hourglass and triple barrel shapes. This was an extension and combination of the stencil printed solder bump technology and micro-ball mounting technology. This approach offers better control of the solder joint height and shape. However, it uses Sn96.5/Ag3.5 alloy (with 221℃ melting temperature) for the inner solder cap, and high lead (Sn10/Pb90) solder, with a melting temperature of 268℃ as the middle solder ball. By using a reflow temperature of 250℃, it allows the middle solder ball to preserve its shape and hence maintain the stacked standoff height (Figure 2.11). Liu and Lu concluded that solder joint shape is the dominate factor determining crack initiation time, and solder joint height is the major factor affecting crack propagation time (Figure 2.11).

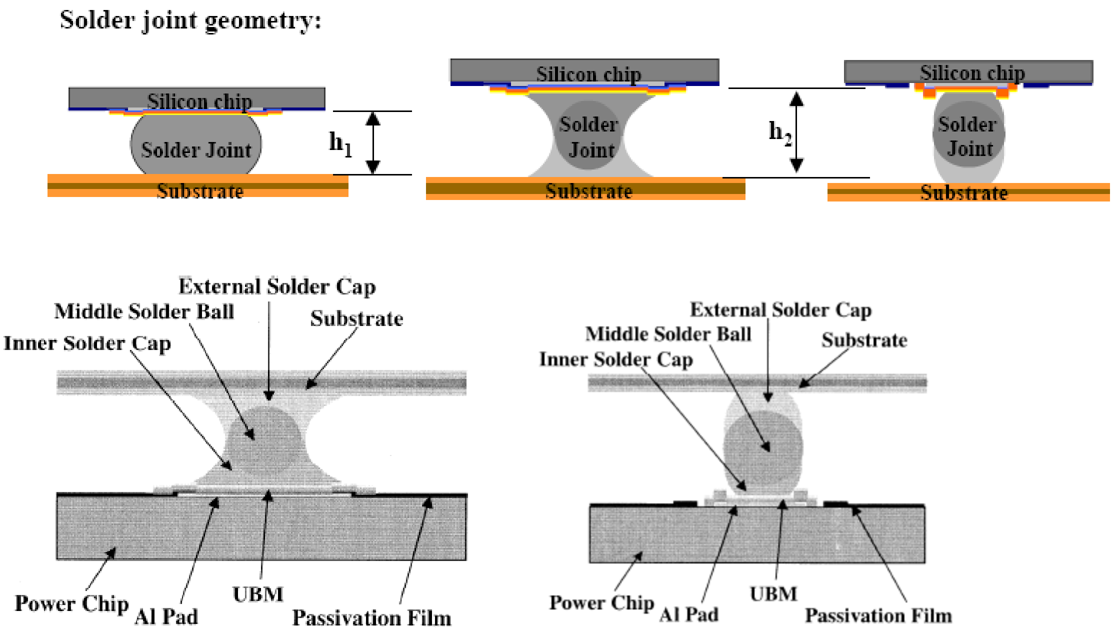


Figure 2.11 High standoff solder joint structure:
(a) Triple-stacked hourglass shape; (b) Triple-stacked barrel shape [52]

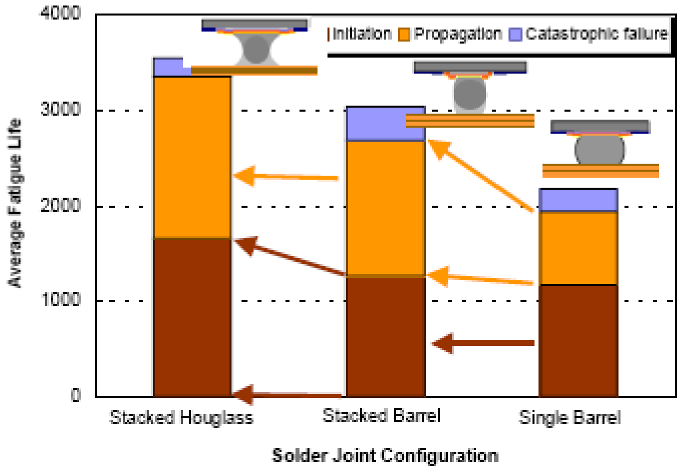


Figure 2.12 Average fatigue life for different type of solder joint configurations [52]

2.4.6 Fine Pitch Pre-Encapsulated Flip Chip

Grieve, et al. [65] at Aguila Technologies proposed an approach to developing stacked solder bumping technology, by using the structure of a small bump stacked on a regular bump (Figure 2.15). This approach utilizes encapsulation of wafers with multiple material layers that are deposited at the wafer-level. The entire wafer is encapsulated using either a dry-film underfill or by stencil printing, while the second layer encapsulation layer is done during the assembly process using a no-flow underfill.

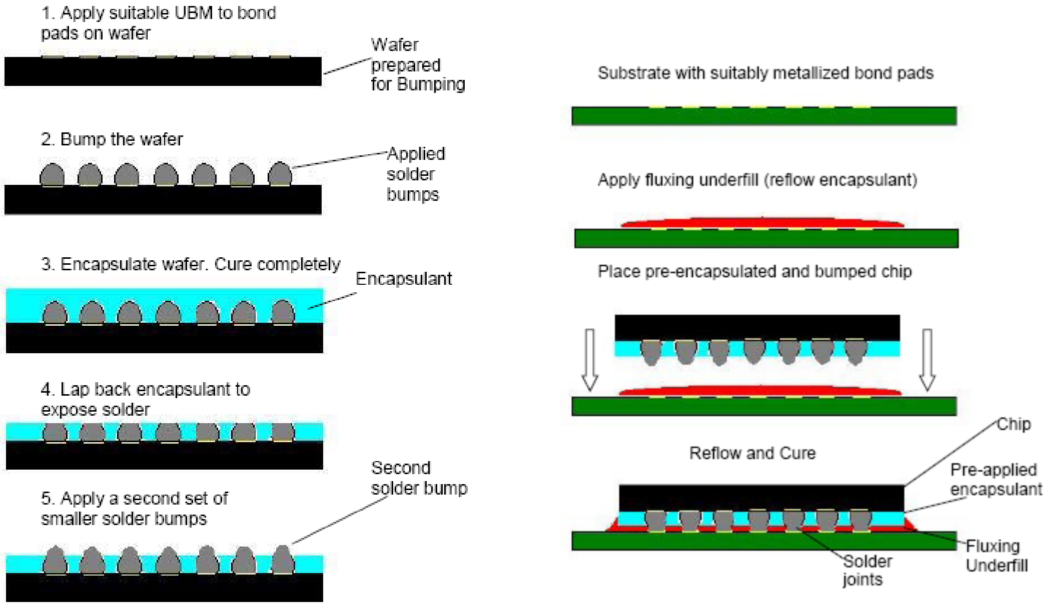


Figure 2.13 (a) Wafer-level encapsulation process diagram and (b) assembly process with pre-encapsulated die [65]

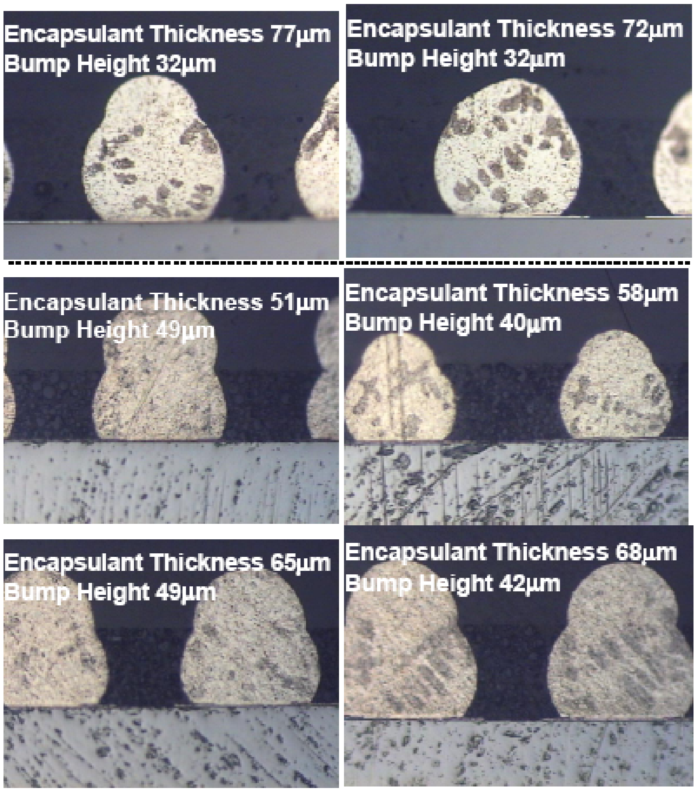


Figure 2.14 Optimized pre-encapsulated, bumped die [65]

2.4.7 Micro-ball Pre-Encapsulated Flip Chip

Tatsumi, et al. [66] at Nippon, Japan presented a double bump flip chip utilizing a micro-ball wafer bumping technique. The wafer was first bumped with micro-balls (Figure 2.15) and then encapsulated with epoxy resin containing silica fillers by using Apic Yamada's wafer-level molding system. The second micro-ball bumping was processed on top of the first bumps. The first and second balls were connected by reflowing to form the double ball bumps. However, no additional underfill was used besides the pre-encapsulated layer used in the package when it was assembled onto the board. Therefore, in thermal cycle tests (-55℃ to 125℃, 15 min dwell), though the double bump showed an improved reliability compared to single bump, they both still had relatively short fatigue lives of 65 cycles and 350 cycles, respectively.

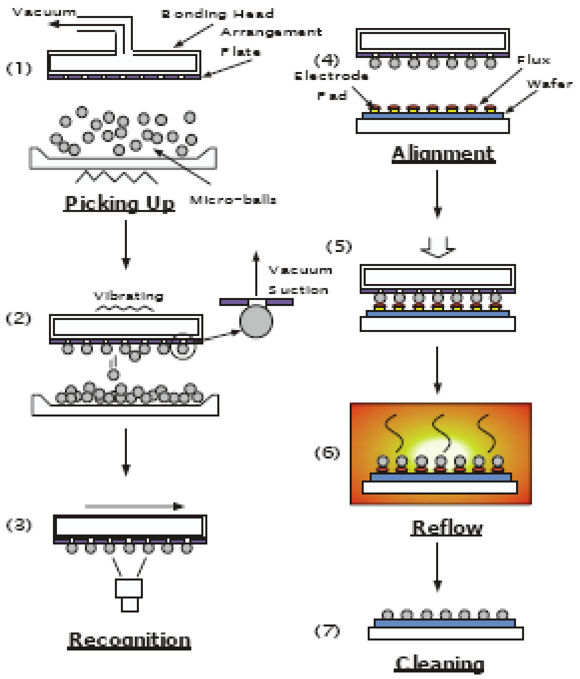


Figure 2.15 Process flow of micro-ball bumping [66]

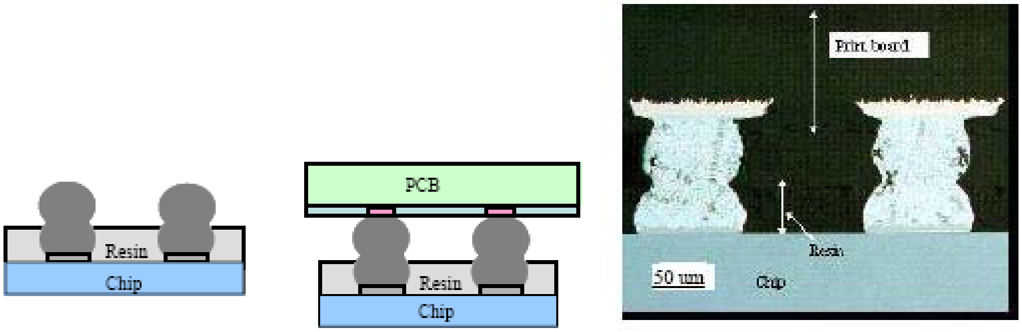


Figure 2.16 Cross-section of the solder joint after assembly on board [66]

2.4.8 Summary

Increasing of the solder joint height using a double bump structure is an effective method to improve the reliability of the package as demonstrated in the literature [52, 61-64], though different implementation approaches to achieve the double bump structure were utilized. Most of these applications do not have underfill fill the gap between the die and the PCB [61,62], or only partially fill the gap closer to the die side [63-68]. This is acceptable for packages with larger solder ball size, such as CSP's. For flip chips, the whole gap between the die and the PCB should be filled with underfill since the CTE mismatch effect increases when the gap height gets smaller.

CHAPTER 3

SOLDER JOINT GEOMETRY PREDICTION BY SURFACE EVOLVER

It is important to design the solder joint geometry because the reliability of the solder joint highly depends on it. Before the experimental implementation of the proposed idea, the double bump solder joint shape was first predicted using software Surface Evolver. The results of the simulation provided a guideline for stencil design utilized in the experiments.

3.1 Review of Solder Joint Prediction Methodology

For a specified solder material and package design, many studies have found solder joint reliability to be highly dependent on the solder joint geometry as characterized by the standoff height, lower/upper contact angles of the joint, solder pad diameter and pad shape, etc [70-76]. Solder joint shape prediction has thus been shown to be a very important tool for use in process development, assembly, and reliability enhancement [74, 75]. The motivation for this study on predicting solder joint shape was to use the predicted results as a guideline for stencil design during the solder bumping process, as well as to generate precise solder joint shapes for double bump flip chips and to provide information that could be imported into finite element modeling software, such as ANSYS for solder joint fatigue analysis.

Current methodologies for predicting solder joint shape can be categorized into three types of algorithms, namely the truncated sphere analytical method, the force-balanced analytical method and the energy based method. All three methods assumed that the final solder joint geometry after reflow can be taken to be the static equilibrium shape of molten solder formed at the junction.

3.1.1 Truncated Sphere Analytical Method

The truncated sphere analytical method is a purely geometrically-based algorithm. There are several different models found in the literature [71, 73, 77], all of which provide a set of equations that describe the relationship between the package physical geometry dimensions, such as solder radius, standoff height, pad radius etc., and the volume computation based on these parameters.

The model developed by Chiang and Yuan is illustrated in Figure 3.1 and Figure 3.2 [73]. This model includes two sets of equations that can be utilized for a stand alone solder ball on the die and a solder joint between the die and the board. In both cases, the height of the solder ball, h, and the radius of the solder ball, R, must be calculated first in order to compute the volume of the solder joint using equations (3-1) to (3-6) [73].

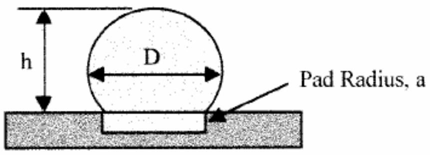


Fig. 1. Solder ball cross section view.

Figure 3.1 Schematic for truncated sphere model [73]

$$h = R + \sqrt{R^2 - a^2} \tag{3-1}$$

$$R = R + \frac{h^2 + a^2}{2h} \tag{3-2}$$

$$V = \pi\left(\frac{h^3}{6} + \frac{h \cdot a^2}{2}\right) \tag{3-3}$$

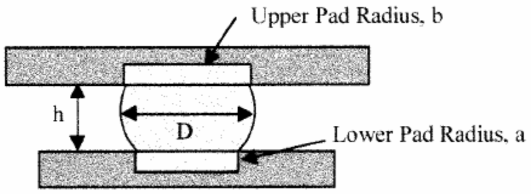


Fig. 2. Solder ball on the board.

Figure 3.2 Schematic for truncated sphere model on board [73]

$$h = \sqrt{R^2 - a^2} + \sqrt{R^2 - b^2} \tag{3-4}$$

$$R = \frac{\sqrt{h^4 + 2h^2(a^2 + b^2) + (a+b)^2(a-b)^2}}{2h} \tag{3-5}$$

$$V = \frac{\pi}{3}\left[\sqrt{R^2 - b^2}(2R^2 + b^2) + \sqrt{R^2 - a^2}(2R^2 + a^2)\right] \tag{3-6}$$

where R is the radius of the solder joint and is equal to D/2. All the other notations are as illustrated in Figure 3.1 and Figure 3.2. The truncated sphere model provides accurate predictions of solder shape within an limited range, but is only suitable for a light weight package [73], because it ignores the effect of gravity on the package.

3.1.2 Force-balanced Analytical Model

The force-balanced analytical solution proposed by Heinrich, and coworker [74, 75] used a Laplace equation to calculate the contours of the solder surface. This model theoretically provides more accurate predictions but is limited by several of its initial assumptions. These assumptions include that the solder joint surface arc remains a perfect circular curve, the gravity density of the solder joint is neglected and the upper and lower pad sizes are similar [74]. The volume and the force are expressed in equations (3-7) and (3-8) [74].

$$V = \pi \int_0^h (R)^2 dz = \pi \int_0^h [R_{arc}(z)]^2 dz \qquad (3\text{-}7)$$

$$\delta F = F_h - \frac{\pi R_h}{2hR_{arc}} \times \left[\mp (R_0 + R_h) - \sqrt{\frac{4R_{arc}^2 h^2}{(R_0 + R_h)^2 + h^2} - h^2} \right] \qquad (3\text{-}8)$$

where V is the solder volume, R(z) is the meridian radius of the arc of the solder surface. F is the unbalanced force on solder joint, and h is the standoff height.

3.1.3 Energy Based Method

The energy based method predicts the solder joint surface by minimizing the surface energy. The implementation of this method utilize a free public software package, Surface Evolver, which was developed by Brakke [78]. Surface Evolver is a computer program that evolves the surface towards an equilibrium shape by minimizing its energy while subjecting it to a set of user-defined constraints. This minimized energy is made up of surface tension, gravitational energy, and energy related to external forces, as indicated in equation (3-9). The solution for the equilibrium state of the solder joint is given when equation (3-10) is satisfied, which implied that the forces due to three energy sources are in balance.

$$E = E_s + E_G + E_{ex} = \iint_s TdA + \iiint_V \rho gzdV - PV \tag{3-9}$$

$$\delta E = 0 \tag{3-10}$$

E is the total energy, E_s is the surface tension energy on the solder surface, E_G is the gravitational energy of the solder, E_{ex} is the energy associated with the external forces, T is the surface tension, A is the area of a facet, ρ is the density of the solder material, g is the gravitational constant, z is the height of a facet, P is the net pressure in the molten solder, and V is the volume of the solder joint.

This method is reported to provide the most accurate predictions compared to experimental results [70, 71, 76]. Unlike the truncated sphere method and the force balanced method, the Surface Evolver method is not limited to applications with a regular shape for the solder joint, but can be applied to all kinds of geometries, which makes it the most widely accepted method for solder joint prediction [71, 79].

In this chapter, the energy based method implemented using Surface Evolver was utilized to predict the shape and height of the second layer solder bump in a double bump flip chip after the bumping process. The effects of the stencil design were examined in terms of the deposited solder volume, and the preferred stencil designs were selected based on the predicted solder joint shape on the board. The self-alignment effect was studied by calculating the restoring forces due to misalignment between the solder joint and pad.

3.2 Model Development

The original geometry was constructed using series definitions of vertex, lines, facets and bodies. The surface in Surface Evolver is actually not a smooth surface, but is divided into groups of triangles called "facets". Each facet is formed by "edges" that connect "vertices". The accuracy of the results depends on how fine the mesh of triangles is; the more triangles used for a single surface, the finer the mesh. Usually, every iteration action in the calculation split one facet into four smaller triangles. As the surface evolves, all the facets are refined to form a more accurate surface based on the smallest energy. User defined constraints determine the final shape of the solder, as well as material constants such as surface tension and density. These constraints can be geometrical constrains of vertex positions or constraints of integrated quantities such as body volumes. The material constants used in this study are listed in Table 3.1.

Table 3.1 Parameters used to calculate flip chip solder joint height

Solder Density	8.510 grams/cm^3
Solder Surface Tension	481 dyne/cm
Gravity Constant	980.665 cm/s^2
Die Pad Diameter (UBM)	0.0102 cm
PCB Pad Diameter	0.0114 cm

In this study for a solder joint between circular pads, a hexagonal prism was used as the starting shape. The applied constraints include: the fixed bottom and top pad, the volume of the body. Other parameter determining the solder ball shape was the bumping site, which is the exposed original solder ball tops and characterized by diameter as shown in Figure 3.3. The rim of the bumping site was also implemented as one of the constraints in Evolver program. The height of the 2nd solder ball was set as the optimized parameter in the script for both double bump die shape prediction and double bump die on the board prediction. In both predictions, 4 refinement steps and 30 iterations were used to generate the solder bump surface.

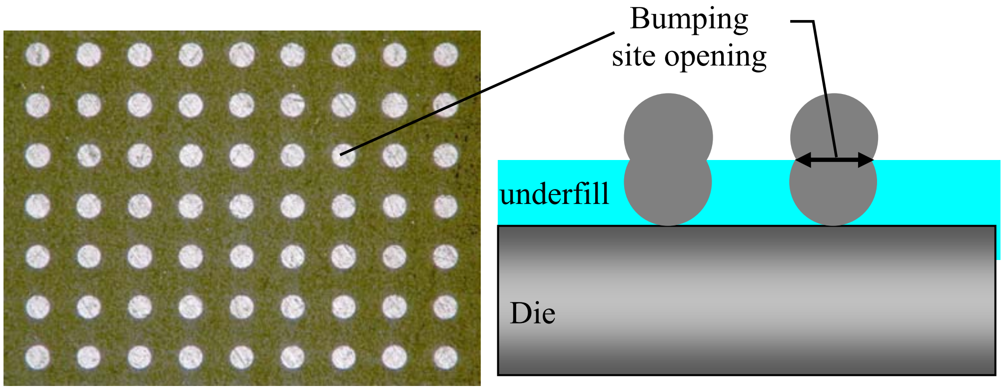


Figure 3.3 Bumping site (a) top view of the polished wafer surface with exposed solder ball top as bumping site (b) schematic side view showing bumping site opening

In the prediction for the stand alone solder balls with double bump geometry, only the second bumped solder ball was constructed using the model, as shown in Figure 3.4, and the original solder bump in the first layer was omitted for simplicity. The original solder bump maintained its original shape and thus had no influence on the shape of the second layer solder balls, because the original solder ball was secured in the cured underfill during the bumping process, and no gravity would affect the solder in the original solder bump, since the die was facing up during reflow.

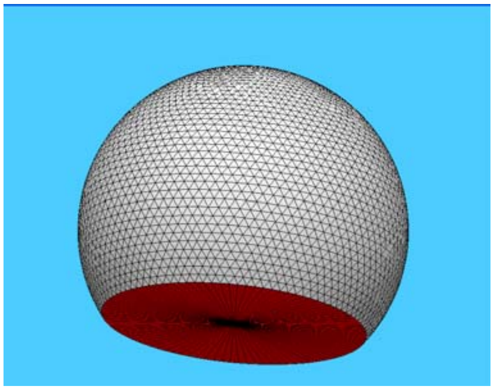


Figure 3.4 The appearance of the predicted 2nd solder bump shape after the bumping process, displayed at an angle to show the 3D effect

In the prediction of the double bump flip chip on the board, both the original solder ball and the second solder ball were constructed using the model, as shown in Figure 3.5. Note that the two solder balls must be defined as a single body in order to predict the shrinkage problem as explained in Section 3.4.

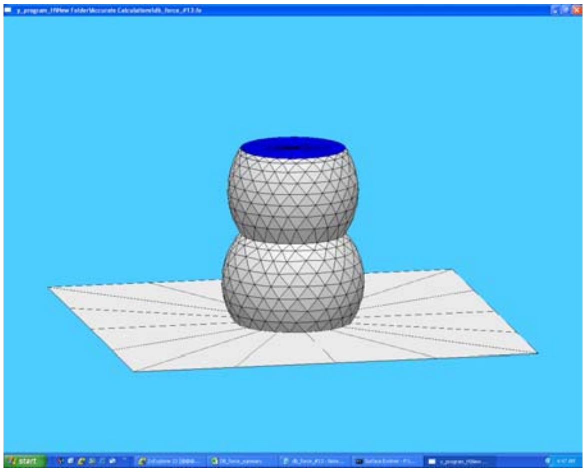


Figure 3.5 The appearance of the predicted shape of the double bump flip chip on the board using Surface Evolver, displayed at an angle to show the 3D effect

3.3 Stencil Designs

In addition to all the material constants, the parameters that determined the second layer bumped solder ball shape are the solder paste volume that is deposited on the coated wafer and the bumping site opening diameter. The deposited volume is directly related to the aperture in the stencil design.

Chemical etching, laser cutting and electroforming are the three most commonly used technique for stencil fabrication. Electroformed stencils provide the best solder release and is the most suitable one for fine feature printing [80]. Therefore the electroformed stencil was selected for use in the solder ball bumping process in this research. The other important factors considered in the stencil design were stencil thickness and the aperture shape and dimensions. Table 4.2 lists the generally adopted rules for stencil design with different stencil types recommended by Cookson Electronics [81].

Table 3.2 Stencil design guidelines for different stencil type [81]

Stencil Type	Ratio of Minimum Aperture Width to Foil Thickness
Chemically Etched	≥ 1.5 : 1
Laser-Cut	≥ 1.3 : 1
Electroformed	≥ 1.1 : 1

For fine aperture sizes, calculating the area ratio is important in stencil design. The area ratio is defined in equation (3-11):

$$AreaRatio = \frac{A_p}{A_w} \qquad (3\text{-}11)$$

where A_p is the area of the aperture opening, and A_w is the area of the aperture wall. A larger area ratio is desired for better transfer efficiency. Area ratios less than 66% tend to not release well, i.e., paste remains on the aperture wall rather than transferring to the die [81]. However, in flip chip wafer bumping applications, the space between the adjacent apertures also needs to be considered to avoid bridging problems.

After considering all of the above factors and careful calculations, the stencil thickness was selected to be 3mil. Six types of aperture opening designs were investigated in this study. These six designs include two aperture shapes, square and round, each of which was modeled for three different sizes from 6.0mil to 7.0mil in terms of diameter or side. The aperture volume and area ratio of each stencil design are listed in Table 3.3. Here all the designs have an area ratio lower than the 66%. To increase the area ratio, the stencil thickness could be reduced, but the aperture opening would have to be increased in order to maintain the desired solder ball volume after reflow. Considering the 10mil pitch of the bumps, a 7.0mil opening diameter leaves only 3mil space in between, which was already close to the limit for good bumping and further reducing this space would increase the risk of solder bridging after reflow.

Table 3.3 Stencil aperture parameters and calculated deposited volume

Aperture Type	Aperture Size (mil)	Area Ratio	V (aperture) (mil^3)	V (solder ball) (mil^3)
Square Apertures	6.0	50.00%	108.00	54.000
	6.5	54.17%	126.75	63.375
	7.0	58.33%	147.00	73.500
Round Apertures	6.0	50.00%	84.82	42.412
	6.5	54.17%	99.55	49.775
	7.0	58.33%	115.45.	57.727

As shown in Figure 3.6, a larger aperture size provides a higher area ratio, and the area ratio was the same for square apertures and round apertures of the same size. The volume of the solder ball after reflow was calculated as being half of the deposited solder paste volume, because the flux and solvent (50% of the paste volume) in the solder paste evaporate during reflow. Square apertures transferred more solder paste than round apertures of the same size, thus producing a larger solder bump.

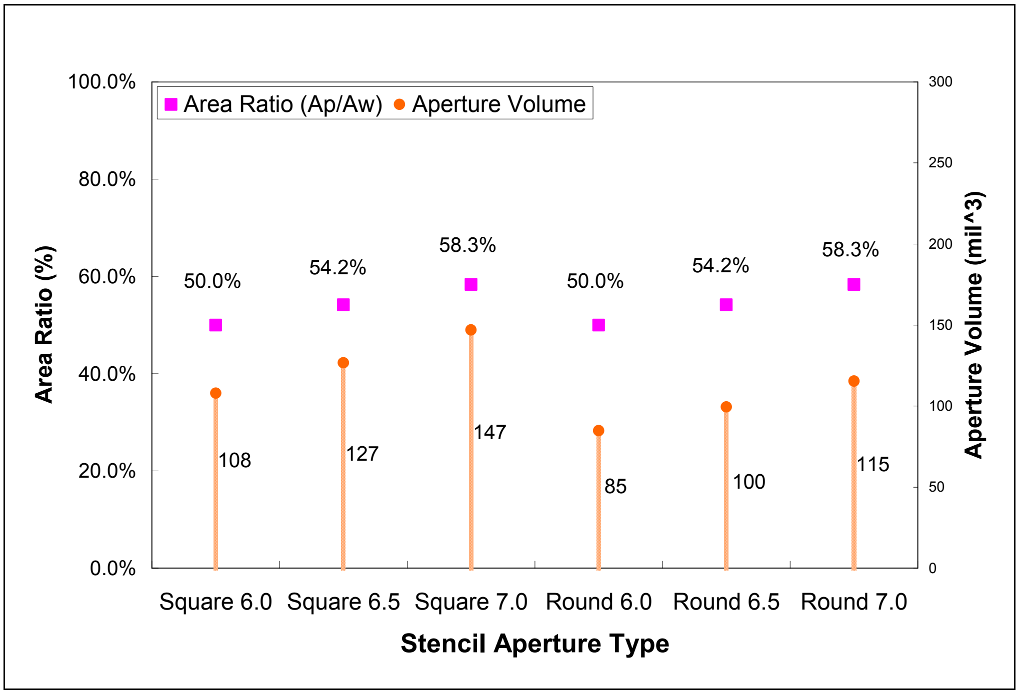


Figure 3.6 Area ratio and aperture volume for different stencil designs

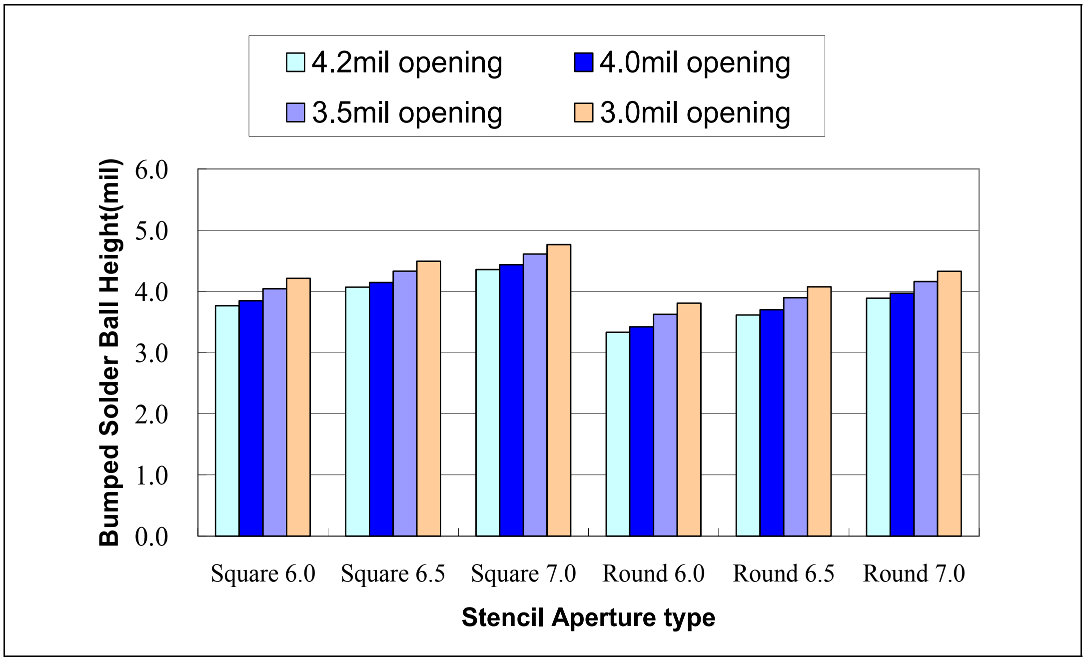
3.4 Shape Prediction of Standalone Double Bump Solder Ball After the Bumping Process

The first application of Surface Evolver simulation was to predict the double bumped solder ball shape after the bumping process. The calculated solder ball volume listed in Table 3.3 was imported into the Evolver model as the input for predicting the corresponding solder joint radius and height after reflow. The height of the second solder ball was given an initial value of 3.15mil but was set as an optimizing parameter in the program, which allowed it to change during the computation in order to achieve a minimized energy.

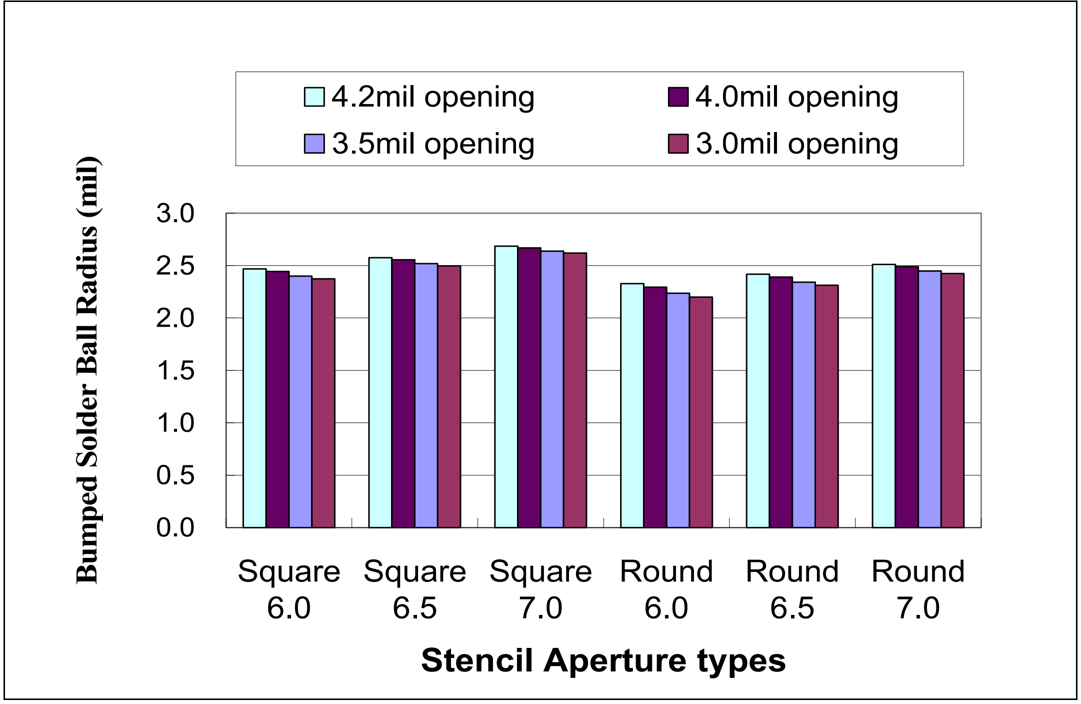
All combinations for the six different stencil designs and four different bumping site openings were calculated and the simulation results are listed in Table 3.4. As expected, the smaller the bumping site opening, the higher the solder ball after the bumping process, as shown in Figure 3.7(a), and vice versa for the solder ball radius, as shown in Figure 3.7(b).

Table 3.4 Prediction of double bumped solder ball shape after reflow for different stencil aperture designs

Stencil			Prediction (Bumping Site DIA=4.2mil)		Prediction (Bumping Site DIA=4.0mil)		Prediction (Bumping Site DIA=3.5mil)		Prediction (Bumping Site DIA=3.0mil)	
Aperture Type	**V(ap) (mil^3)**	**V(Solder) (mil^3)**	**h (mil)**	**R (mil)**	**h (mil)**	**R (mil)**	**h (mil)**	**R (mil)**	**h (mil)**	**R (mil)**
Square 6.0	108.000	54.000	3.7642	2.468	3.8476	2.444	4.0413	2.400	4.2116	2.373
Square 6.5	126.750	63.375	4.0664	2.575	4.1460	2.555	4.3310	2.519	4.4931	2.497
Square 7.0	147.000	73.500	4.3573	2.685	4.4330	2.668	4.6107	2.637	4.7653	2.619
Round 6.0	84.823	42.412	3.3304	2.327	3.4190	2.294	3.6260	2.235	3.8095	2.200
Round 6.5	99.549	49.775	3.6149	2.417	3.7002	2.391	3.8980	2.342	4.0730	2.313
Round 7.0	115.454	57.727	3.8888	2.511	3.9707	2.489	4.1607	2.448	4.3275	2.424



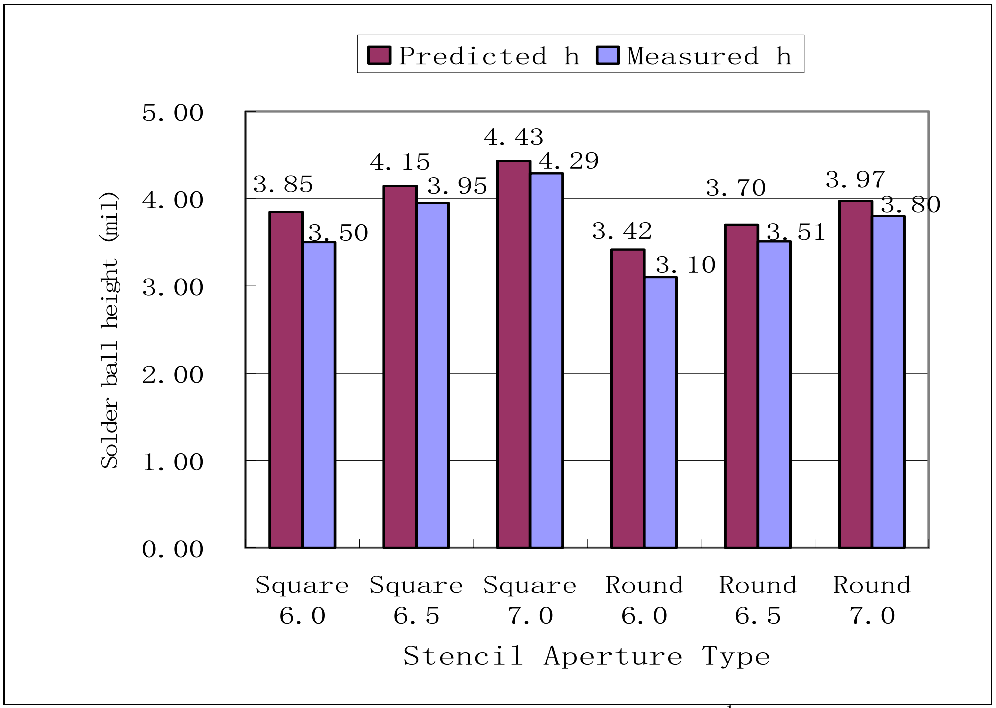
(a) Predicted solder ball height



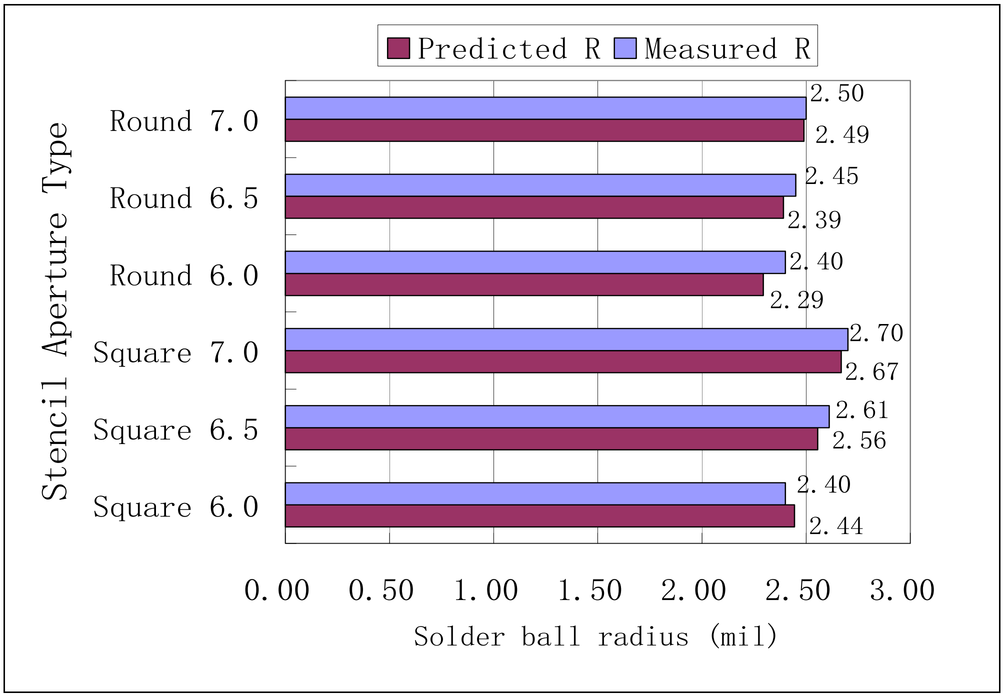
(b) Predicted solder ball radius

Figure 3.7 Predicted solder ball shape for different stencil aperture types

The predicted results were compared with the measured results for a bumping site equal to 4.0mil, as shown in Figure 3.8. Both measurements of the second solder ball height and radius were done with a measuring microscope system. Calibration of radius measurement was performed by measuring the pitch (solder ball center to solder ball center distance) prior to measurement of each bumped wafer. As shown in Figure 3.8, the measured solder ball height was slightly lower than the predicted value for all stencil types, while the radius of the solder ball was close to the predicted value.



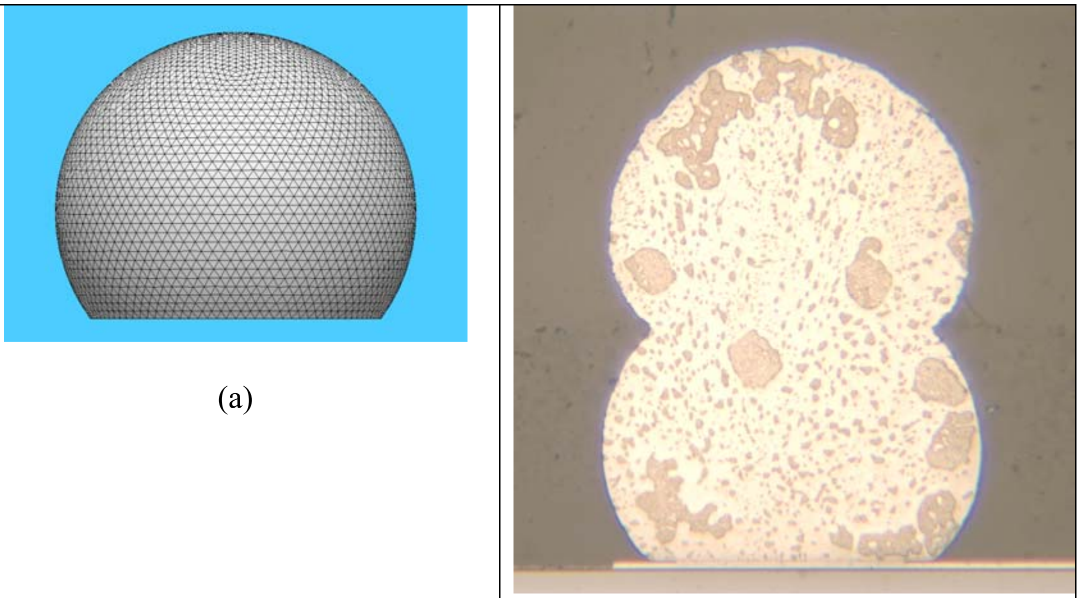
(a) Predicted value and measured value of the 2nd solder ball height



(b) Predicted value and measured value of the solder ball radius

Figure 3.8 Predicted value and measured value of the solder ball height and radius after bumping process for different stencil types (bumping site diameter=4.0mil)

Figure 3.9 (a) shows an example of a predicted solder bump shape, and Figure 3.9 (b) shows the cross-sectional picture of a double bump flip chip die obtained experimentally. The image indicates a close fit with the predicted shape.


(a)

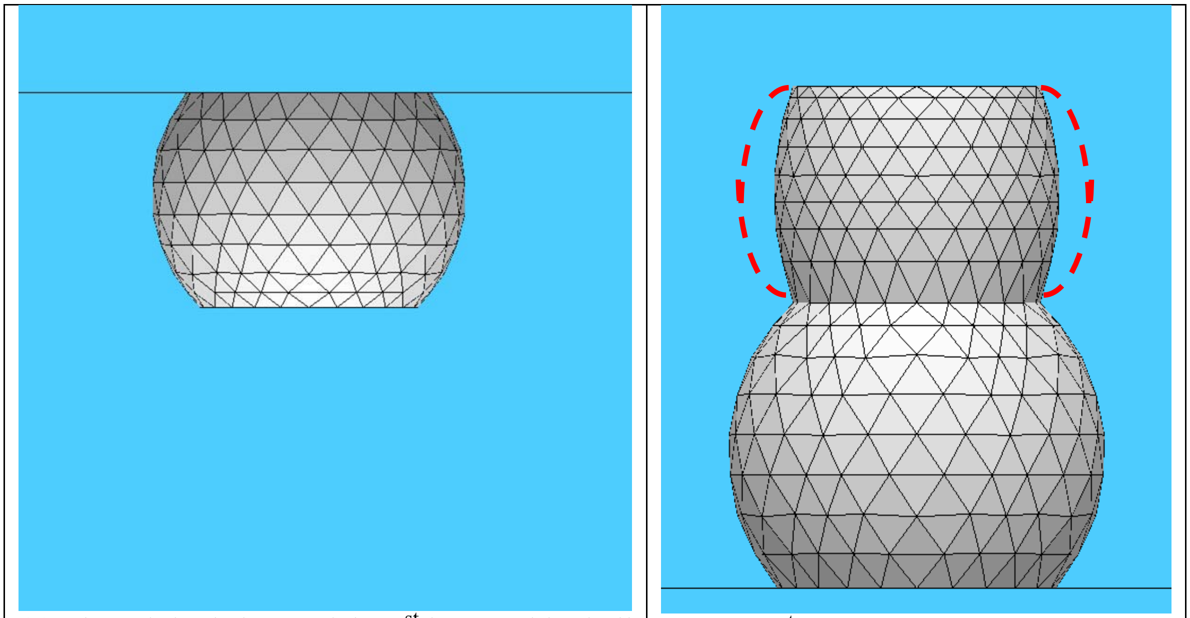
(b)

Figure 3.9 (a) The appearance of the solder bump shape predicted using Surface Evolver, (b) the experimentally obtained cross-sectional image of a flip chip die with double bump structure

3.5 Shape Prediction of Double Bump Flip Chip on Board

Shape prediction for a double bump flip chip mounted on a board after the assembly process was conducted in order to predict the standoff height of the double bump flip chip. The prediction results also provided a guideline for stencil design selection, and were imported into the ANSYS program for a finite element analysis of the solder joint fatigue with a precise geometry definition.

The criteria used to select the stencil designs was to avoid the "shrinkage" problem of to the original solder ball (die side) as illustrated in Figure 3.10. Figure 3.10 (a) shows the original shape of the solder ball in the first layer (die side), which is a perfect double truncated sphere shape. However, after bumping a second solder ball onto the original solder ball, the original solder ball at the die side shrinks during reflow assembly to the substrate due to the presence of the second solder ball. Solder flow from the upper ball (die side) to the lower ball (board side) when melted under the force of gravity and surface tension creates a smaller than original upper ball and a larger lower ball shape. This shrinkage creates a gap between the solder ball and the cured underfill surrounding it, as shown in the Figure 3.10 (b) where the dashed line represents the underfill interface. This cavity reduces the underfill protection for the solder ball and degrades the reliability of the package. Thus, stencil design considerations must take this into account to avoid any shrinkage of the upper solder ball.

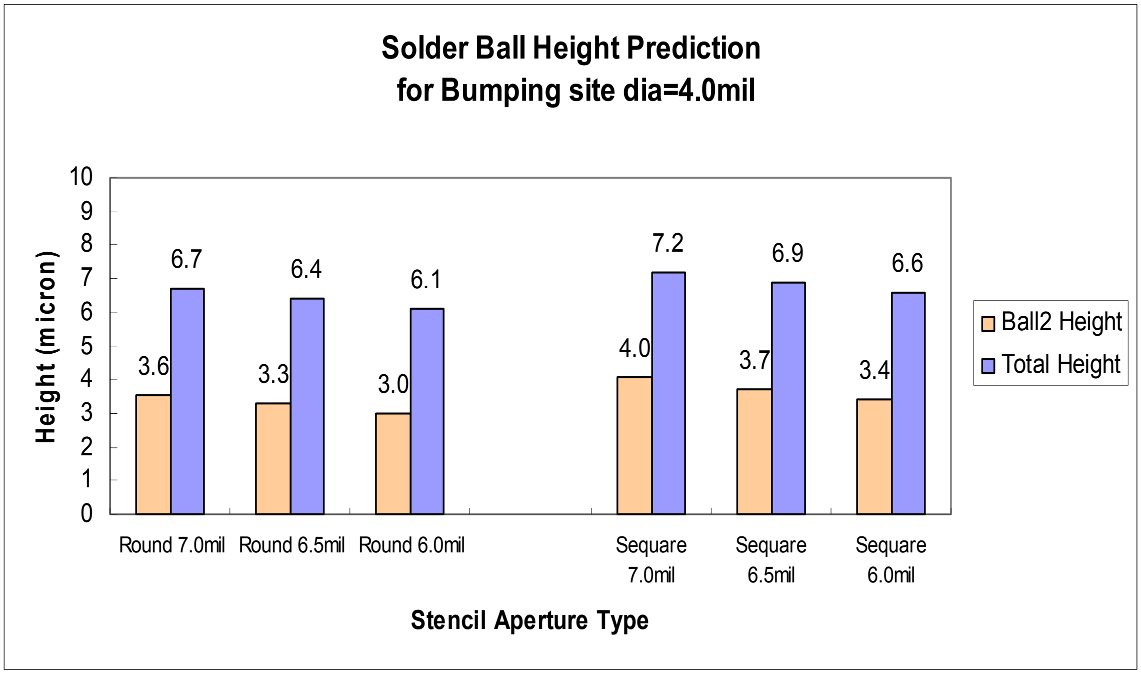


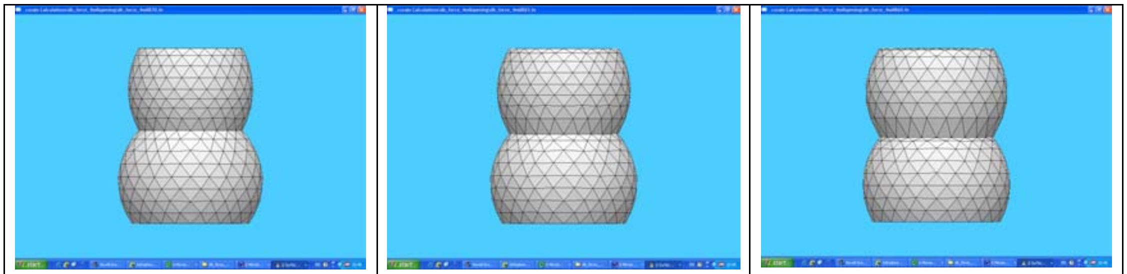
(a) The original shape of the 1st layer solder ball (die side) before the second ball is bumped onto it

(b) The 1st layer solder ball shrinks with a 2nd solder ball present during assembly

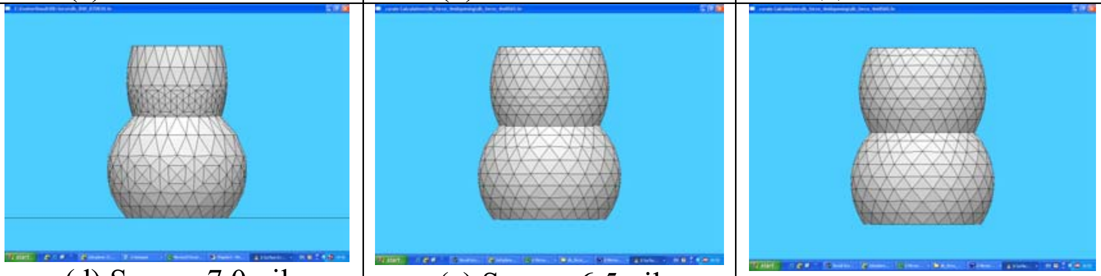
Figure 3.10 The shrinkage problem for the original solder ball during the assembly process

Predicted double bumped solder ball height for all stencil types with 3 different bumping sites were plotted with respect to the total solder ball height (two stacked solder balls) in Figure 3.11 to Figure 3.16. The predicted shape of the double bump solder balls are also shown from a side view. Applying the shrinkage criteria, Table 3.5 summarizes the acceptable stencil designs for three bumping site opening diameters. None of the stencil designs with square apertures qualified for the double bump process, due to shrinkage of the original solder ball. The preferred stencil designs were round apertures with 6.0mil and 6.5mil diameters for both the 4.0mil and 3.5mil bumping site openings, and round apertures with a 6.0mil diameter for the 3.0mil, with the best choices being round apertures with 6.0mil and 6.5mil diameter for a 4.0mil bumping site opening.



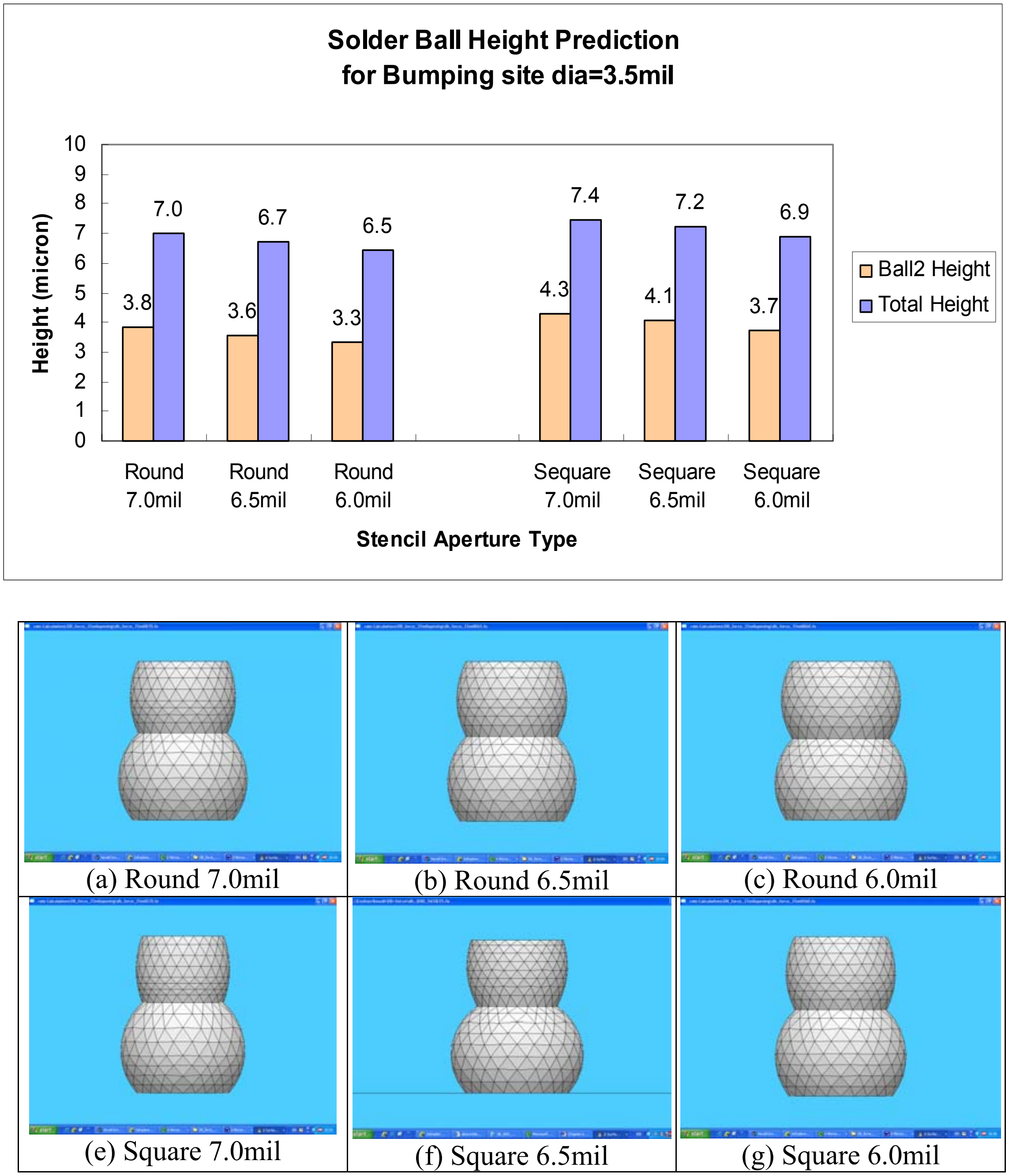


(a) Round 7.0mil (b) Round 6.5mil (c) Round 6.0mil



(d) Square 7.0mil (e) Square 6.5mil (f) Square 6.0mil

Figure 3.11 Solder joint shape and height prediction for different stencil aperture types with a bumping site opening diameter =4.0mil



(a) Round 7.0mil

(b) Round 6.5mil

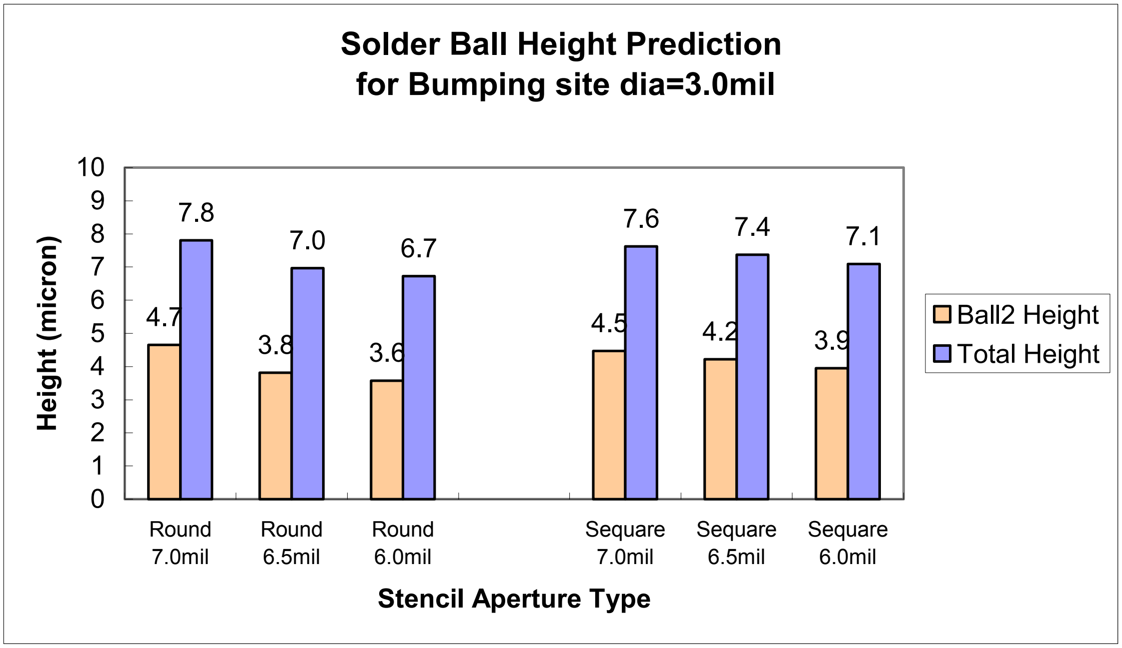
(c) Round 6.0mil

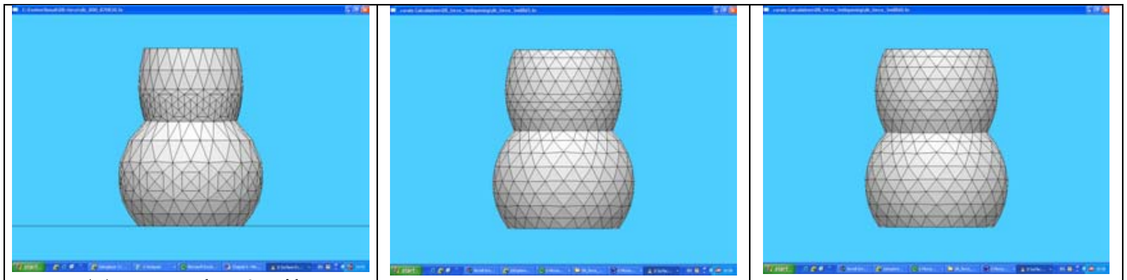
(e) Square 7.0mil

(f) Square 6.5mil

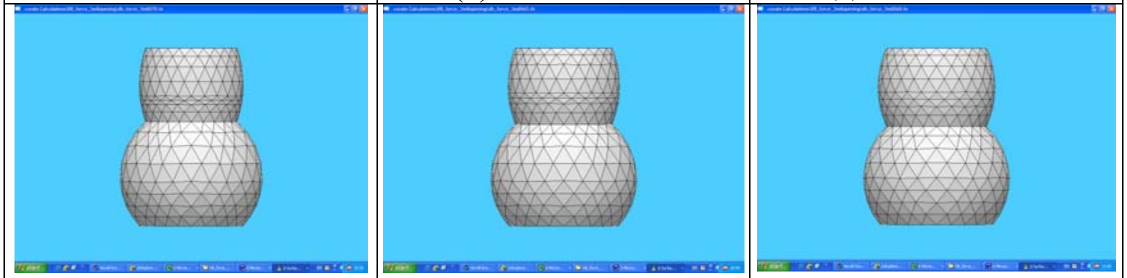
(g) Square 6.0mil

Figure 3.12 Solder joint shape and height prediction for different stencil aperture types with a bumping site opening diameter =3.5mil





(a) Round 7.0mil (b) Round 6.5mil (c) Round 6.0mil



(d) Square 7.0mil (e) Square 6.5mil (f) Square 6.0mil

Figure 3.13 Solder joint shape and height prediction for different stencil aperture types with a bumping site opening diameter =3.0mil

Table 3.5 Preferred stencil designs for three different bumping site opening diameters

Site opening diameter / Aperture Types	DIA=4.0mil	DIA=3.5mil	DIA=3.0mil
Round 7.0mil	shrinkage	shrinkage	shrinkage
Round 6.5mil	3.3mil/6.4mil	3.6mil/6.7mil	shrinkage
Round 6.0mil	3.0mil/6.1mil	3.3mil/6.5mil	3.6mil/6.7mil
Square 7.0mil	shrinkage	shrinkage	shrinkage
Square 6.5mil	shrinkage	shrinkage	shrinkage
Square 6.0mil	shrinkage	shrinkage	shrinkage

* The cell represented for preferable design was marked in shade and filled in terms of solder ball height (2nd solder ball height/total solder ball height).

3.6 Shape Prediction of the Restoring Force for a Double Bumped Flip Chip Mounted on a Board

Solder balls on a regular flip chip have the ability to self-center to the die, which means the component self-aligns to the pad during reflow even if they are misplaced during the placement process, as shown in Figure 3.14. This is due to the force, namely restoring force, generated from the surface tension of the molten solder ball during reflow. The restoring force is opposite to the direction of displacement, and it always tries to pull the solder joint back to the self-centered position.

The Evolver program can be used to calculate the restoring force due to the misalignment. The original Evolver program used in the previous shape prediction was modified to imply the misalignment by offsetting the lower pad constraint location. Parameter y was defined as the offset due to this misalignment. After achieving a steady state with the desired constraints, the parameter y is changed to y + δy, evolved to a minimum of energy, then changed to y- δy and evolved again. These two achieved minimum energy values were used to calculate the restoring force utilizing a central difference formula as:

$$F_y = -\frac{E(y+\delta y) - E(y-\delta y)}{2\delta y} \tag{3-12}$$

where E(y) is the minimum energy for offset value y. δy was chosen to be $1e^{-5}$ in the restoring force calculation.

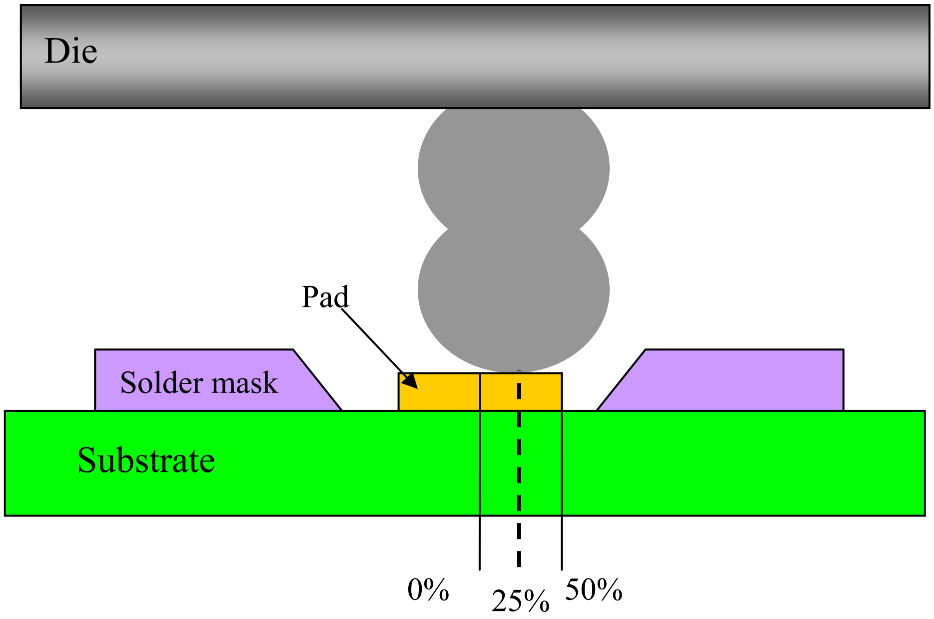


Figure 3.14 Definition of offset percentage

As shown in Figure 3.15, when the offset increases, the restoring force initially increases. This indicates that the further the solder joint is misaligned from the initial position, the larger the force pulling the solder joint back into the self-centered position. When the offset is greater than 38.9% of the solder diameter, this restoring force starts to increase slowly. The maximum restoring force occurs at about 44.4% offset, which is 2mil displacement and the restoring force at this position is about 0.008 dynes. Figure 3.16 shows the corresponding solder joint shapes for different offset.

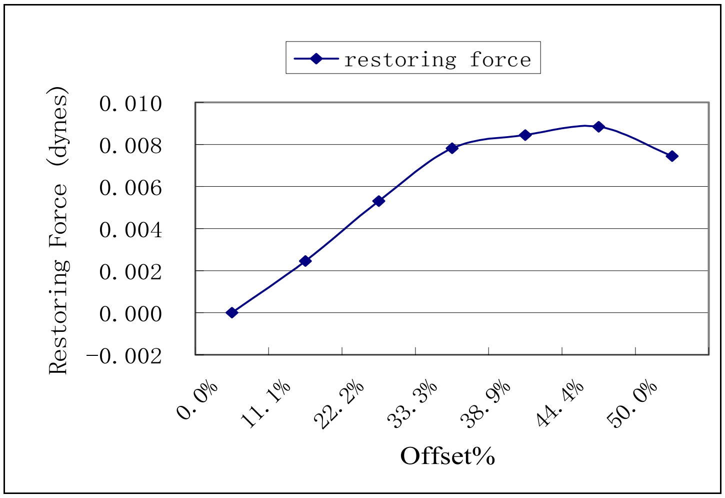
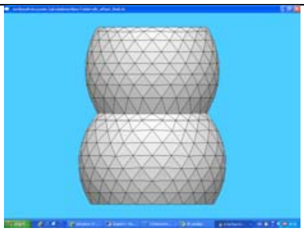
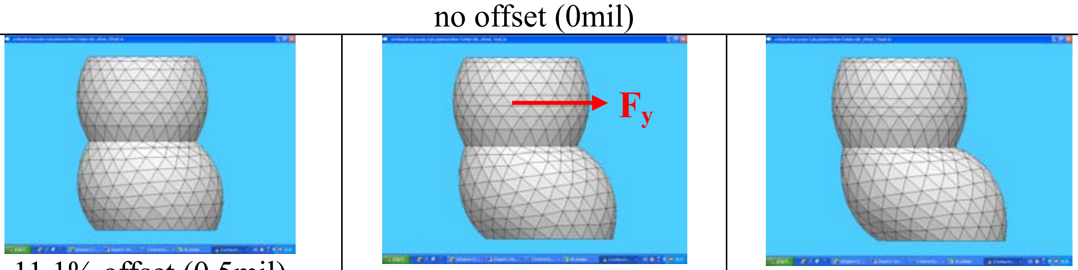


Figure 3.15 Restoring force vs. offset % for double bump flip chip



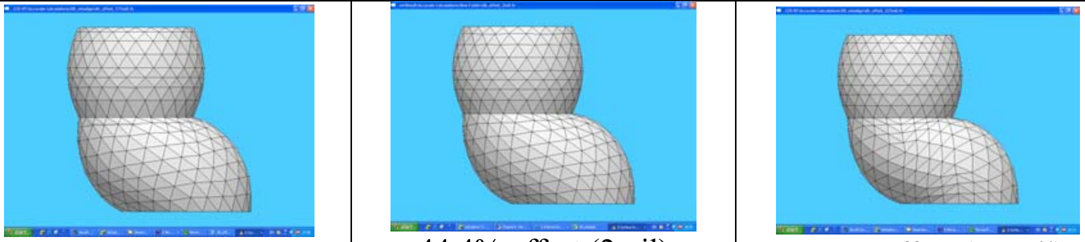
no offset (0mil)



11.1% offset (0.5mil)

22.2% offset (1mil)

33.3% offset (1mil)



38.9% (1.75mil)

44.4% offset (2mil)

50.0% offset (2.5mil)

Figure 3.16 Solder joint cross-sectional shape for different offset % with 4.5mil pad diameter and 5mil solder joint diameter

CHAPTER 4

EXPERIMENTAL PROCEDURES AND RESULTS

4.1 Double Bump Flip Chip Die Fabrication

In order to produce a higher standoff for the existing solder joint, a new process was developed to fabricate a flip chip with a double bump solder joint structure. Figure 4.1 illustrates the steps for this process. An underfill layer with low CTE was applied at the wafer level and cured. The encapsulated wafer was then polished to expose the original solder ball top, which served as the bumping site for the second layer of solder balls. The second layer of solder balls were formed on the first layer of solder balls using the stencil print technique, followed by reflow. Finally, the double bumped wafer was cleaned and diced into individual dies for later use. More details of the process are given in this chapter.

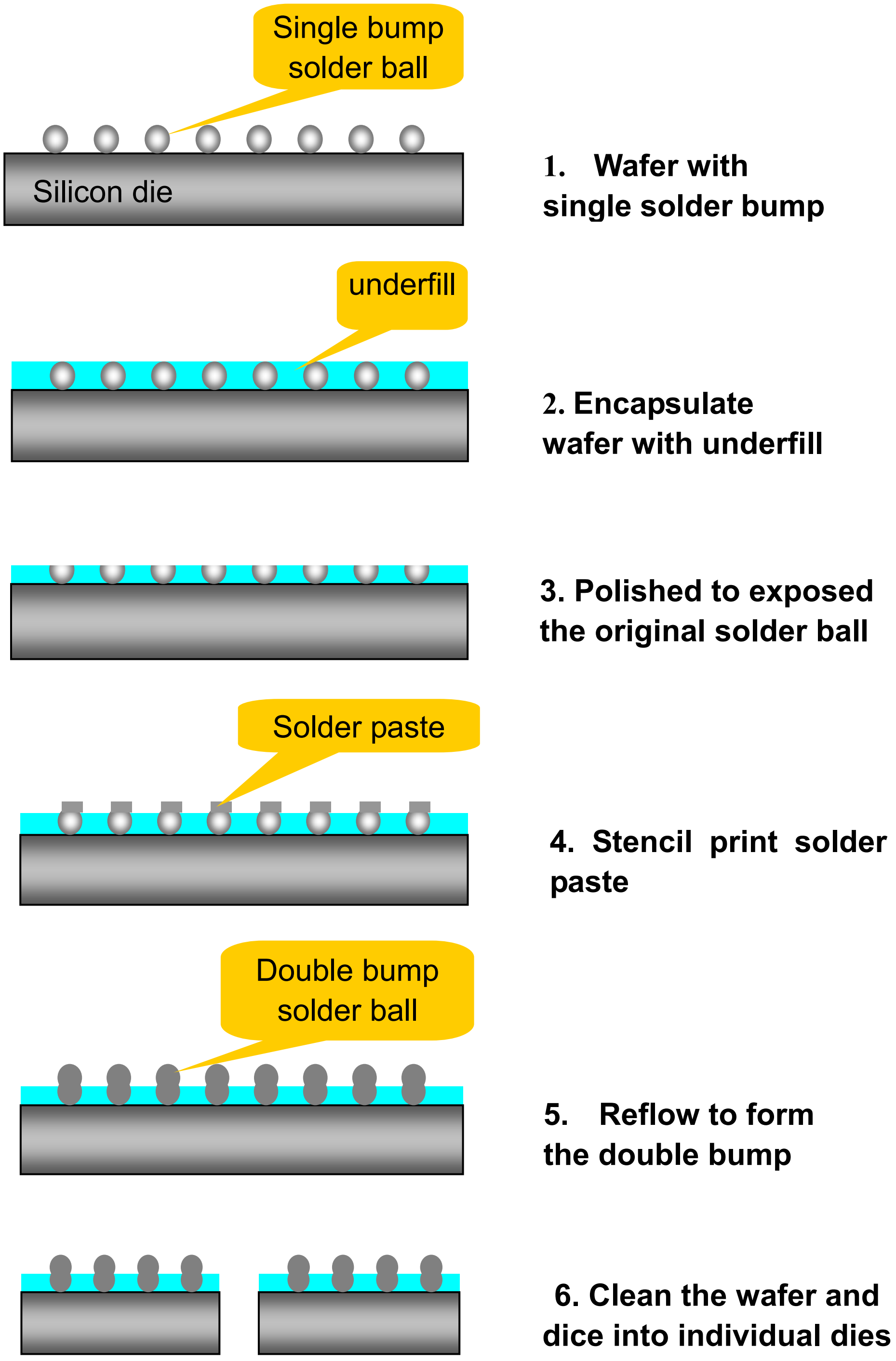


Figure 4.1 Illustration of the double bump flip chip die fabrication procedure

4.1.1 Test Vehicles

The FA10 flip chip test wafer from Delphi was utilized as the platform in this study. The test chip dimensions were 200x200mil (shown in Figure 4.2), with 317 bumps in an area array. The solder ball material was eutectic SnPb solder with a melting temperature of 183℃. The daisy chain circuit on the test chip was used to check the integrity of the solder bumps after the chip was assembled to the printed circuit board.

Due to cost considerations, wafer squares were used for all tests rather than the whole wafer. The wafer squares were obtained by dicing the original 5 inch wafer into 800x800mil squares, each containing 16 individual FA10 200x200mil dies, as shown in Figure 4.2.

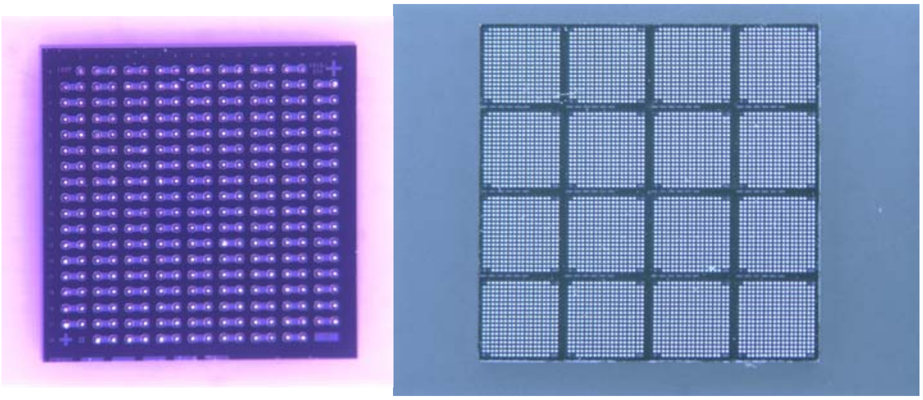


Figure 4.2 Test vehicles: (a) 200x200mil FA10 die (b) 800x800mil wafer square

4.1.2 Coating Process

Lord CircuitSAF® ME-525 underfill was used as the underfill layer applied at the wafer level. ME-525 is a flip chip underfill encapsulant with high fracture toughness and high T_g (Table 4.1). Since this underfill material was originally developed for capillary flow underfill applications, it has both a low viscosity and a low CTE (22ppm/℃) matched with the solder. There are two methods that were investigated to apply the underfill coating. The underfill can be printed onto the wafer surface using a stencil with a rubber squeegee gently sweeping across the solder ball tops. This results in hills and valleys in the surface topology of the underfill layer: valleys between solder balls and hills at the solder balls. Furthermore, in order to achieve a particular coating thickness, the viscosity of the underfill material must on occasion be adjusted to a higher value. The coating can also be achieved by flipping the wafer square down into a layer of underfill that has been dispensed on a Teflon surface with a special tool that holds and compresses the wafer in the underfill during curing. The second method was chosen for this application in order to use the commercialized underfill ME-525 without any material development. After coated onto the wafer, the underfill was cured for 30 minutes at 150℃. Figure 4.3 shows the top surface after curing of the wafer coated using second method. The small white dots are the original solder ball tips.

Table 4.1　Properties of the underfill used for the wafer level underfill layer

T_g	Modulus (E)	Viscosity	CTE
150℃	4GPa	6,000 MPa.s @25℃	22ppm/℃

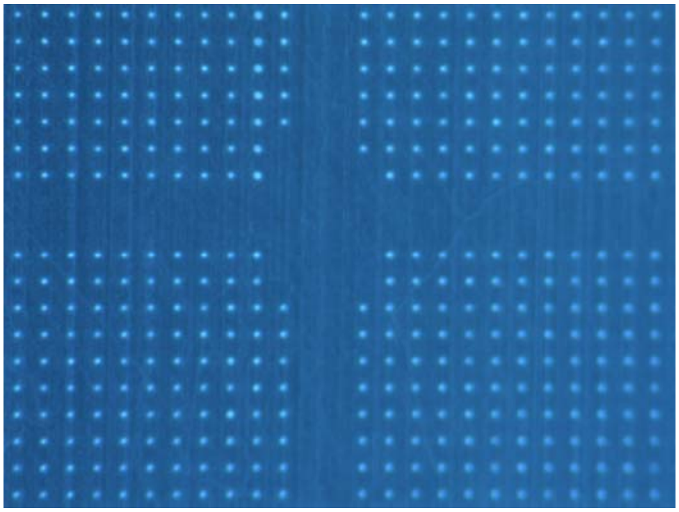


Figure 4.3　Wafer surface after encapsulation with the underfill layer

4.1.3　Polishing Process

In order to bump the second layer of solder balls onto the coated wafer, a bumping site must be provided in order to achieve reliable joining of the first and second layer of balls. The original solder ball tops thus served as the bumping sites in this application. As shown in Figure 4.3, the original solder ball tops are mostly covered in the underfill with only a very small tip exposed. Mechanical polishing using a grind wheel and grinding paper was performed on the encapsulated wafer. A flat, smooth surface was obtained after step polishing using grinding paper with 400 grit, 600 grit, 800 grit and 1200 grit.

The extent of wafer polishing was found to be a major factor in achieving a successful second bumping. For 5.0mil diameter initial solder balls, the optimized bumping site diameter was found to be in the range of 4.0mil to 4.2mil based on the simulation results using Surface Evolver software (Chapter 3) and validated by the experimental results. Further polishing to increase the exposed solder bump area yields a reduced standoff height as shown in Figure 4.4.

Due to variations in the initial bump height and bump diameter, the exposed area after polishing also varied across the wafer square. Parallelism during polishing is another important factor. Tilting during the polishing stage leads to non-uniform bumping site diameters, thus resulting in non-uniform second solder bump heights which often leads to electrical opens when the chips are assembled onto boards. However, careful control of the polishing process can result in a relative uniform bumping site opening diameter with variation less than 0.3mil across the wafer square. After polishing, the dies must be cleaned and dried thoroughly prior to the solder paste printing process.

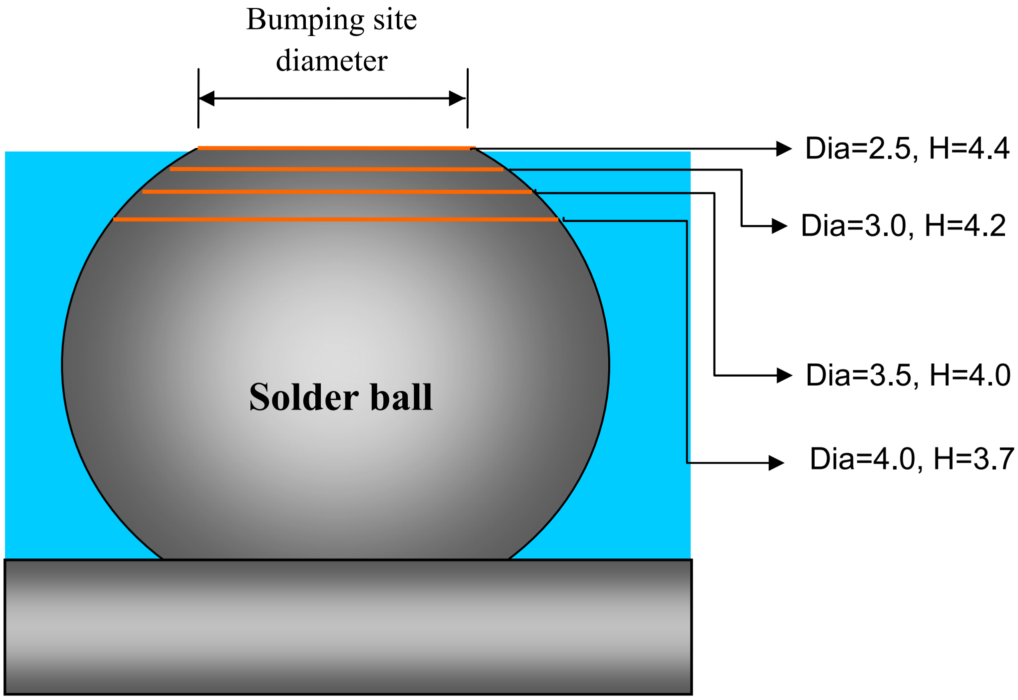


Figure 4.4 Schematic of the cross-section of the coated die showing the exposed area of the bumping site after polishing (all Diameter and H shows in unit: mil)

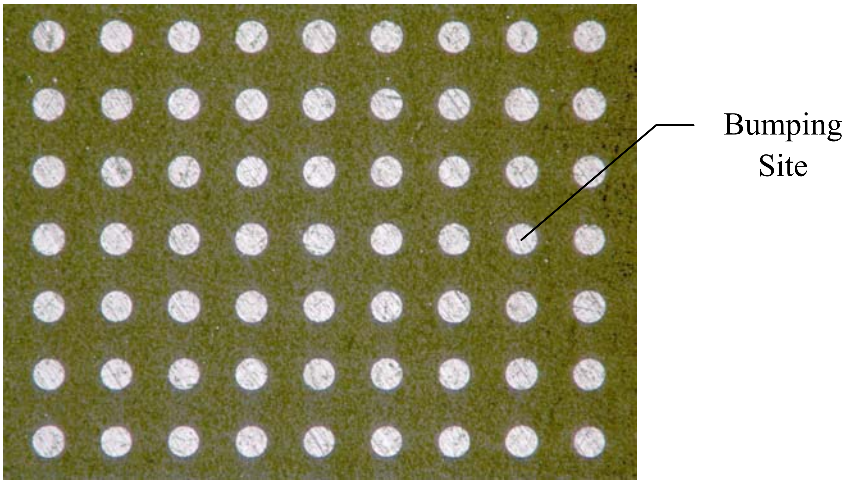


Figure 4.5 Coated wafer surface after polishing with exposed solder ball top served as the bumping site for the 2nd layer solder ball bumping

4.1.4 Second Layer Solder Bump Printing Process

Several processes have been successfully demonstrated for use in flip chip solder ball bumping in the literature, including plating, wire bumping, evaporation, dispensing and printing [82]. Among these techniques, stencil printing was chosen for the solder ball bumping in this application, since it is the simplest bumping method. Recent advances in stencil technology and pastes now allow it to be used in fine pitch applications.

Stencil printing processes typically involve three steps. The solder paste is first pushed into the apertures in the stencil by a squeegee. It then makes contact with the bumping sites on the polished wafer. Finally, the paste is transferred to the wafer while the stencil retracts. The rheology of the paste, the print speed, and the separation between the stencil and the substrate all affect the volume, geometry and uniformity of the printed solder paste.

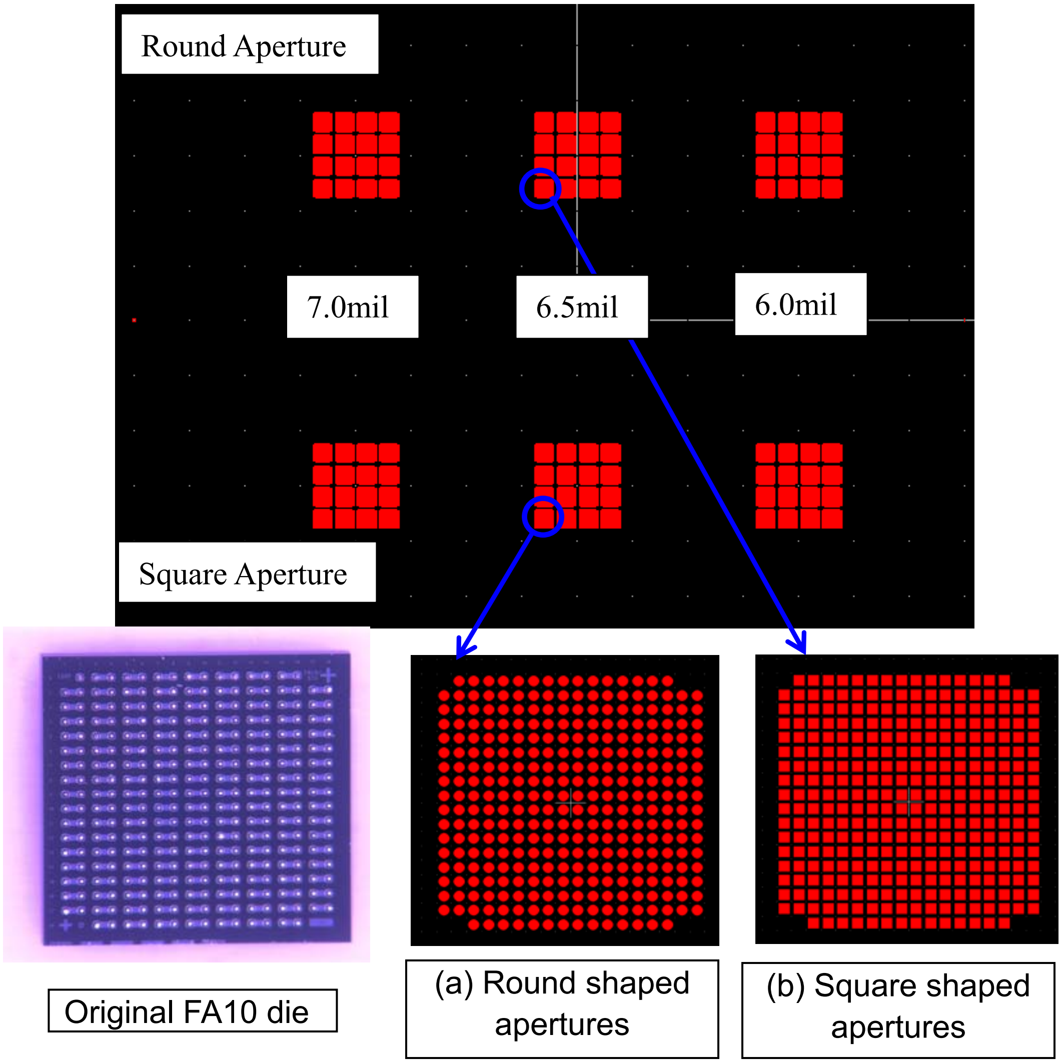
4.1.4.1 Solder Paste

A high grade solder paste was required for this fine feature printing application. A type-6 solder paste with a maximum particle size of 20 um or finer (no-clean) paste was used for the double bump process.

4.1.4.2 Stencil Design

An electroformed stencil was chosen for use in this application because its smooth, tapered aperture walls offer the optimum paste release characteristics. The use of an electroformed stencil maximizes both the solder paste deposit and its ease of release, which is preferred in stencil designs with very fine apertures.

The stencil design is the most critical factor in achieving the desired bump geometries for a given assembly design. The solder volume and bump geometry is mainly determined by the stencil aperture opening and stencil thickness. The aperture pattern of the stencil was designed to match the pattern of an FA10 area array flip chip wafer, as shown in Figure 4.6. Two aperture shapes were investigated: round aperture for three patterns positioning at the upper part of the stencil, and square aperture positioning at the lower part of the stencil. Each aperture shape included 3 different sizes, 7.0mil, 6.5mil and 6.0mil. All these six stencil designs were placed onto one 12”x12” stencil with a thickness of 3.0mil. By adjusting the wafer placement on the work-holder each stencil design could be used independently. The relationships between the final solder ball geometry and bumping site opening size, aperture size and shape were studied extensively using both experimental data and simulations using Surface Evolver software as described in Chapter 3.



Original FA10 die

(a) Round shaped apertures

(b) Square shaped apertures

Figure 4.6 Six stencil patterns on one stencil for double bumping process with apertures matching the FA10 die

All six stencil designs were tested to evaluate the printing quality. Based on the simulation results in Chapter 3 (Table 3.5), Table 4.2 lists the preferred stencil designs determined by simulation. Round 6.0mil diameter and 6.5mil diameter apertures should produce satisfactory double bump shapes. The yield was found to be poor for the 6.0mil round aperture and the bumped solder balls were not uniform in size under the microscope inspection. The round 6.5mil apertures showed much better printing results. This difference in printing quality is related to the different area ratio of the two stencil aperture designs. The round 6.5mil apertures have a higher area ratio compared with that of the round 6.0mil apertures. As explained in Chapter 3.3, the higher the area ratio, the better the transfer efficiency is. With the 6.0mil round apertures, a certain amount of solder paste adhered onto the vertical wall of the stencil aperture when the stencil was retracted away from the wafer, which resulted in a much smaller solder ball or even on occasion, a missing solder ball. Therefore the bumped solder ball after reflow showed different sizes due to this solder paste volume variation. The non-uniformity of the solder balls across the die will causes placement yield problems since the die is sitting on the 3 highest solder balls during the placement and a much smaller solder ball will not be able to touch the pad, causing no wetting after reflow. As for the other stencil designs, the square 6.0mil aperture showed similar non-uniformed solder ball size due to the same low area ratio (50%); the square 6.5mil, square 7.0mil and round 7.0mil showed good printing quality, but these stencil designs were eliminated due to the potential "shrinkage" of the original solder ball during later assembly based on the simulation results in Chapter 3.

Therefore, stencil design with round 6.5mil apertures was selected among all the stencil designs. Both 3.5mil and 4.0mil bumping sites using round 6.5mil apertures showed good yield. However, the 4.0mil bumping site was chosen, since the 4.0mil opening has a higher tolerance for alignment during the printing process.

Table 4.2 Preferred stencil designs for three different bumping site opening diameters

Aperture Types	Area ratio (A_p/A_w)	Solder paste volume (mil^3)	Bumping site opening diameter	Predicted solder ball height 2^{nd} ball h / total h	Yield
Round 6.5 mil	54.17%	99.55	4.0mil	3.3mil/6.4mil	Good, easier for alignment
			3.5mil	3.6mil/6.7mil	Good
Round 6.0 mil	50%	84.82	4.0mil	3.0mil/6.1mil	Poor yielding, non-uniform bump height
			3.5mil	3.3mil/6.5mil	
			3.0mil	3.6mil/6.7mil	

4.1.4.3 Printer Setups and Process Parameters

An MPM TF100 Printer with vision system was utilized for all of the printing process studies. During the course of the printing process, the printer parameters must be appropriately defined, as they regulate the risks of defective double bumped dies, such as bridging between adjacent solder balls, and non-uniform solder ball size. The parameters established for the printer in this application are listed in Table 4.3.

Table 4.3 Printer setup and printing parameters

Height of stencil above wafer	0 (contact printing)
Down stop force applied on the squeegee	7-9 lb
Squeegee type(metal or polymer) and hardness	Polymer, 90 dorameter
Printing Speed	0.5”/sec
Angle of squeegee with respect to stencil surface	45 degrees

When all process parameters are set, the vision alignment system of the printer verifies the precise alignment of the openings in the stencil to the bumping site openings. Misalignment of the stencil aperture to the bumping site opening on the coated wafer can cause solder balls that shift away from their original bumping site onto the underfill surface. The solder paste is stencil printed through the openings in the stencil and deposited onto the wafer. Each printed wafer surface was inspected under a microscope to eliminate prints with bridging and significant non-uniformity defects.

4.1.5 Reflow

The printed wafers were sent through a reflow oven to form the second layer of solder balls. A Heller 1800 reflow oven with nine heating zones was utilized for this application. A typical SMT reflow profile (Figure 4.6 and Table 4.4) was developed according to the specifications of the solder paste. Nitrogen was utilized in the reflow oven as a protective atmosphere to prevent oxidation of the solder ball surface.

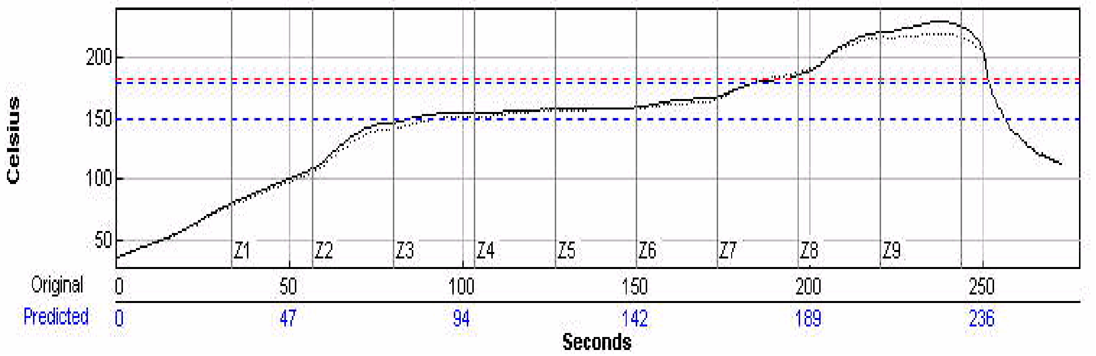


Figure 4.7 Reflow profile used for bumping process

Table 4.4 Reflow profile parameters

Max. Rising Slope	1.9 ℃/sec
Soaking time (150-180 ℃)	98.7 sec
Reflow Time (above 183 ℃)	59.3 sec
Peak Temperature	229.8 ℃

Two types of defects were discovered in the formation of the solder balls. The first one of these was solder bridging as shown in Figure 4.8. A bridge is the unwanted formations of an electrical connection (short) between two conductive points due to the solder extending (bridging) over the nonconductive material. The causes of solder bridges were found to be either the excess solder paste deposition or warped substrates. The excess solder paste deposition often occurs when the downstop of the stencil were not properly set to be fully in contact with the substrate or the force applied on the squeegee were too high, causing solder paste bleed underneath the stencil. A warped substrate due to incorrect polishing in the previous polishing process usually appears to have lower die thickness on the corners of the wafer squares. These corners were easier to get excessive solder paste deposition and have solder bridging as shown in Figure 4.8.

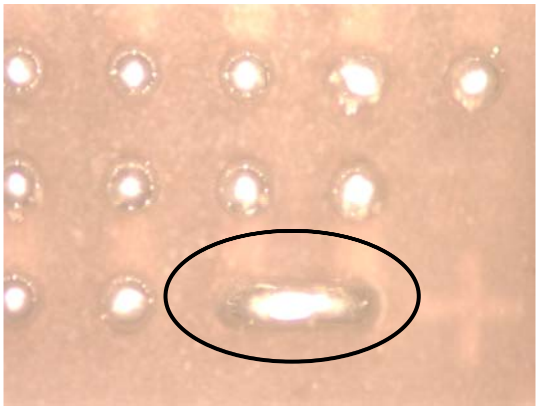
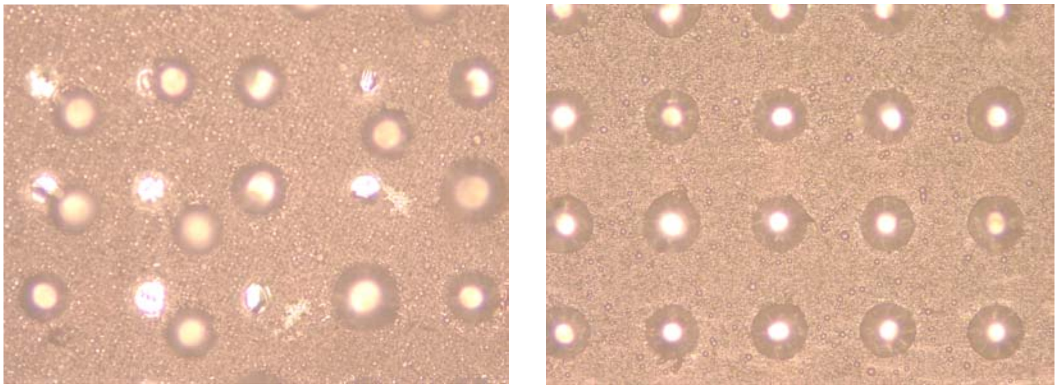


Figure 4.8 Solder bridging after reflow

Other defects were caused by the shifting of the solder balls during reflow as shown in Figure 4.9. The major reason for this phenomenon is a misalignment between the stencil aperture and desired bumping site. It was also found that shifting is more likely to occur when the bumping site diameter is too small.



(a) Secondary bump shifted from the original bump

(b) Secondary bump aligned with the original bump

Figure 4.9 Shifted solder balls and aligned solder balls after second layer solder bumping

Larger bumping sites and better uniformity across the wafer help to avoid such defects since the solder reflows on heating into a molten, low viscosity liquid, that has a tendency to self-align. Figure 4.10 shows cross-sectional pictures of the double bump flip chip after reflow. The height of the solder joint was dramatically increased comparing to a regular FA10 die.

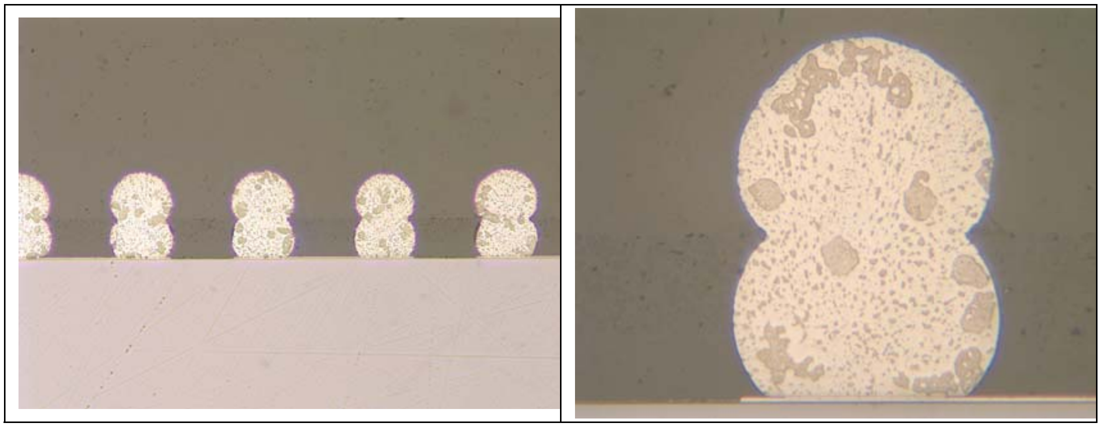
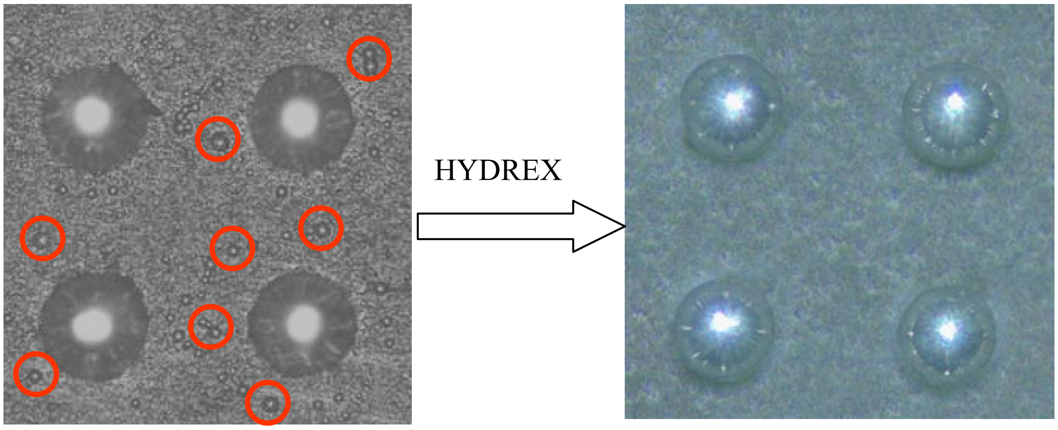


Figure 4.10 Cross-section of the double bumped solder ball after bumping with total bump height of 6.68mil

4.1.6 Surface Cleaning after Reflow

A certain amount of flux residue or solder paste residue was left on the cured underfill surface between solder balls after reflow, as shown in Figure 4.11(a). These residues would cause problems in later assembly processes, such as impeding the flow of the underfill, generating voids, and causing poor adhesion between the two underfill layers [83]. Furthermore, trapped flux residue in the assemblies may also degrade the reliability performance of the components. Thus, a cleaning step after double bump formation is necessary.

After reflow, the wafers were immediately cleaned using a 10% concentration of HYDREX DX aqueous cleaner. HYDREX DX is a high performance alkaline cleaner with exceptional wetting properties and detergency, which can be used to remove water soluble flux and paste residues as well as rosin and many no-clean flux and paste residues after reflow. The diluted HYDREX solution wash bath was heated to 55°C~ 65°C prior to the cleaning process. The dies were submerged into the bath with gentle stirring for about 2 minutes and were rinsed in the DI water. The air dried, cleaned dies were inspected under an optical microscope to check the cleanness of the surface. After cleaning (Figure 4.11), the surface of the die is much cleaner, indicating that the residues were successfully removed from the underfill surface by the cleaning process.



(a) before surface clean

(b) after surface clean

Figure 4.11 Double bumped die surface before and after cleaning

4.1.7 Singulation Process

A Micro Automation 1500 wafer dicing saw (Assembly Technologies) was used to singulate the coated double bump wafer into individual dies. The saw uses a closed circuit TV system with a split image to align the wafers before cutting.

Several challenges were encountered during this process. First of all, the loading effect of the underfill had to be taken into account. Secondly, the underfill layer did not reflect light in the same way as the silicon wafer, which introduced difficulties in the alignment during dicing. An enhanced light source was supplied to the wafer surface at a 45 degree angle in order to intensify the reflected images of the solder balls for alignment.

A special blade, the "Whisper Dice" flange-mounted resinoid blade from Dicing Technology, was utilized for singulation of the coated wafer. This blade was designed to cut very hard materials, and was expected to accommodate the loading effect due to the underfill. Blades with different thickness were tested and the smallest usable blade thickness was found to be 3mil. As a rule of thumb, the kerf width is approximately 1.3 times the blade thickness. As the saw street space from solder ball edge to solder ball edge was 25mil, this kerf width loss of 3.9mil was acceptable.

Table 4.5 compares the parameters utilized in the regular wafer dicing and coated wafer dicing. The cutting speed was reduced from 250mil/sec to 75mil/sec to reduce the loading on the blade and to avoid blade breakage.

Table 4.5 Comparison of the different specifications for regular wafer dicing and coated wafer dicing

	Original FA10 wafer	FA10 wafer with underfill coating
Blade Type	Sintered Blade	Resinoid Blade
Spindle speed	20000RPM-25000RPM	20000RPM-25000RPM
FWD cut speed	250mil/sec	75mil/sec
Film Height	2~3mil	2 ~3 mil
Wafer Thickness	35-40mil (must >27mil)	40-45mil (must >32mil)

After dicing, all the singulated dies had to be cleaned and dried. The dies were then ready for assembly and could be stored at room temperature, since the underfill layer was fully cured.

4.2 Assembly Process Development for No-flow Fluxing Underfill

In order to evaluate the reliability performance of the double bump flip chip, the fabricated double bump dies were assembled onto boards utilizing a no-flow fluxing underfill. The no-flow underfill eliminates the need for flow time and cure time after reflow, leading to a much higher throughput in the assembly process. Regular single bump flip chips were also assembled on boards with the no-flow underfill for comparison purposes. The process was developed and optimized for best results. Figure 4.12

illustrates the process used for double bump flip chips assembled with no-flow fluxing underfill.

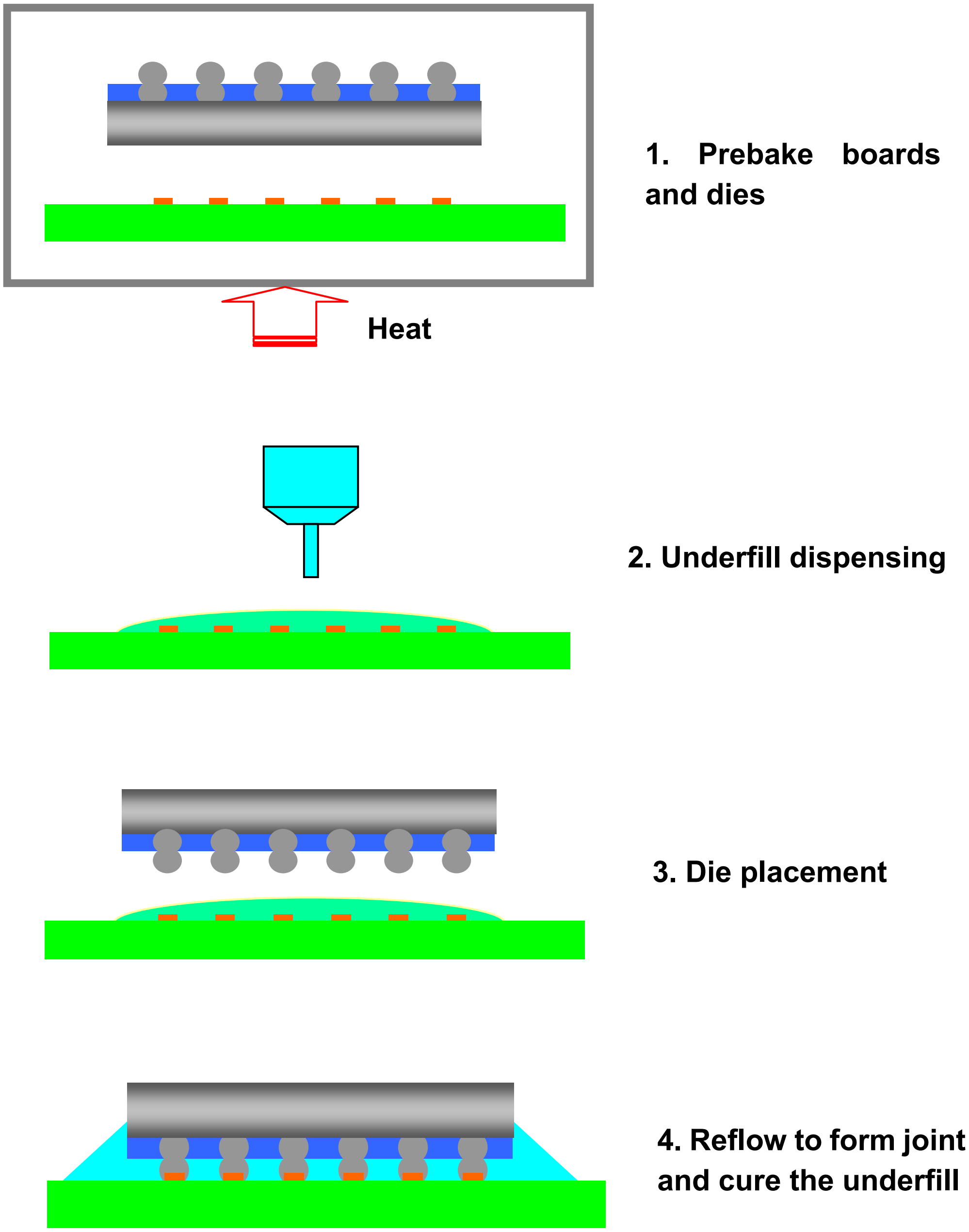


Figure 4.12 Illustration of the assembly process using no-flow fluxing underfill for double bump flip chips

4.2.1 Test Vehicle Description

The test vehicle consisted of area array dies mounted on a FR-4 substrate with eight sites. The board dimensions were 5”x3”x0.035”, and the board was made of high T_g (>185℃) laminate. The trace metallization on the boards was copper, electroless nickel, and immersion gold. Photographs of the board are shown in Figure 4.13. Two types of test dies were used in the assembly for comparison, one was a regular FA10 die with area array daisy chain, and the other was a double bumped FA10 die fabricated as described previously.

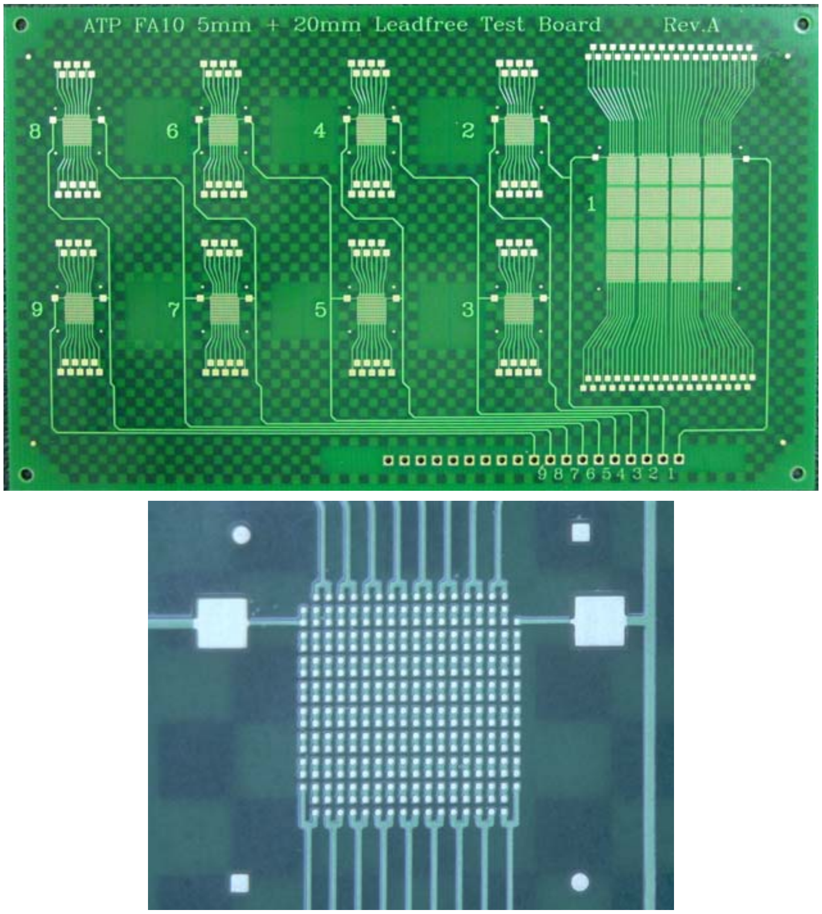


Figure 4.13 (a) The test board; (b) closer picture of the individual bonding site

4.2.2 Dehydration of Board and Die

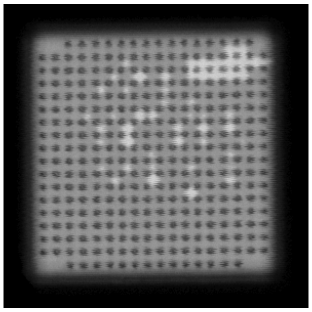
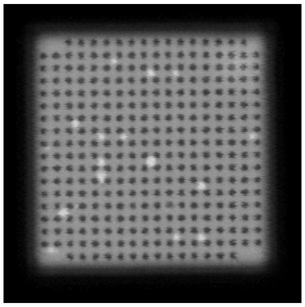
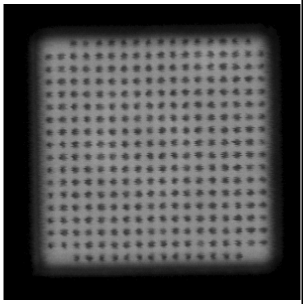
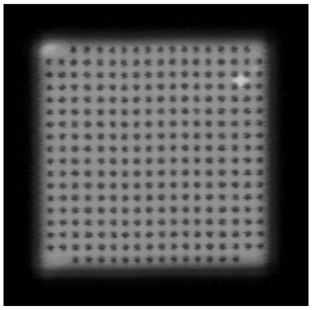
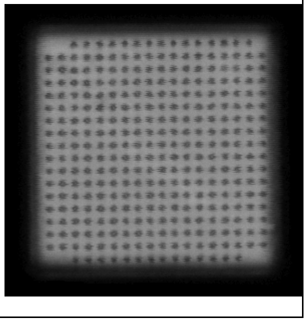
It has been well documented that the hydroscopic nature of PCB makes it absorb and retain moisture from the ambient air [84]. During the reflow process, the moisture captured in the PCB will be released in the form of gas when the PCB enters the high temperature zone. This gas will flow out into the gelling underfill layer above the PCB causing voids in the underfill and sometimes even floating the die. The solution to avoiding such defects is to prebake the PCB prior to assembly for a specified time at a sufficiently high temperature to drive the moisture out of the PCB.

Generally, silicon dies do not need pre-baking since the silicon does not absorb moisture. However, the double bumped flip chip die has an underfill coating as its exposed surface, rather than a silicon surface. In addition, these dies were exposed to water during the dicing and cleaning processes prior to assembly. Thus, pre-baking the dies was also performed in preparation for assembly and the baking time and temperature were investigated for best results. This also applied to regular FA10 dies that has been exposed to an aqueous environment prior to assembly, such as cleaning or dicing processes.

Two test matrixes containing sets of different prebake temperatures and prebake times were utilized in order to determine the most effective time-temperature combination for the board baking and die baking conditions. After reflow, the boards were scanned using acoustic microscopy to identify voids.

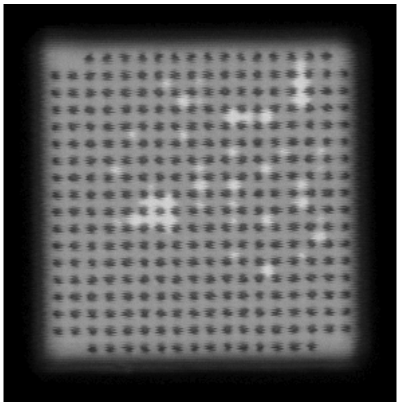
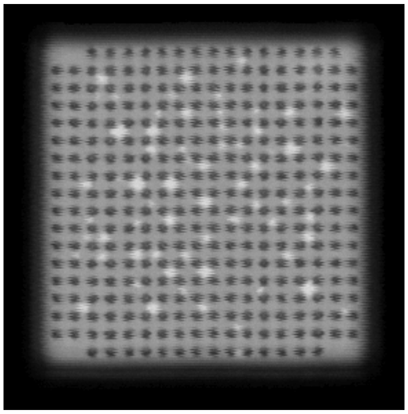
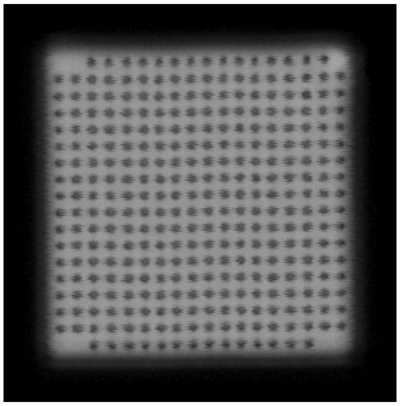
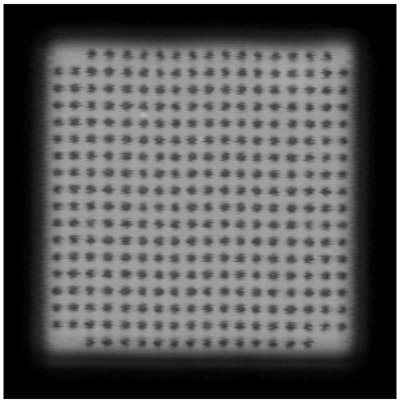
The first test matrix was designed to test the baking condition for the boards. All the dies tested in this experiment were baked thoroughly for over 48 hours at 125℃ to eliminate any possible residual moisture in the die. Dot dispensing was utilized for all assemblies using the same dispensing parameters. Two baking temperatures and two baking times were chosen for the test. The results, listed in Table 4.6, clearly indicate that moisture trapped on the boards was a major source of voids. However, an 18 hour bake at either 80℃ or 125℃ eliminated all the moisture, producing void free assemblies.

Table 4.6 Examples of the CSAM pictures under different prebake conditions for boards assembled with regular single bump flip chip dies

Baking Temp.	No bake	3 hours	18 hours
80℃			
125℃			

The second test matrix, similar to the previous one, was carried out to determine the optimum prebake condition for the elimination of moisture trapped in the die. The prebake temperatures tested were lowered to 60 ℃ and 90 ℃ in order to avoid damage to the silicon die. Prebake durations of 3 hours and 18 hours at these two temperatures for both regular FA10 dies and coated double bumped dies were tested. The test results again proved that the prebake significantly reduced the moisture level in both types of dies. An 18 hour bake at 90 ℃ was sufficient to drive out the moisture for both dies (Table 4.7).

Table 4.7 CSAM picture for dies with prebake and without prebake for both regular FA10 dies and coated double bumped dies

Regular FA10 die (single bump)	Coated Double bumped dies
 without prebake	 without prebake
 prebake at 90 ℃ for 18 hours	 prebake at 90 ℃ for 18 hours

4.2.3 Underfill Dispensing

Dispensing was performed with a Cam/Alot dispense system using a 23-gauge needle. Kester SE-CURE® 9126 was used in both double bump die assemblies and regular single bump die assemblies. This is a no-flow fluxing underfill with a single component liquid made up of epoxy-based flux and underfill. Properties are given in Table 4.8. The Kester 9126 formula requires no needle or board heating during the dispensing step according to manufacturer's datasheet.

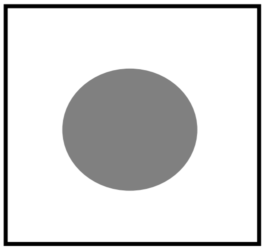
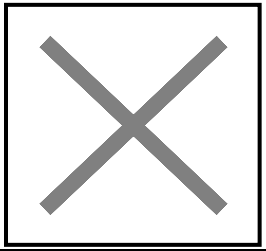
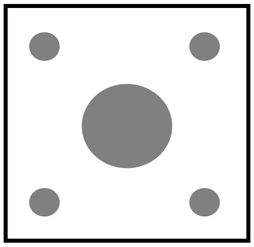
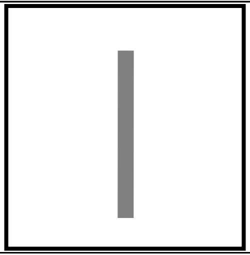
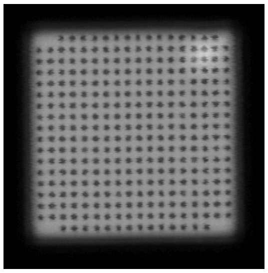
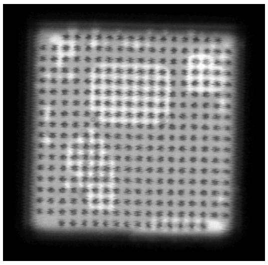
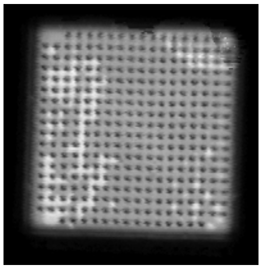
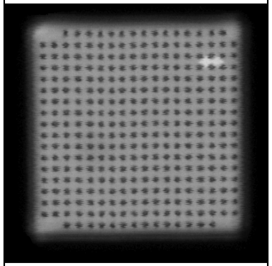
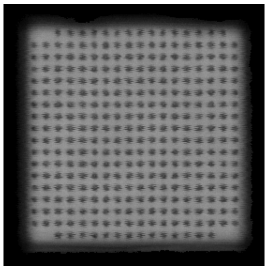
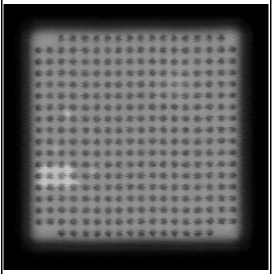
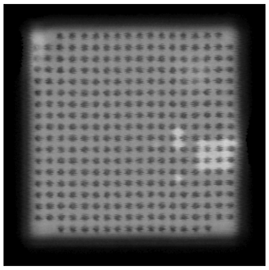
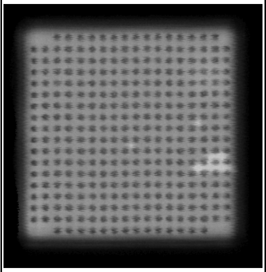
Table 4.8 Material properties for Kester 9126 no-flow underfill

Viscosity	CTE	Tg
3500 poise	70ppm/℃	70 ℃

The dispensing of the no-flow underfill must create appropriate volumes and shapes that allow all bumps to be fluxed during reflow and then form proper fillets. Voids should be few and as small as possible. When designing the dispensing process, viscosity must be taken into account in order to optimize the settings. For high viscosity materials, an insufficient distance of the dispense nozzle from the board tends to prevent the material from exiting the nozzle, resulting in a smaller volume of encapsulant being dispensed onto the board. High viscosity materials also tend to form a "tail" as the needle retracts, which may causes underfill strings deposition on undesirable locations. However, none of these problems are serious for Kester 9126 underfill when using a 23-gauge needle, with the needle height set to be 0.012 inch from the board.

The dispensing pattern has a major influence on the underfill dispensing results. The simplest dispensing pattern is a single drop at the center of the site, where higher dispensing rates can be achieved since the dispensing head is not moving. Other alternative dispensing patterns tested include a cross pattern, a center line pattern and a five dot pattern with one in the center and four at the corners, as shown in Table 4.9.

Table 4.9 Voiding study of different dispensing patterns

Dot (Glob)	**Cross**	**Five dot**	**Center line**
			
			
			

The CSAM pictures of the assemblies for different dispensing patterns showed that the single dot pattern and center line pattern generated far fewer voids than the other two patterns. The cross pattern produced several void-free assemblies, but occasionally formed large voids. This is because multiple flow fronts converged during the placement and trapped air voids inside. Thus, the single dot pattern has a clear advantage due to uniform symmetric flow that passes from the center towards the edges in all directions.

Based on the dispensing pattern study results, the dot pattern appears to offer the best results and was thus chosen for use in the remaining process parameter studies. The amount of underfill dispensed for each site was another key parameter which needed to be controlled. It was studied using a DOE designed matrix along with the die placement factors of placement force and dwell time when holding the chip in compression as discussed in the next section.

4.2.4 Die Placement

The next step in the process after underfill dispensing was to place the die into the deposited underfill on the board. A typical flip chip placement process can be divided into five steps.

a) The die is picked up by the placement machine and vision recognition performed by the vision system in order to match the die to the desired bond site on the board based on the board fiducials.
b) The placement machine places the die on the board at a pre-set z direction velocity.
c) The placement machine measures the force exerted by the underfill material on the chip using the placement head force sensor and keeps moving down until a programmed force limit is reached.
d) The placement machine then maintains this force for a specified length of time, the dwell time, while still holding the chip
e) The placement machine releases the chip after the dwell time, and starts to retracts to its original position.

The key controllable process variables in this placement process are the placement speed in step (b), the placement force limit in step (c) and the dwell time in step (d). Unlike in a conventional capillary flow process, these process variables are observed to be critical in the no-flow underfill process because inappropriate selection of those parameter values could lead to a significant process yield loss. For instance, if the die does not make contact with the pad during placement with a small placement force, the die would have the tendency to float during the reflow, causing low yield. A certain dwell time can help the die perimeter wet wetted with the fluxing underfill and form the fillet around the edge.

In this study, the pick and placement of the die was accomplished using an ACM automatic pick and placement machine. This machine allows only the control of the placement force and dwell time during placement, with no control over the placement speed. Therefore, a factorial DOE design was constructed to optimize the process. Three factors were chosen for studies: placement force, dwell time and the dispensed underfill volume.

A full factorial DOE matrix for 3-level 3 factors requires at least 27 runs. In order to reduce this requirement, an L_9 Taguchi DOE Design was carried out for optimization of the assembly process.

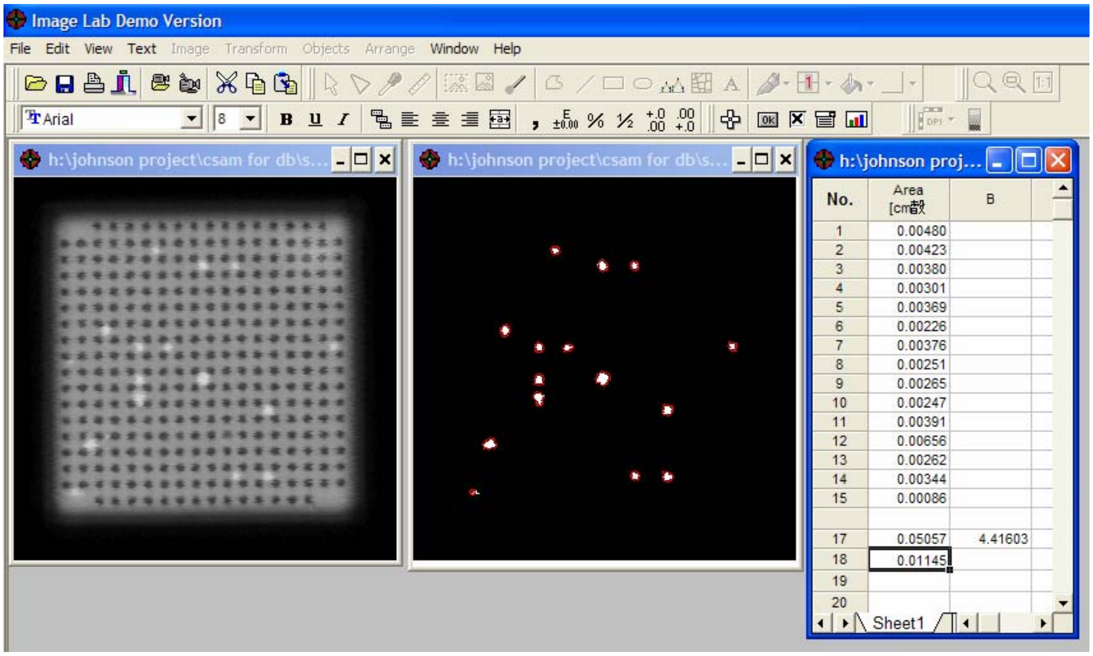
Table 4.10 The DOE design matrix for assembly process parameters

Run #	DOE Factors			Response
	Underfill amount	Placement force	Dwell Time	Area% of voids
1	2 mg	1N	1 s	12 10 24
2	2 mg	5N	3 s	21 12 8
3	2 mg	9N	5 s	15 17 17
4	4 mg	1N	3 s	3 0 6
5	4 mg	5N	5 s	0 0 1
6	4 mg	9N	1 s	4 2 3
7	6 mg	1N	5 s	5 7 6
8	6 mg	5N	1 s	8 8 3
9	6 mg	9N	3 s	1 4 2

All the assemblies using the nine process parameters combinations in Table 4.10 had 100% electrical continuity. Therefore, the responses of the DOE design were defined as the distribution of voids in the assemblies. The CSAM images were analyzed in order to obtain the amount of voiding, reported in terms of the percentage of the total die area using the software ImageLab.

Figure 4.14 gives an example of the void analysis process for a CSAM picture of the assembly. The CSAM picture was loaded into the program, and then was converted into a black and white image showing the void area by changing the contrast threshold in ImageLab. Then all the white area in this converted image was identified and the area of each void was calculated by the software as listed in the spreadsheet on the right side of Figure 4.14. The total area of the die was pre-calculated using the same technique in ImageLab. Thus the area percentage of the total voids in the die can be obtained.

The void area percentages for different process parameter combinations obtained from the calculation using ImageLab were recorded, and used as the response input for factorial analysis using MINITAB statistical software. The results are listed in Table 4.10 for each run of each combination. The response table of the analysis is shows in Table 4.11 with the criteria set to be the smaller the better. The results shows that the underfill weight factor was the most important factor, ranked number 1, while the dwell time has the weakest influence on the voiding response. The main individual effect of three factors is plotted in Figure 4.15, which shows that the best results were obtained at underfill weight level 2, placement force level 3 and dwell time level 2. A combination of process parameters of 4mg underfill, a 9N placement force and a 3 second dwell time was the best combination among all the combinations with the current level setting. All the subsequent assemblies were made using this combination of process parameters.



No.	Area [cm敤	B
1	0.00480	
2	0.00423	
3	0.00380	
4	0.00301	
5	0.00369	
6	0.00226	
7	0.00376	
8	0.00251	
9	0.00265	
10	0.00247	
11	0.00391	
12	0.00656	
13	0.00262	
14	0.00344	
15	0.00086	
17	0.05057	4.41603
18	0.01145	
19		
20		

Figure 4.14 Voids detection and area calculation using software ImageLab

Left: CSAM picture of the assembly; Center: Captured voids areas; Right: datasheet of the area for each void and the calculated area percentage of the void vs. die area

Table 4.11 Response table based on criteria: smaller is better for voids% response

Level	**UF Weight**	**Placement Force**	**Dwell Time**
1	-23.9980	-17.2569	-15.7770
2	-10.2002	-17.4729	-14.5210
3	-13.5632	-13.0317	-17.4634
Delta	13.7978	4.4412	2.9424
Rank	1	2	3

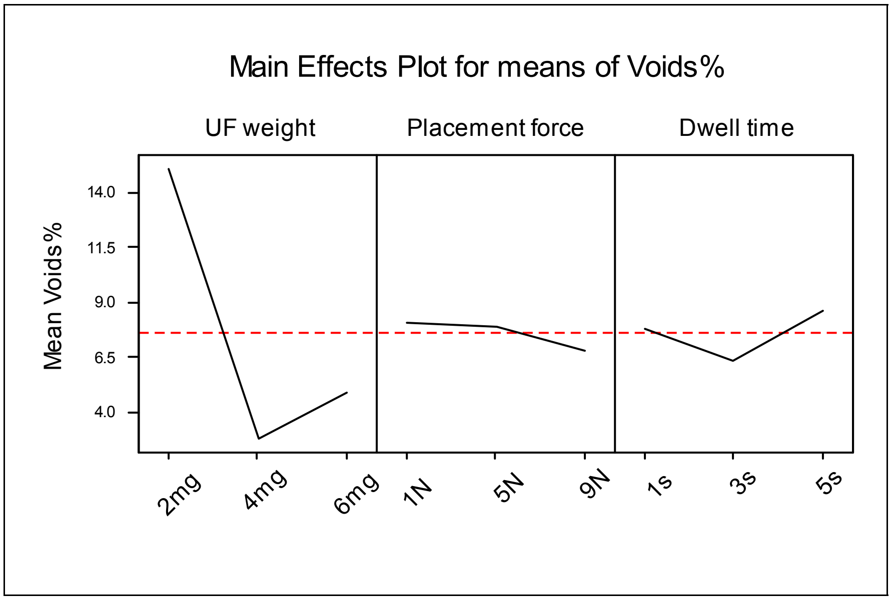


Figure 4.15 The main effect plot for three factors with voids% as the response

The interaction effects of the three factors were also analyzed in Figure 4.16. The interaction plot between placement force and dwell time (Figure 4.16, second plot in right column) shows three response curves crossed with each other, indicating the most strong interaction between placement force and dwell time factors among all three interaction combinations. The interaction plot between underfill weight and placement force (Figure

4.16, first plot in center column) shows parallel response curves, indicating the least interaction between underfill weight and placement factors. The interaction between underfill weight and dwell time (Figure 4.16, first plot in right column) is relatively hard to judge since it shows the similar trends when underfill weight changes, but when dwell time is 3 second, the 4mg and 6mg underfill weight response curve interacted with each other, indicating some small interaction.

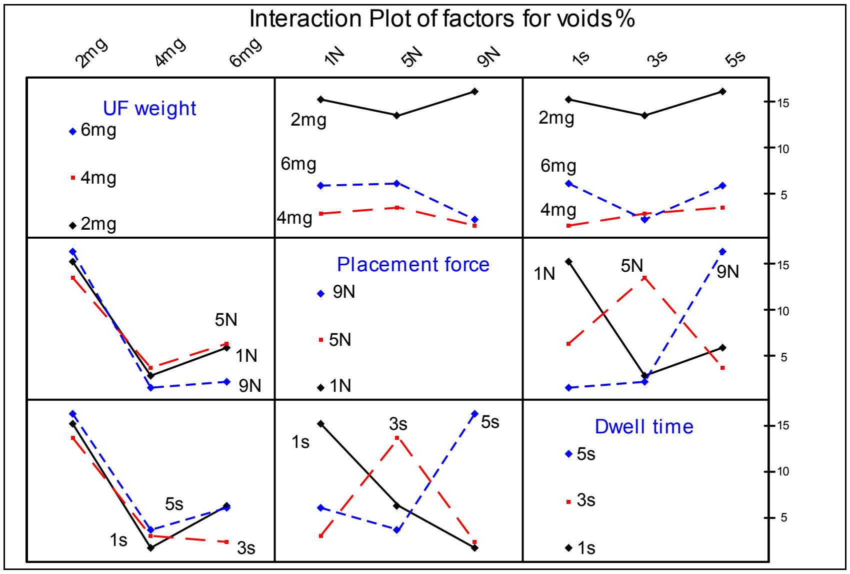


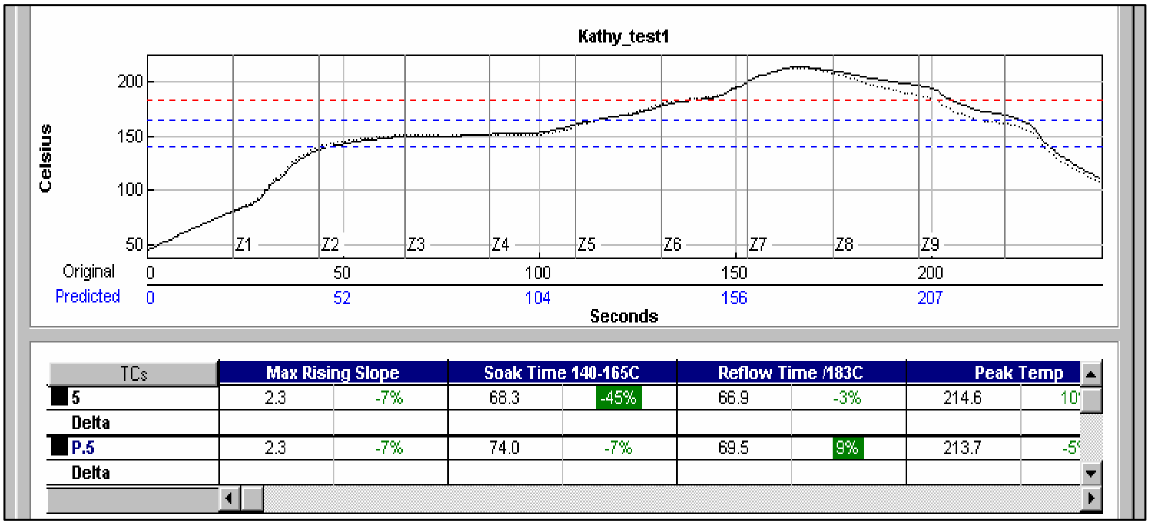
Figure 4.16 2-factor interaction plots for all combinations with the three evaluated factors: the underfill weight, placement force and dwell time

4.2.5 Reflow Process

Reflow is a most complicated process in optimizing for assemblies with no-flow fluxing underfill because several processes take place simultaneously and their requirements sometimes conflict. The underfill material must retain its flowability before melting of the solder bumps so it will not prevent joints from forming and the die from collapsing. However, the underfill must be cured by the time the board comes out of the reflow oven. Some of these materials are sensitive to variations in the reflow profile, resulting in poor wetting of the solder ball onto the pad if the temperature ramps up too high too soon.

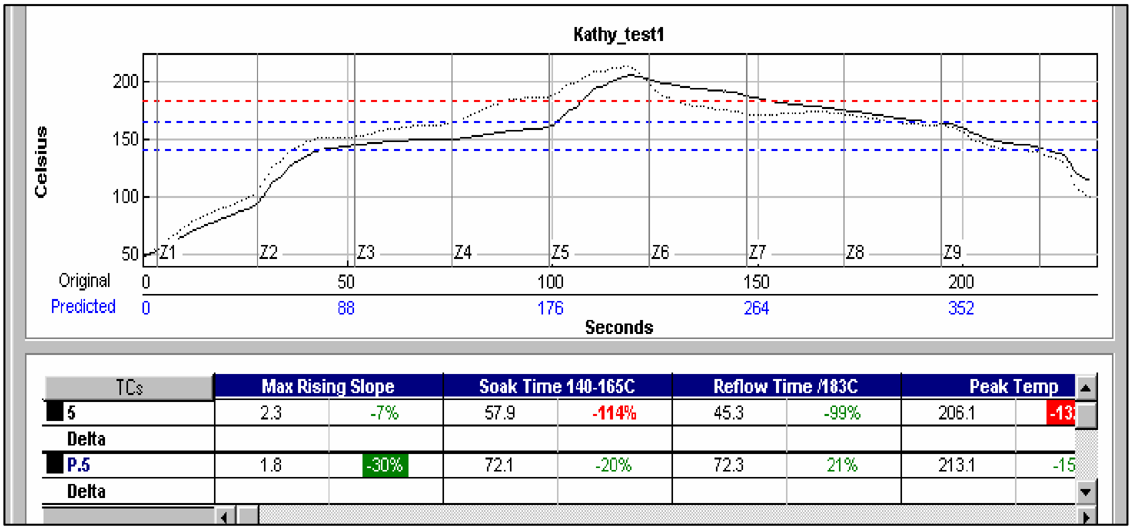
Two reflow profiles were developed for optimization. The first reflow profile was compatible with the traditional SMT profile (Figure 4.17). This profile would be preferred if other surface mount components are to be built at the same time. The other profile, namely the "early peak profile" has a shorter soaking time and shorter reflow time (Figure 4.18). This reduces the time available to activate the flux, but provides a longer curing time after the solder is wetted onto the pad. These two profiles are plotted in Figure 4.19 for comparison and the characteristic parameters, such as measured soaking time, reflow time and peak time were listed in Table 4.12. The temperature zone settings and conveyor speed for these two profiles are listed in Table 4.13.

The "early peak" profile caused non-wetting problem for the assemblies as shown in Figure 4.20. The quick temperature ramp up to reflow temperature left insufficient time for flux activation. Without proper fluxing, the solder exhibited a poor wetting. Therefore, the standard SMT profile was chosen to use for both single and double bump flip chip assemblies. Figure 4.21 shows the good wetting double bump flip chip after reflow.



TCs	Max Rising Slope		Soak Time 140-165C		Reflow Time /183C		Peak Temp	
5	2.3	-7%	68.3	-45%	66.9	-3%	214.6	10
Delta								
P.5	2.3	-7%	74.0	-7%	69.5	9%	213.7	-5
Delta								

Figure 4.17 Reflow profile no.1 (“standard”)



TCs	Max Rising Slope		Soak Time 140-165C		Reflow Time /183C		Peak Temp	
5	2.3	-7%	57.9	-114%	45.3	-99%	206.1	-13
Delta								
P.5	1.8	-30%	72.1	-20%	72.3	21%	213.1	-15
Delta								

Figure 4.18 Reflow profile no.2 (“early peak”)

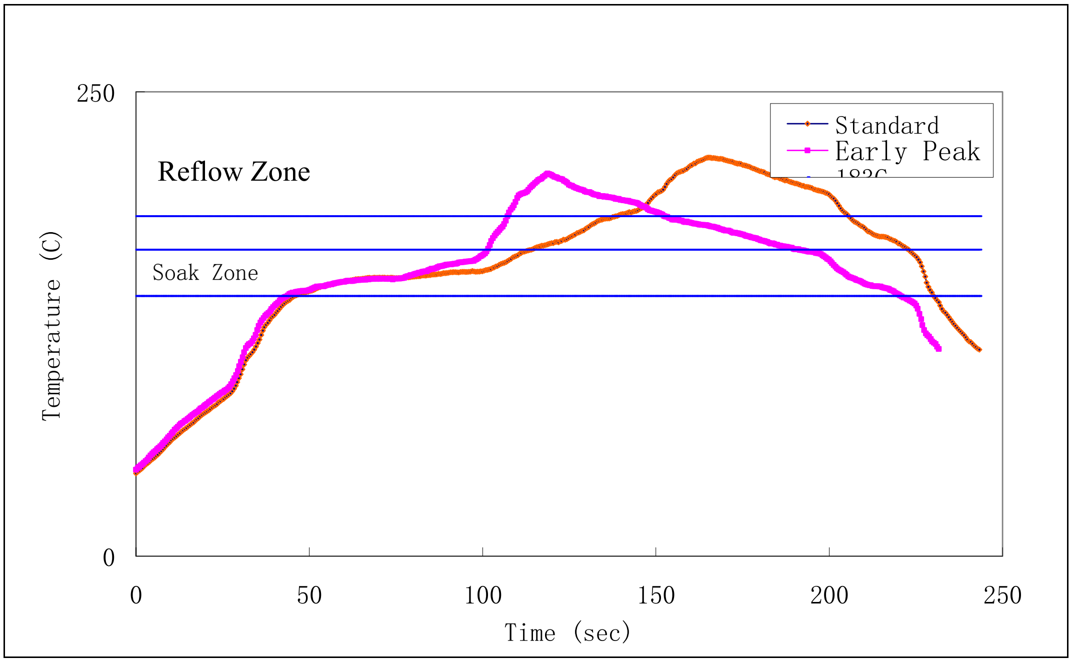


Figure 4.19 The two different types of reflow profiles

Table 4.12 Characteristic properties of the two reflow profiles

	Standard SMT Profile	Early Peak Profile
Max Rising Slope	2.3 ℃/sec	2.3 ℃/sec
Soak Time (140 to 165 ℃)	68.3 sec	57.9 sec
Reflow Time (>183 ℃)	66.9 sec	45.3 sec
Peak Temperature	214.6 ℃	206.1 ℃

Table 4.13 Reflow oven temperature zone settings for the standard SMT and early-peak profiles (℃)

Standard SMT Profile								
Zone 1	Zone 2	Zone 3	Zone 4	Zone 5	Zone 6	Zone 7	Zone 8	Zone 9
235	170	154	155	180	200	240	195	160
Conveyor Speed :67 cm/minute								
Early-Peak Profile								
Zone 1	Zone 2	Zone 3	Zone 4	Zone 5	Zone 6	Zone 7	Zone 8	Zone 9
235	170	152	157	240	185	175	162	140
Conveyor Speed :67 cm/minute								

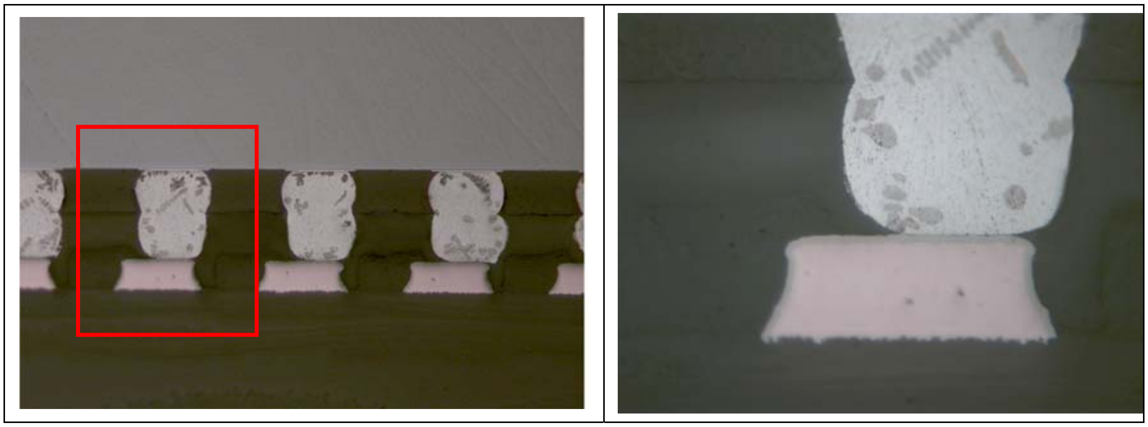


Figure 4.20 Non-wetting solder joint of double bump flip chip

After establishing the proper underfill dispensing process and reflow profiles, a set of double bump flip chip assemblies were built as well as a set of single bump flip chip assemblies. Figure 4.21 shows the cross-sectional pictures of single bump and double bump assembled on the board. The standoff height of the assemblies was increased significantly by using the double bump structure.

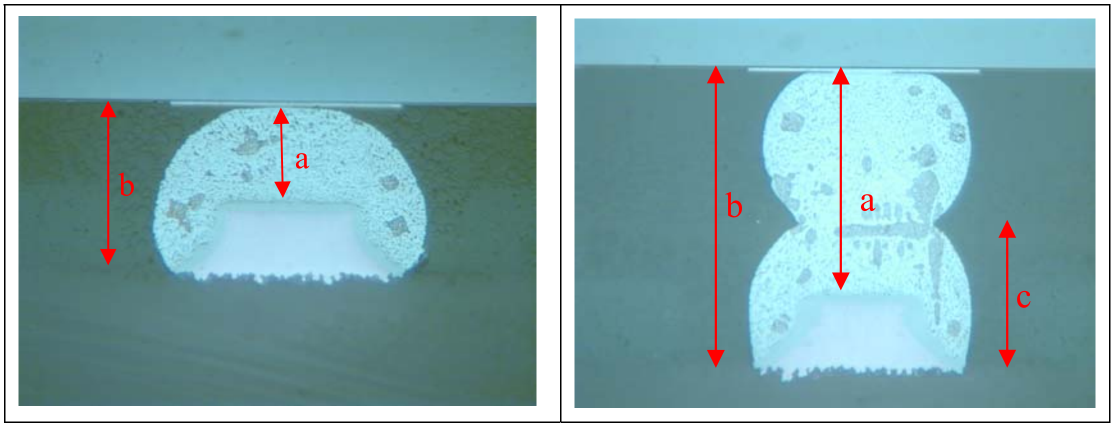


Figure 4.21 Cross-sectional pictures of single bump flip chip on board and double bump flip chip on board

4.3 Reliability Tests

Common reliability tests of electrical components include thermal cycling, thermal shock, temperature humidity aging, etc. Both thermal cycling and thermal shock tests consist of subjecting the parts to the specified low temperature then subjecting the same units to the specified high temperature for a specified number of cycles. These tests are performed in order to determine the ability of the parts to withstand cyclical temperature exposures in real applications. The thermal shock test is similar to the thermal cycling test, but more severe as it also tests the parts’ resistance to sudden changes in temperature. Either air or liquid can be used as the medium for transferring thermal energy, as indicating in the nomenclature for the air-to-air thermal shock test (AATS), liquid-to-liquid thermal shock test (LLTS) and air-to-air thermal cycling test (AATC).

Air-to-air thermal shock test was carried out to test the reliability of the double bumped flip chip assemblies. A Ransco air-to-air thermal shock environmental test chamber was used for this reliability test. It features dual compartment thermal shock chambers with a temperature range of -70°C to 200°C . The elevator feature provides a transfer time of less than 5 seconds between the hot and cold zone chambers.

The thermal shock conditions used in this study were from -40°C to 125°C, with 15 minutes dwell time at each temperature extreme. Figure 4.22 shows the time-temperature profile measured by a thermocouple attached to the test board during cycling.

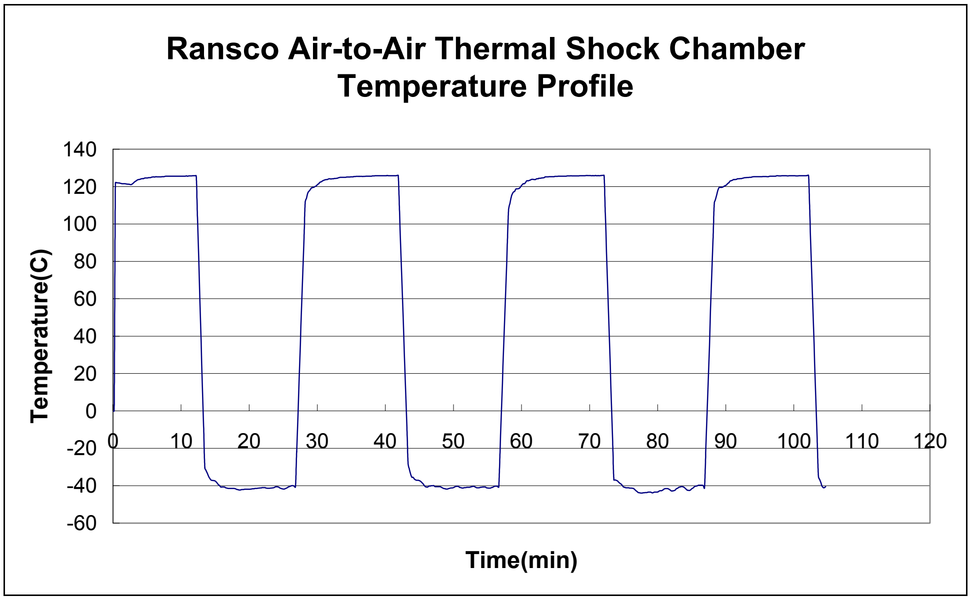


Figure 4.22 Air-to-air thermal shock chamber temperature profile

(-40℃ to +125℃)

After the assembly process, two types of test vehicles were subject to the accelerated thermal shock test. One group was assembled with regular single bump flip chip and the other group was assembled with double bumped flip chips. Each group contained 28 measurements, i.e. die sites. Kester's no-flow fluxing underfill SE-CURE9126 was used for assembling these test vehicles. All of the boards were electrically tested and then subjected to C-SAM examination prior to the reliability test to be qualified as known good parts.

In-situ monitoring is necessary for accurate thermal shock or thermal cycle testing of underfilled flip chip dies. It was observed that the failures (open circuits) recorded at the extreme temperatures still appeared to have good electrical connections at room temperature. This is because at the high temperature, the expansion of the underfill opens and expands the possible crack in the solder joint, while at the room temperature, the contraction of the underfill holds the cracked solder joint back together, maintaining electrical continuity.

In order to in-situ monitor the resistance of each daisy chain of the test die, every boards were wired outside of the chamber to a LabView controlled measurement system, which employed a Keithly 2000 multimeter, a Keithly 7002 switchbox and a custom-designed interface board. During the thermal shock test, the resistance of each daisy-chained flip chip assembly was measured and monitored real-time for failure. A failure was defined as 5 occurrences of resistance over the threshold resistance. The threshold resistance was set to 30 ohms. The maximum resistance of the flip chip assembly was below 15 ohms before they were placed in the thermal shock chamber.

The tests were terminated at 1750 cycles by which time 15 of the double bump flip chip assemblies and the entire regular single bump flip chip assemblies had failed. After completing the thermal cycling test, a Weibull analysis was conducted in order to characterize the failure data. The Weibull distribution plot of the AATS test results for the two types of assembly are presented in Figure 4.23. Table 4.14 summarizes the first failure, characteristic life and the shape parameter for both assemblies.

It can be concluded from the reliability results, that the double bump flip chip assemblies had a higher characteristic life than that of the single bump flip chip assemblies, and thus better reliability. The improvement was about 43% for the double bump structure. The first failure of the double bump flip chip assemblies occurred at 1112 cycles, which is also markedly higher than 851 cycles, where the first failure occurred among the regular single bump flip chip assemblies.

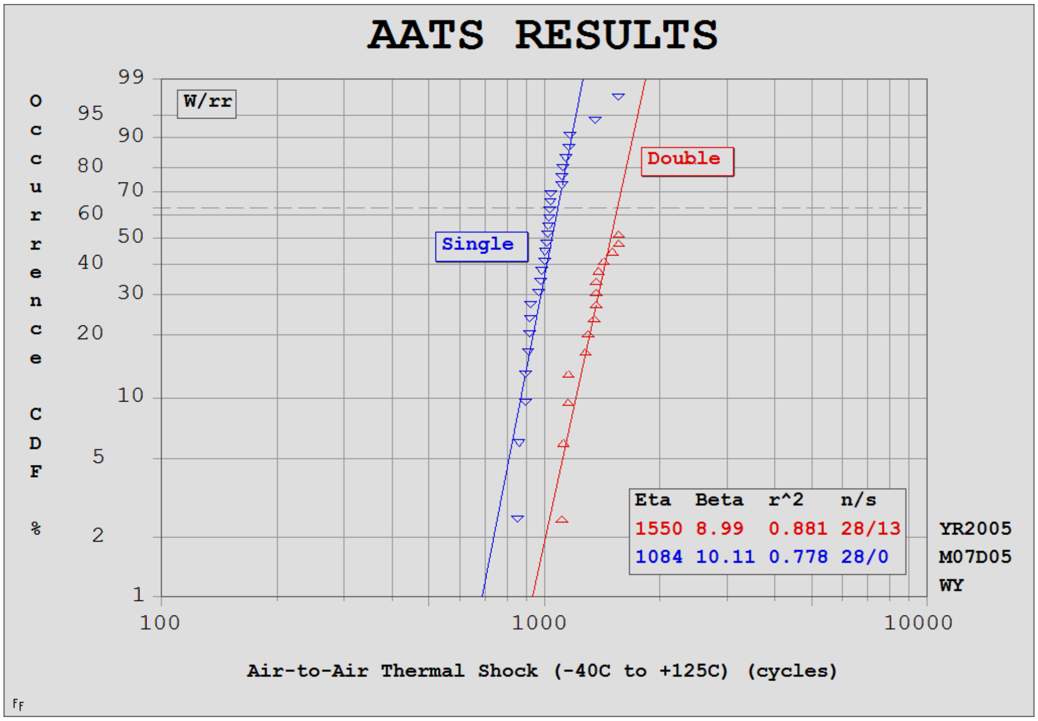


Figure 4.23 Weibull plot of the thermal shock test data for single bump flip chip and double bump flip chip assemblies

Table 4.14 Summary of AATS test results

	First Failure	Characteristic Life	Shape Parameter
Single	851 cycles	1084 cycles	10.11
Double	1112 cycles	1550 cycles	8.99

4.4 Failure Analysis

All the tested assemblies were examined using acoustic microscope (CSAM) after the tests to detect underfill delamination. Cross-sections of the related solder bumps in the failed dies were inspected in order to analyze the failure mechanisms involved.

Solder joint fatigue is the major failure mode in flip chip. Solder fatigue cracks account for the electrical failures observed in AATS testing. Normally, solder joint fatigue is observed at locations close to the UBM (under bump metallurgy). Figure 4.24 shows a failed single bump solder joint with the crack at the interface between the solder joint and the die. However, a new crack location was found at the middle of the solder joint in the double bump structure. Figure 4.25 shows the cross section of a double bump solder joint with a fatigue crack appearing at the interface of the two stacked solder balls. This crack location indicated that the middle rim of the interface of the two stacked solder balls was the location where strain most concentrated and experience the most plastic deformation. This is due to the different CTE of the two underfill layers, which results in a significant shear strain at the interface of the two underfill layers. This shear strain can lead to local solder fatigue at the interface. A finite element analysis using ANSYS with similar findings for double bump flip chip was conducted and will be described in detail in Chapter 5.

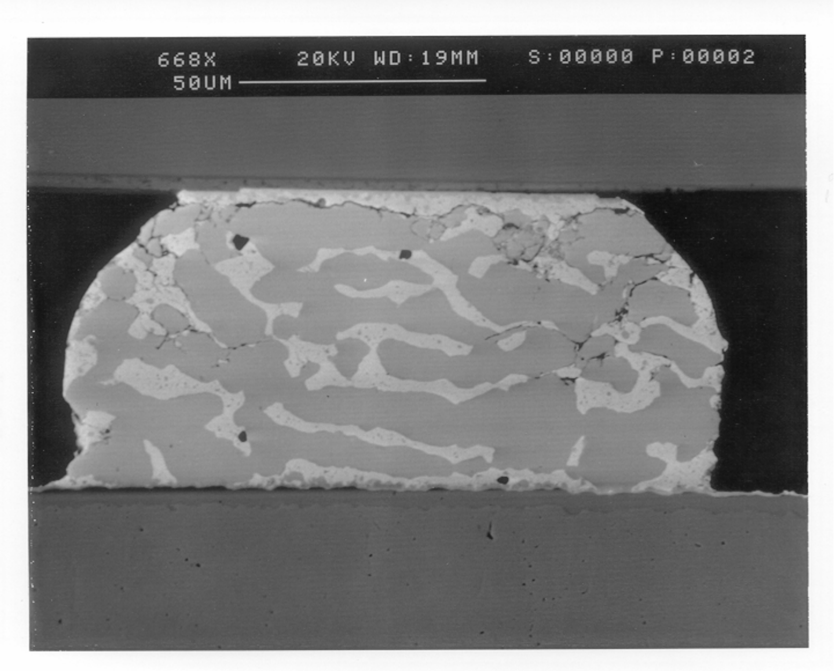


Figure 4.24 Cross-sectional picture of a failed single bump die with a crack at the interface between solder and die

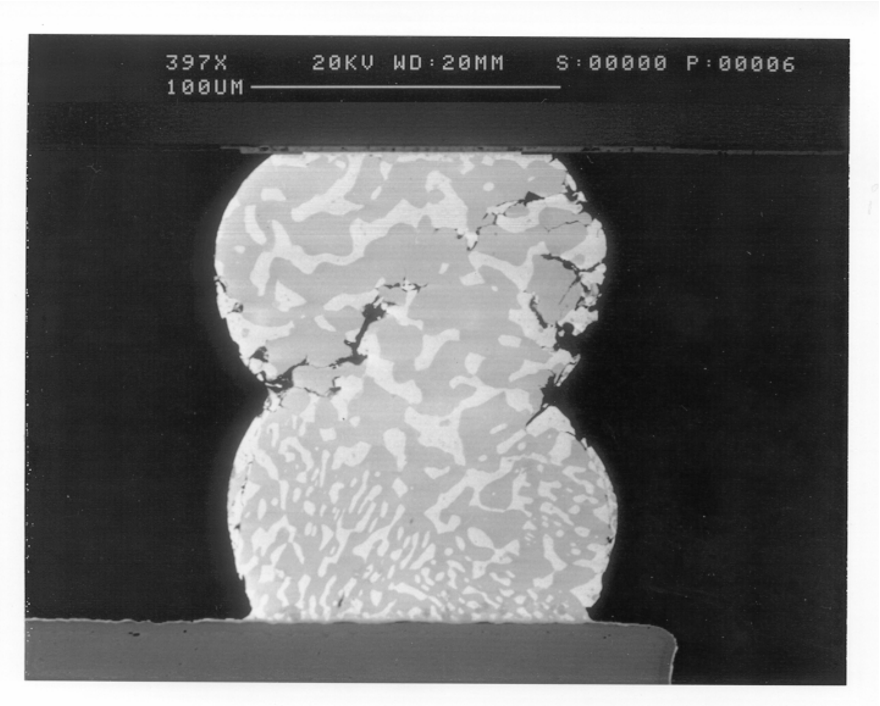


Figure 4.25 Cross-sectional picture of a failed double bumped die with a crack at the center neck of the solder joint

Underfill adhesion is also critical to the reliability of a flip chip assembly, however, underfill delamination is another dominant failure mode in flip chips. Figure 4.26 shows a typical delamination of double bump flip chip detected after reliability test. The delamination generally appears around the corners or edges of the chip. In order to locate the location of the delamination of the underfill, a TAMI (Tomographic Acoustic Multiple Imaging) scan technique was applied during the acoustic microscope scanning of the double bump flip chip after 1750 thermal shock cycles. A sequence of the acoustic scanning pictures at different depths of the assembly were acquired during the TAMI-scan as shown in Figure 4.27. As indicated in Figure 4.27(b), the delamination occurred at the interface of the die and the first layer of underfill.

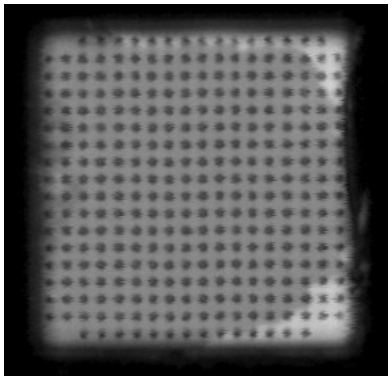
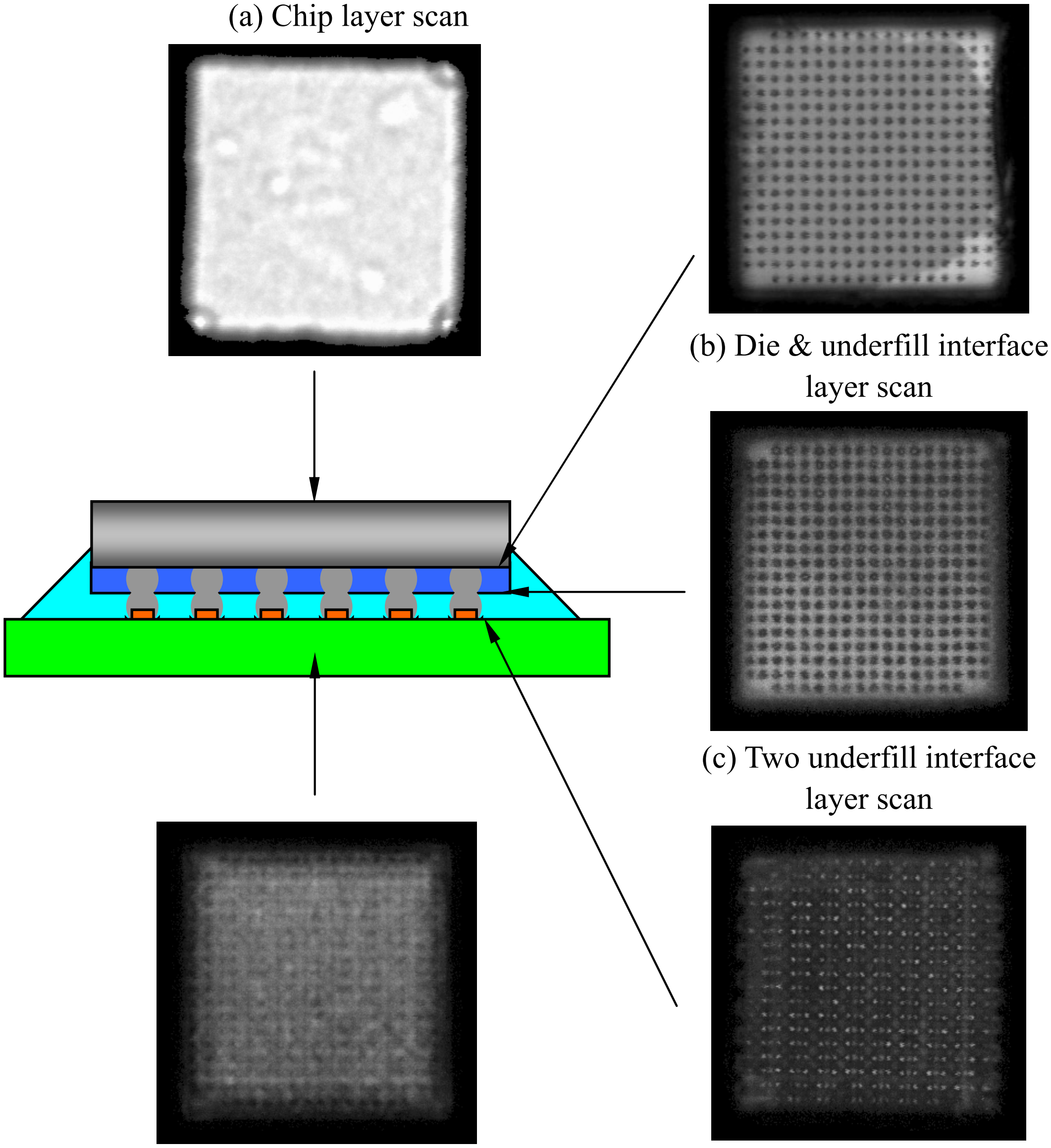


Figure 4.26 CSAM picture of the double bump flip chip after 1750 cycles showing underfill delamination


(a) Chip layer scan

(b) Die & underfill interface layer scan

(c) Two underfill interface layer scan

(d) Pad layer scan

(e) Board inner layer scan

Figure 4.27 TAMI scan pictures at different layers using the scanning acoustic microscope, showing the delamination occurs at the interface of the die and the 1st underfill layer

For further confirmation, the double bump flip chips with detected delamination were cross-sectioned and polished for cross-sectional inspection. The delamination of the underfill is usually very small and difficult to find in the cross-sectional pictures. For easier inspection, the test board was kept in thermal shock chamber for additional 450 cycles after electrical monitoring was terminated at 1750 cycles. Figure 4.28 shows a solder joint failure at the die side due to the underfill delamination.

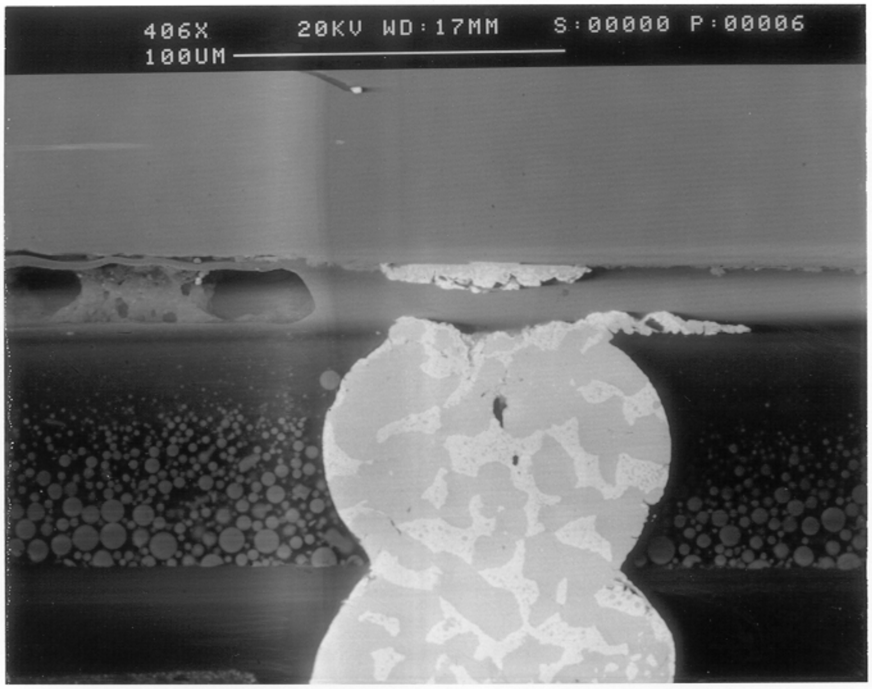


Figure 4.28 Failures due to the underfill delamination at the interface of the die and the 1st underfill

CHAPTER 5

FINITE ELEMENT ANALYSIS

As discussed in previous chapters, the reliability of flip chips relies heavily on the materials properties of all the components in the flip chips, as well as the solder joint geometry and standoff height. In order to further understand the double bump solder joint design characteristics, a nonlinear finite element model was established for thermal cycling analysis, and the results compared to a standard single bump flip chip with the same material properties and dimension definitions. The lifetime of the solder joint was predicted based on the simulation results obtained from the finite element model calculations.

5.1 Literature Review for Finite Element Modeling

To understand the failure modes of flip chips and provide better estimates of reliability, it is important to understand the local strain / stress distribution. As the I/O density becomes higher and solder bump size and pitch get smaller, it is almost impossible to directly measure the thermal strains or stresses experimentally. The reason is that this kind of fine measurement requires the experimental tools that offer an extremely high resolution of measurement. The other difficulty is experiment time. As with the other experiments commonly used to investigate time-dependent behavior, it is very time consuming to perform this type of measurement for electronic packages. A typical thermal cycling experiment usually takes 4 months to accomplish 2000 cycles. Hence, the use of finite element (FE) modeling, as an alternative method for reliability studies of electronic package has gained a lot of attention.

An FE model commonly requires five steps:

1) Details of the component geometry must be known

2) A constitutive model for each material in the package must be defined. This model requires information on properties such as elastic modulus, CTE, Poisson ratio and nonlinear properties like creep and viscoplasticity.

3) An appropriate geometry model must be established with a suitable mesh applied to divide the bulk material into small elements. The mesh construction also directly affects the accuracy of the simulation results.

4) The thermal loads applied to the model must represent those in a real thermal cycling experiment. Sub-steps need to be included taking model convergence into consideration.

5) Suitable reliability models need to be applied in order to predict the thermal fatigue life of solder joints using the simulation results.

5.1.1 Approaches to Establishing a Geometric Model

Analyzing the complete assembly using the finite element method is generally prohibitive in terms of computer cost and time. Various finite element modeling techniques have therefore been developed for use in predicting the fatigue life of a solder joint in BGA, CSP and flip chip assemblies. These common modeling techniques can be classified into the following categories.

A. Two Dimensional Model

A two dimensional finite element model is the simplest of all model approaches. It offers the advantage of minimal computing time and requires much less effort to create. However, as one would expect, it loses the third dimension details and generally results in very inaccurate results.

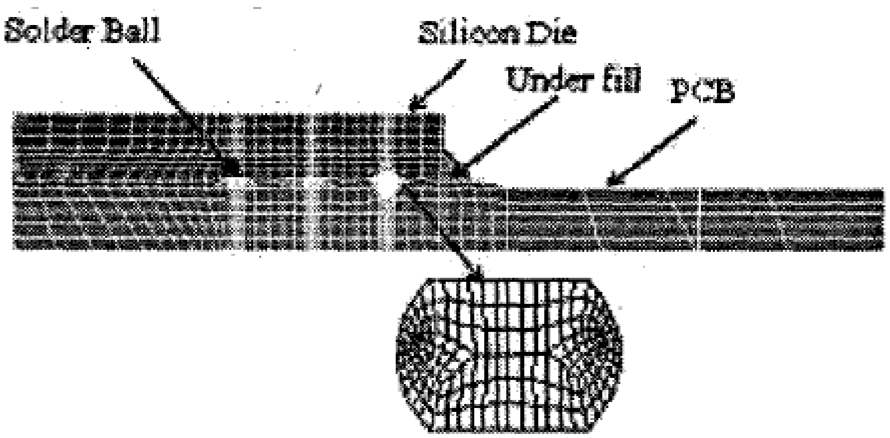


Figure 5.1 Two dimensional finite element model with free mesh at the solder ball [86]

B. Three Dimensional Model

3D model simulations are more realistic than the 2D models, but more computing time and resources are needed because of the dramatic increase in the complexity and element number. Unlike 2D models, the 3D models usually require more advanced techniques in the process of modeling implementation. There are divided into three categories as slice model, global model, and global model with submodel.

B1) Slice Model

This commonly accepted modeling technique uses only a slice of the assembly in order to save computation time. This model assumes the opposite plane to be parallel to the symmetry plane. This may, however, result in an over estimation of the shear loading on the solder joint during temperature cycling, resulting in a conservative prediction of the solder fatigue life [87].

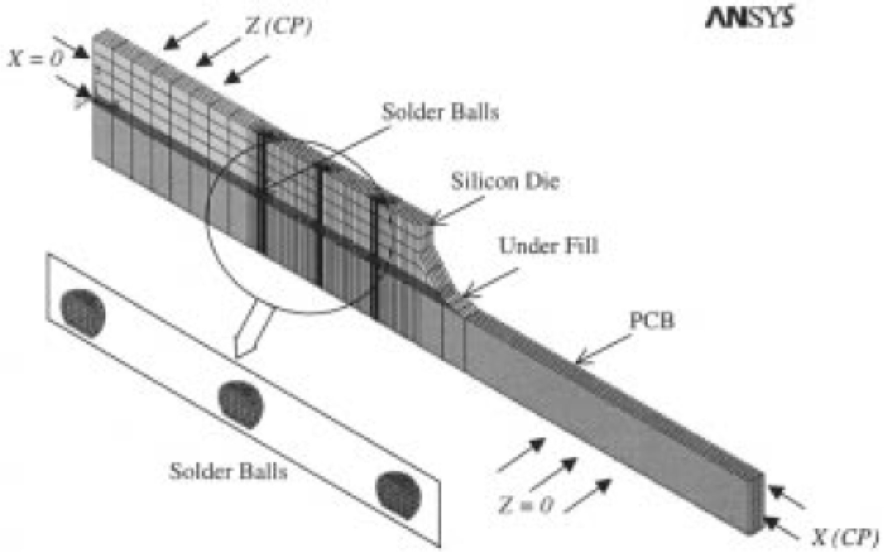


Figure 5.2 Nonlinear slice model [88]

B2) Global Model

The global model may utilize either one quarter or one octant [89] of the whole chip, taking advantage of the package symmetry. To reduce the computing time, the global model is usually modeled using a coarse mesh, except at the critical solder joint, which is modeled with a much finer mesh for analysis. This approach avoids the assumptions associated with the boundary conditions of the slice model. However, the mesh generation process for this approach becomes very time consuming, because the fine mesh associated with the critical solder joint(s) must match the coarse mesh of the remaining solder joints, die, and substrate, etc.

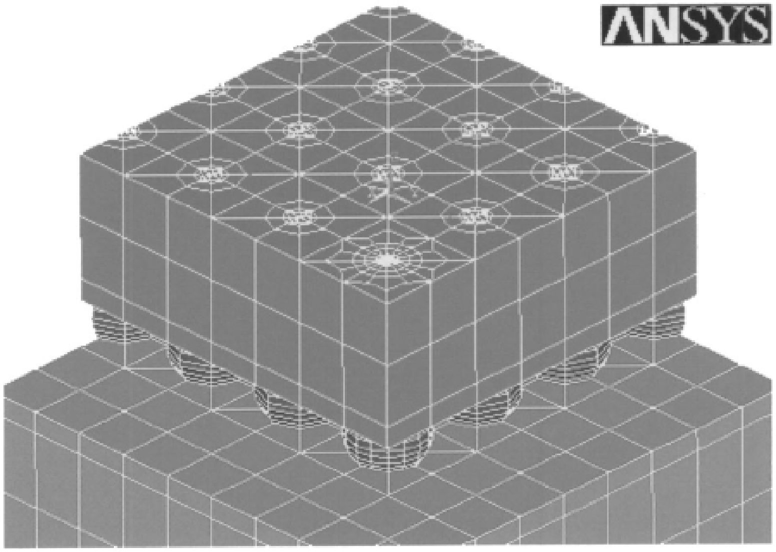


Figure 5.3 Finite element octant model with one critical solder joint in refined mesh [90]

B3) Global Model with Submodel

An alternative to the preceding approach is to use a model of the package on a global scale to identify the critical solder joint for subsequent nonlinear submodeling. As shown in Figure 5.4, the submodel of the critical solder joint has a much finer mesh and can utilize a far more detailed geometry. The displacement at the boundary of the critical position is first obtained from the simulation results of the global model. These displacement results are used as the boundary conditions for the submodel.

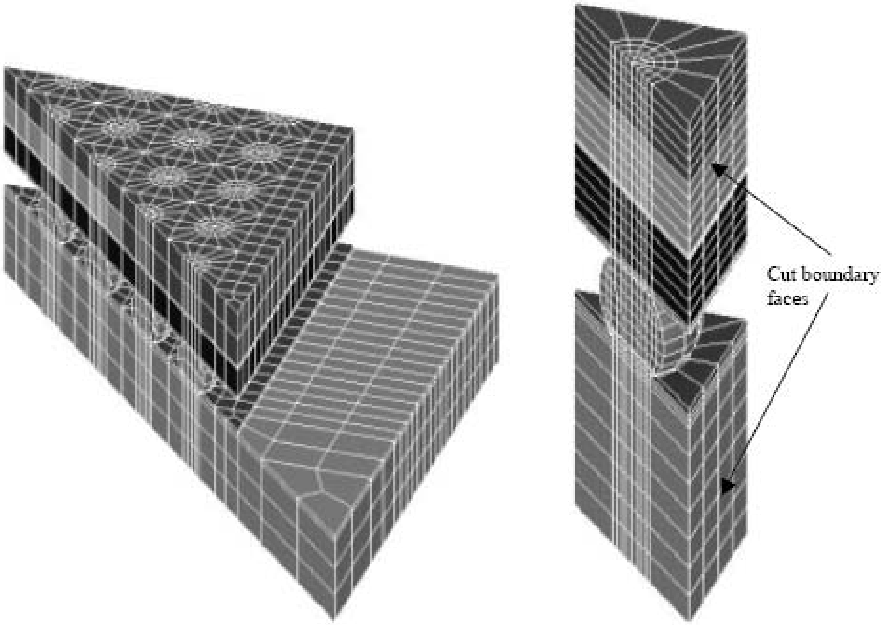


Figure 5.4 (a) Octant global model with coarse mesh (b) refined submodel with refined mesh [89]

5.1.2 Different Material Constitutive Models for Eutectic Solder

The most effective model reported in the literature, showing a very good fitting capability, used a dual description divided into an elastic/plastic model and a creep model[91]. The elastic/plastic model is usually defined as either time-dependent or temperature dependent. For the steady state creep model, two alternative models, the sinh law model and the double power law model are often used. The most popular application of the sinh law model was implemented into the form of Anand viscoplastic model, described below.

A. Anand's Viscoplastic Model For Eutectic Solder

The viscoplastic constitutive law introduced by Anand (1982) is a frequently used material model for the definition of solder material behavior. The steady state creep of solder can be expressed by a sinh law relationship of the form [92]:

$$\dot{\varepsilon}_{cr} = C_{ss}[\sinh(\alpha\sigma)]^n \exp\left(\frac{-Q_a}{kT}\right) \tag{5-1}$$

where $\dot{\varepsilon}_{cr}$ is the steady state strain rate, T is the absolute temperature, σ is the applied stress, Q_a is the apparent activation energy, n is the stress exponent, α prescribes the stress level at which the power law dependence breaks down, and C_{ss} is a constant.

The sinh law model is often applied to the models built in ANSYS in the form of Anand's model. The reason is that ANSYS uses Anand's constitutive model as the standard option to describe viscoplastic elements. It is convenient for the user since the user does not need to modify the source code.

There are two basic features in Anand's model that are applicable to an isotropic rate-dependent constitutive model for metals. First, there is no explicit yield surface, so the instantaneous response of the material is dependent on its current state. Secondly, a deformation resistance variable "s", is used to represent the isotropic resistance to inelastic flow of the material. Anand's model consists of a flow equation and three evolution equations that relate the inelastic strain rate to the rate of deformation resistance [93].

Flow equation:

$$\frac{d\varepsilon_p}{dt} = A\left[\sinh\left(\xi\frac{\sigma}{s}\right)\right]^{\frac{1}{m}} \exp\left(-\frac{Q}{kT}\right) \tag{5-2}$$

Evolution equation:

$$\frac{ds}{dt} = \left\{ h_0 \left(|B| \right)^a \frac{B}{|B|} \right\} \bullet \frac{d\varepsilon_p}{dt} \tag{ 5-3}$$

where

$$B = 1 - \frac{s}{S^*} \tag{ 5-4}$$

with

$$s^* = \hat{s} \bullet \left[\frac{d\varepsilon_p / dt}{A} \exp(\frac{Q}{kT}) \right]^n \tag{ 5-5}$$

where $d\varepsilon_p$ = effective inelastic deformation rate,

σ = effective Cauchy stress,

s = deformation resistance,

S^* = saturation value of deformation resistance,

$\hat{s}$ = time derivative of deformation resistance,

T = absolute temperature,

Q = activation energy,

R = universal gas constant,

A, ξ = material constant,

m = strain rate sensitivity of stress,

n = strain rate sensitivity of saturation (deformation resistance)

The inelastic strain rate in Anand's definition of material is dependent on the temperature and stress, as well as the loading rate. As seen in the above equations, there were nine material parameters involved in the Anand's model. Note that equation (5-1) is already included in the Equation (5-2) of Anand's model for the viscoplastic elements, where the parameters Css, α, and n correspond with A, 1/m and ξ/s, respectively.

B. Other Sinh Law Creep Models For Eutectic Solder

In contrast to the Anand model, other models incorporate plasticity along with the creep. Examples are Wiese's model [94], Darveaux's model [92] and Feustel's model, as listed in Table 5.1, Table 5.2 and Table 5.3, respectively.

Table 5.1 Wiese's sinh law creep model [94]

$\dot{\varepsilon}_{cr} = C_1\left[\sinh(C_2\sigma)\right]^{C_3}\exp\left(-\frac{C_4}{T}\right)$			
C1 (sec^{-1})	C2 (MPa^{-1})	C3	C4(kJ/mol)
10	0.2	2	44.9

Table 5.2 Darveaux's sinh law creep model [95]

$\dot{\varepsilon}_{cr} = C_{4t}\frac{G}{T}\left[\sinh\left(\alpha_t\frac{\sigma}{G}\right)\right]^n\exp\left(-\frac{Q}{kT}\right)$			
Shear Modulus: $G=1.9 \times 10^6-8.1 \times 10^3(T-273)$ (psi)			
C_{4t} (K/sec/psi)	α_t	Q (eV)	n
0.114	751	0.548	3.3

Table 5.3 Feustel's double power law creep model [96]

$\dot{\varepsilon}_{cr} = B_1 \exp\left(\frac{-H}{kT}\right)\left(\frac{\sigma}{E}\right)^{n_1} + B_2 \exp\left(\frac{-H}{kT}\right)\left(\frac{\sigma}{E}\right)^{n_2}$				
Young's Modulus: E=0.088T(℃)+32GPa				
B1(1/sec)	B2 (1/sec)	H (eV)	n_1	n_2
1.7×10^{12}	8.9×10^{24}	0.468	3	7

5.1.3 Life Prediction Analysis Models

Numerous solder fatigue models have been reported in the literature [92, 97-105]. These models can be categorized into two major groups, one of which is strain-stress based, and the other is energy based. Strain range and energy density are used as the failure parameters.

5.1.3.1 Strain-based Approach

Strain-based models have been extensively used to characterize the low cycle fatigue behavior of solder alloys. These models predict solder fatigue life by correlating the total strain (or plastic or creep strains) to cycles to failure. In this category, almost all strain based models originated from the Coffin-Manson relation. The Coffin-Manson model was a major breakthrough in the field of low cycle fatigue and has many applications [100].

A. Original Coffin-Manson Model

The original Coffin-Manson model [97] consists of plastic strain and elastic strain, as shown in the equation:

$$\varepsilon_{tot} = \varepsilon_{plastic} + \varepsilon_{elastic} = MN_f^Z + \frac{G}{E} N_f^v \tag{ 5-6}$$

where N_f is the number of cycles to failure, ε_{tot} is the total strain, $\varepsilon_{plastic}$ is the plastic strain, $\varepsilon_{elastic}$ is the elastic strain, E is the elastic modulus, and M, Z and v are material constants.

B. Solomon's Isothermal Cycling Model

Solomon [98] modified the Coffin-Manson relation into the following equation:

$$\Delta\gamma_p N_f^\alpha = \theta \tag{ 5-7}$$

where $\Delta\gamma_p$ is the applied shear strain range, N_f is the number of cycles to failure, and α and θ are solder material constants which are functions of temperature. These constants can be empirically determined from shear strain-cycling tests under certain conditions (temperature and cycling frequency) by constructing a logarithmic plot of N_f versus γ_p for different solder materials. For Sn40Pb60 solder, approximately,

$$N_f = 1.292\Delta\gamma_p^{-1.96} \tag{ 5-8}$$

This modified model assumes the plastic strain is the control parameter for low cycle fatigue (<10,000 cycles). The limitations of this model, which are also true for the following models, are the difficulty in defining $\Delta\gamma_p$ and N_f. All solder joints have large gradients in $\Delta\gamma_p$, because of their complex geometry and boundary conditions.

C. Engelmaier's Modified Coffin-Manson Model

Engelmaier [99, 100] proposed a solder joint fatigue life prediction model based on the total shear strain range.

$$\overline{N}_f = \frac{1}{2} \cdot \left(\frac{\Delta\gamma}{2\varepsilon'_f} \right)^{\frac{1}{c}} \qquad \text{5-9)}$$

where N_f is the median fatigue life in cycles, $\Delta\gamma$ is the applied shear strain range. ε'_f is the fatigue ductility coefficient, which is 0.325 for eutectic solder. C is the fatigue ductility exponent, which can be described by the following equation:

$$C = -0.442 - 6 \times 10^{-4} T_m + 1.74 \times 10^{-2} \ln(1 + f) \qquad (\text{5-10})$$

The exponent takes into account the mean temperature (T_m in ℃) and the cyclic frequency of the temperature loading. The c value for a typical -40℃ to +125 ℃ thermal cycling test with 30 minutes cycling time is -0.49149.

D. Another Modified Coffin-Manson Model

The equation given below is another modified Coffin-Manson model [101]. This model also considers the cyclic frequency and the extreme cycling temperature to be among the factors that contribute to the fatigue life.

$$N_f = \left(\frac{A}{\Delta\gamma_p} \right)^m f^n e^{-\beta / kT_{\max}} \qquad (\text{5-11})$$

where N_f is the number of cycles to failure, $\Delta\gamma_p$ is the cyclic plastic shear strain of a solder joint, A is the solder material constant which is function of temperature, f is the cyclic frequency, T_{max} is the maximum temperature during the cycle, β is the empirical "activation energy", k is the Boltzmann constant, and n is the empirical constant related to cyclic frequency. Constant m is the test condition constant, which is related to the cycles to failure and the shear strain. And m was reported to vary from 1.89 to 2.5 [101]. The predictions obtained from this model are conservative because it only considers the shear plastic strain as the prominent parameter for thermal fatigue life.

5.1.3.2 Energy-based Approach

A strain-based model is required in order to obtain a single inelastic strain value in a solder joint, but it is much easier to calculate strain energy density from the stress-strain hysteresis loops for any type of solder joints. Therefore, in recent years, energy-based low cycle fatigue models have been increasingly used to predict the fatigue life of solder alloys. Most of these models use the strain energy density accumulated in the solder joint during the thermal cycling instead of the strain energy to avoid having to take into account the volume difference.

Several finite element base analysis methodologies using energy based methods have been proposed which predict solder joint fatigue life. A few are briefly reviewed in the following sections.

A. Darveaux's Energy Based Model

Darveaux's model [92] predicts solder joint fatigue life by using the accumulated inelastic strain energy density per thermal cycle as the failure parameter. ΔW_{ave} is a volume-weight averaged inelastic energy density defined as:

$$\Delta W_{ave} = \frac{\sum \Delta W_i \bullet V_i}{\sum V_i} \qquad (5\text{-}12)$$

where ΔW_i is designated as the plastic work density in the i^{th} element and V_i is the volume of the element.

Darveaux correlated the crack initiation time (N_0) and crack growth rate (da/dN) to ΔW_{ave} as follows:

$$\text{Crack Initiation: } N_0 = K_1 \Delta W_{ave}^{K_2} \qquad (5\text{-}13)$$

$$\text{Crack Growth: } \frac{da}{dN} = K_3 \Delta W_{ave}^{K_4} \qquad (5\text{-}14)$$

The characteristic life can be calculated as:

$$\text{Characteristic fatigue life: } N_f = N_0 + \frac{a}{da/dN} \qquad (5\text{-}15)$$

Where a is the characteristic crack length to failure, which generally corresponds to the pad diameter or solder joint diameter.

Constants K_1, K_2, K_3 and K_4 were decided by experimental measurement of the crack length destructively throughout the thermal cycle tests. The values for these constants of different geometry models are given in reference [92].

B. Shi's Frequency Dependent Model

Shi, et al. [102] developed a temperature and frequency dependent energy based fatigue life model for eutectic solder given by

$$\left[N_f \nu^{(k-1)}\right]^m \left[\frac{\Delta W}{\Delta \sigma_f}\right] = C \qquad (5\text{-}16)$$

where Δσ is the stress range and ΔW is the energy density in MPa, calculated from the hysteresis stress strain loop. The constants are C=1.69 and m=0.7, (k-1) =-0.1 for $v \geq 10^{-3}$ Hz or (1-k) =0.59 for $v < 10^{-3}$ Hz.

C. Clech's SRS Model

Clech [103, 104] developed the Solder Reliability Solutions (SRS) model to predict surface mount area array and chip scale assembly reliability. The SRS model uses simplified one-dimensional structural physical models to estimate the solder joint stress/strain due to global CTE mismatch, which refers to shear deformation of solder joints driven by thermal expansion differentials between the components and the board. A correlation of the SRS solder joint failure data produces the following equation:

$$\frac{N_f}{A} = \frac{6.149 \times 10^7}{\Delta W^{0.998}} \qquad (5\text{-}17)$$

Where: N_f is the characteristic life, A (in^2) is the solder crack area for fully cracked electrically open solder joints, which is also the minimum solder joint load bearing area, ΔW is the cyclic inelastic strain energy per unit volume, obtained as the sum of strain energies due to global and local CTE mismatches, and C is a model calibration factor, which ranges from 0.4 to 2.7. This model accounts for both global and local CTE mismatches.

D. Zahn's Power Law Model

Zahn [105] used a power law fit equation to predict the fatigue life of the solder joint, as shown in the following equation:

$$N = C_1 (\Delta W_{avg})^{C_2} \qquad (5\text{-}18)$$

where N is the calculated cycles to failure, which depends on the life prediction coefficients being used. Life prediction coefficients for 1st failure, the median life (N_{50}) and characteristic Life ($N_{63.2}$) are provided using four different critical solder ball joint interface thicknesses and element divisions. Table 5.4 lists the parameters for eutectic solder and lead free solder. Experimental reliability test data validate the predictions, showing that this model provides a reasonable degree of accuracy.

Table 5.4 Prediction parameters used for models with a 0.0127mm element layer for calculations

	C1	C2
63Sn37Pb		
1st failure	881.01	-0.7407
N_{50}	991.46	-0.8946
$N_{63.2}$	1098.30	-0.8827
95.5Sn4Ag0.5Cu		
1st failure	1754.5	-0.1486
N_{50}	2419.1	-0.1390
$N_{63.2}$	2441.3	-0.1558

5.2 Implementation (Modeling Procedure)

ANSYS 7.0 software was used to carry out the simulation in this study. The construction of a finite element model includes a consideration of the material properties, the geometry and mesh, and the loading profile. In addition, the careful selection of the element type is also essential due to its effect on the simulation accuracy.

Compared to 2D models, 3D simulations are more realistic, especially when the flip-chip lacks symmetry, or when local 3D geometric features have a strong influence on the results. However, a 3D model obviously requires a much larger number of elements and nodes for analysis; hence a parametric study of 3D models may need considerable computing resources. In order to avoid the need for an advanced computer CPU, more memory and long computing times to simulate a 3D model that includes complicated structural details, some simplifications are necessary.

Several assumptions were specified in the simulation, namely, that there were (a) no initial residual stress; (b) no transient heat transfer; (c) uniform temperature distribution in the package; and (d) temperature loads were gradually applied.

The life prediction of solder joints under thermal cyclic loading normally requires a two-stage analysis, whose flowchart is illustrated in Figure 5.5:

(1) a 3D finite element analysis to compute the stress, strain or energy field;

(2) a life prediction analysis.

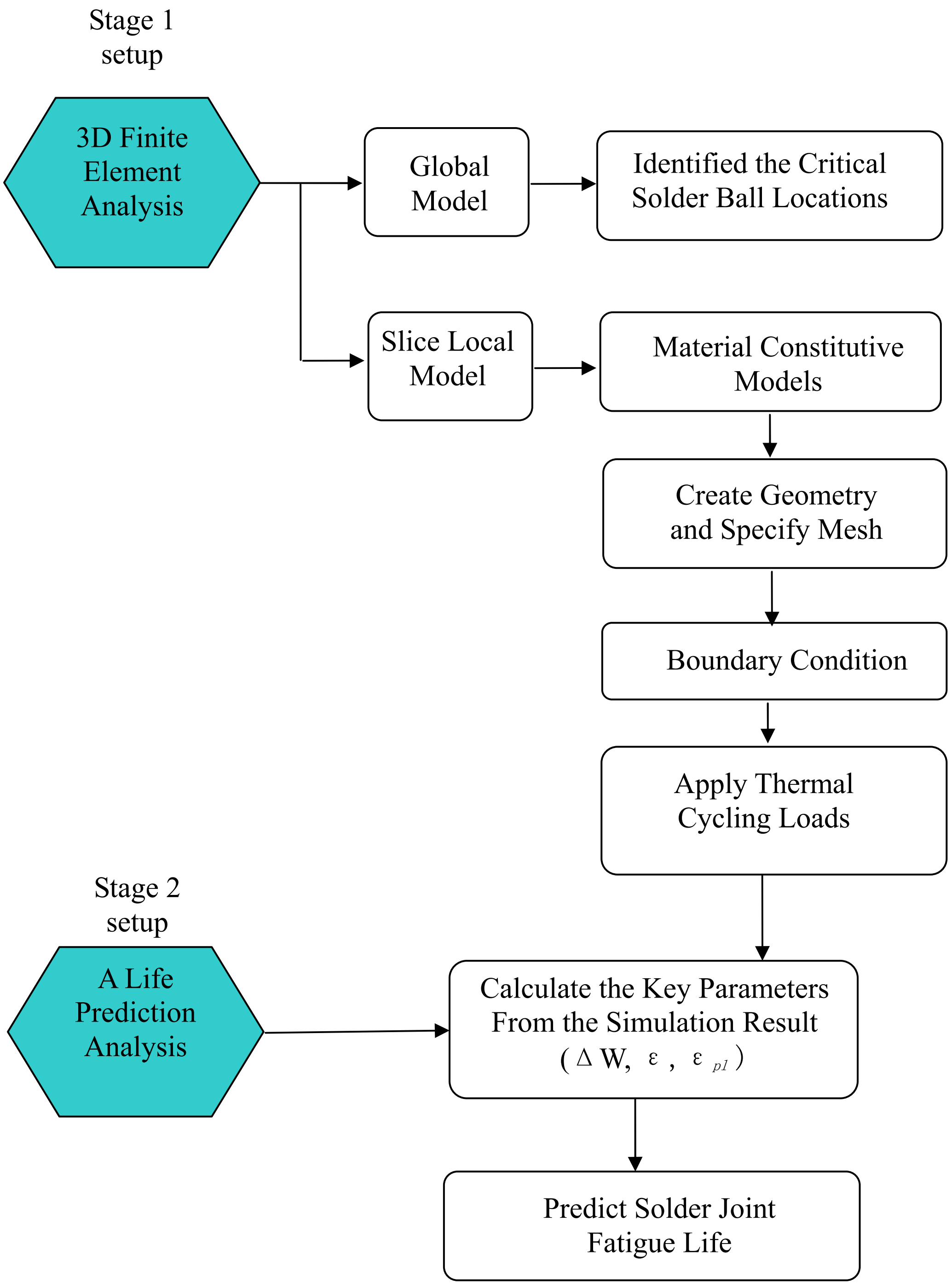


Figure 5.5 Flow chart outlining the steps in solder joint fatigue life prediction using the FEA method

5.2.1 Material Properties

A summary of the material properties used in the simulation is shown in Table 5.5. The silicon die, FR-4 and underfill were considered to be purely elastic. The silicon die was assumed to be isotropic, and the FR-4 substrate was considered to be orthotropic, while both are assumed to be temperature - dependent and elastic during the loading history. Most of these parameters are temperature-dependent. The CTE of the silicon increases when temperature increases as shown in Figure 5.6 plotted from the governing equation listed in Table 5.5. The modulus of the PCB decreases when temperature increases in all directions as shown in Figure 5.7 plotted from the governing equation listed in Table 5.5.

Table 5.5 Summary of material properties of model

Material	Elastic Modulus (MPa)	Shear Modulus (MPa)	Poisson's Ratio	CTE (1/K)
Silicon Die	162716		0.28	$-5.88x10^{-6}+$ $6.26x10^{-8}\ xT$ $-1.6x10^{-10}xT^{2}$ $+1.5x10^{-13}\ xT^{3}$
PCB	EX=27924-27xT EY=12204-16xT EZ=27924-27xT	GXY=5500-7.3xT GXZ=12600 -16.7xT GYZ=5500-7.3xT	0.39 (X) 0.11 (Y) 0.39 (Z)	CTE(X)=$16\ x10^{-6}$ CTE(Y)=$84\ x10^{-6}$ CTE(Z)=$16\ x10^{-6}$
Solder (creep)	30000@278K 27143@323K (see Table 5.3)		0.35	$24.5\ x10^{-6}$
Solder (Anand)	75842-152 x T		0.35	$24.5\ x10^{-6}$
Material	Elastic Modulus (MPa)	Glass Transition Temperature Tg (℃)	Poisson's Ratio	CTE (1/K)
Underfill (Fluxing)	2600 (< Tg) 20 (>Tg) [106]	70	0.35	$70\ x10^{-6}$
Underfill (Capillary)	7100	150	0.35	$22\ x10^{-6}$

*T=Temperature in Kelvin

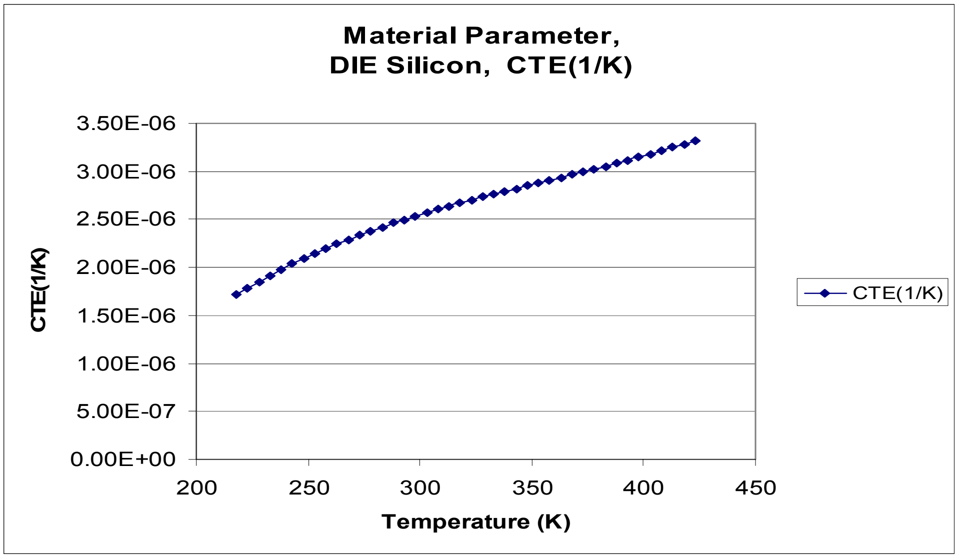


Figure 5.6 Temperature-dependent material properties for silicon die

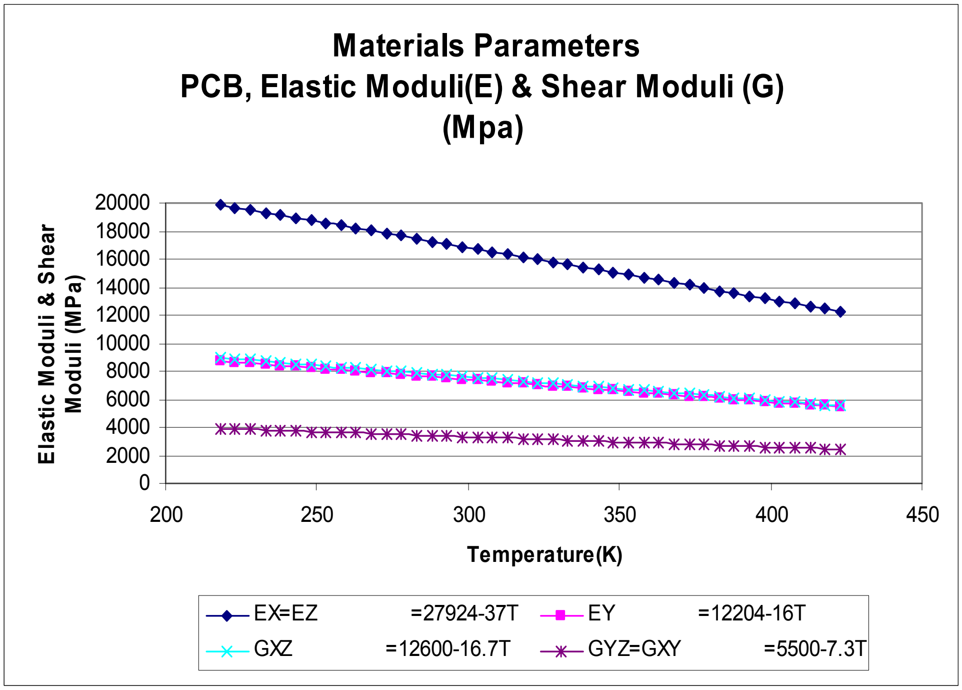


Figure 5.7 Temperature-dependent material properties for PCB

For the solder ball material properties, it is a much more complicated story. There is no commonly accepted constitutive law for describing the properties of solder materials. Furthermore, many of the sources of material data differ from one another. Among the various time-dependent constitutive laws for solder joints in electronic packages, two types of models are most commonly used. One is elastic-plastic-creep analysis and the other is viscoplastic analysis. These two different constitutive models were employed in this study to represent the inelastic deformation of solder. In elastic-plastic-creep analysis a separate constitutive model of time-independent plasticity and time-dependent steady-state creep is used, whereas in viscoplastic analysis, the Anand model is used.

A. Elastic-Plastic-Creep Constitutive Model

This model consists of a dual description divided into an elastic-plastic model and a creep model. The time independent elastic-plastic description is characterized by 6 parameters ($\sigma_1, \sigma_2, \sigma_3, \varepsilon_1, \varepsilon_2, \varepsilon_3$) as shown in Figure 5.8. The values for these six parameters for the solder used in this model were listed in Table 5.6.

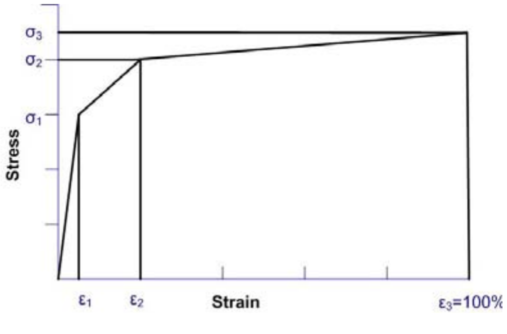


Figure 5.8 General elastic plastic Model

Table 5.6 Strain vs. stress data points for Sn-Pb eutectic solder [91, 105]

Temperature (K)	Strain	Stress (MPa)
278	7.00×10^{-4}	21
	3.00×10^{-3}	4
	1.00×10^{0}	600
323	7.00×10^{-4}	19
	3.00×10^{-3}	31
	1.00×10^{0}	400

Implicit creep, a multi-linear isotropic hardening rule (MISO) of ANSYS, was chosen to define the creep behavior of solder in this study. Its creep equation can be described as follows:

$$\dot{\varepsilon}_{creep} = C_1[\sinh(C_2\sigma)]^{C_3} e^{-C_4/T} \tag{5-19}$$

Table 5.7 Creep equation coefficients (time-dependent plasticity)

C1 (sec^{-1})	C2 (MPa^{-1})	C3	C4 (K)
10	0.2	2	5401.2

B. The Anand Viscoplastic Constitutive Model

This material model is available in the ANSYS material library (referred to as Anand's viscoplastic model). The governing equations are listed as follow:

Flow equation:

$$\frac{d\varepsilon_p}{dt} = A\left[\sinh\left(\xi\frac{\sigma}{s}\right)\right]^{\frac{1}{m}} \exp\left(-\frac{Q}{kT}\right) \tag{5-20}$$

Evolution equation:

$$\frac{ds}{dt} = \left\{h_0(|B|)^a \frac{B}{|B|}\right\} \bullet \frac{d\varepsilon_p}{dt} \tag{5-21}$$

where

$$B = 1 - \frac{s}{S^*} \tag{5-22}$$

with

$$s^{*} = \hat{s}\bullet\left[\frac{d\varepsilon_p / dt}{A}\exp(\frac{Q}{kT})\right]^{n} \qquad (5\text{-}23)$$

Darveaux determined the values of the constants in the Anand model for eutectic tin-lead solder by curve fitting a series of stress-strain data at various temperatures and strain rates [92]. The Anand model constants proposed in his study were used in this study, as shown in Table 5.8.

Table 5.8 Anand's model constants by Darveaux [92]

ANSYS	Constants	Value	Unit	Description
C1	$S_{\circ}$	12.41	MPa	Initial value of deformation resistance
C2	Q/k	9400	1/K	Activation energy / Boltzmann's constant
C3	A	4.00E6	1/sec	Pre-exponential factor
C4	ξ	1.50	dimensionless	Multiplier of stress
C5	M	0.303	dimensionless	Strain rate sensitivity of Stress
C6	$h_{\circ}$	1378.95	MPa	Hardening/softening constant
C7	$\hat{s}$	13.79	dimensionless	Coefficient of deformation Resistance saturation value
C8	n	0.07	dimensionless	Strain rate sensitivity of saturation (deformation resistance) value
C9	a	1.30	dimensionless	Strain rate sensitivity of hardening or softening

5.2.2 Geometry and Mesh

A diagonal slice nonlinear model was utilized in this study, as illustrated in Figure 5.9. This is a commonly accepted modeling technique, which uses only a diagonal slice of the assembly in order to reduce computation time. This technique is usually applicable to packages having octant symmetry. The slice passes through the thickness of the package, capturing all major components and a full set of solder joints.

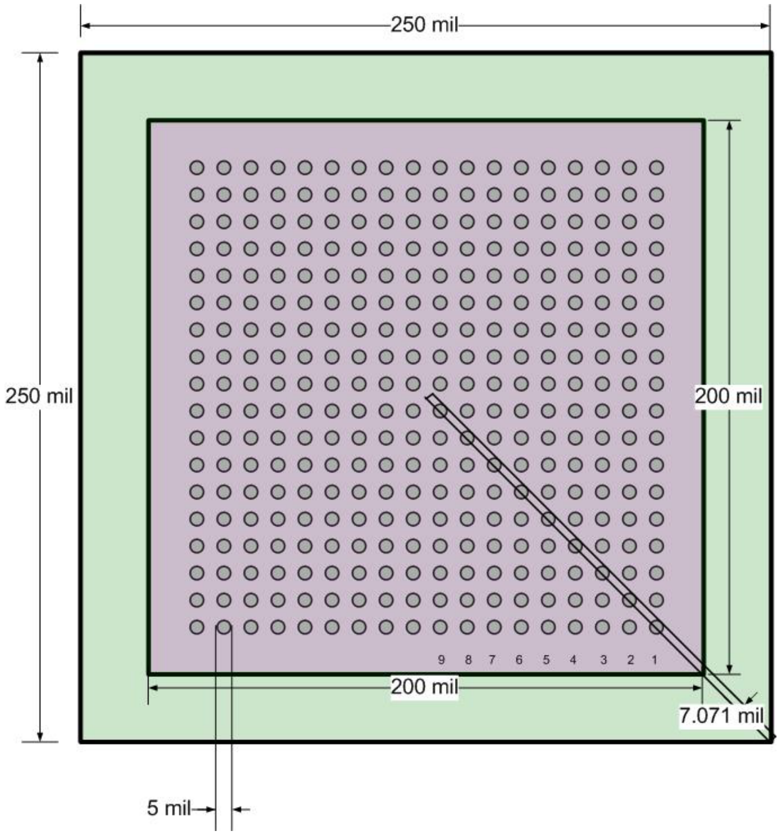


Figure 5.9 Top view of the flip chip package and the slice region represented by the 3D diagonal slice model is marked in the lower corner of the diagram

A. Geometry

The chip (FA10 2x2) having area array solder bumps with a 10 mil pitch used in this study was 200 mil x 200 mil in area and 20.87mil in thickness. The PCB was assumed to be 1.25 times bigger than the die size. All the UBM (under bump metallurgy), solder mask and copper traces were omitted for simplicity. All the component dimensions are listed in Table 5.8.

The outer most solder ball from the package center along the slice length was identified as the critical solder ball joint. As indicated in Figure 5.9, the diagonal slice model is considered to use the least amount of elements while still capturing the most critical solder joint.

Table 5.9 Model size parameters

	Parameters	**Global model size**	**Diagonal slice model size**
Die (Si)	Size	200mil x 200mil	$200\sqrt{2}$ mil x $5\sqrt{2}$ mil
	Thickness	20.87mil=530 micron *without solder ball height	20.87mil=530 micron
PCB (FR4)	Size	250mil x250 mil *defined as 1.25 x die size	$250\sqrt{2}$ mil x $5\sqrt{2}$ mil
	Thickness	31.9 mil = 830 micron	31.9ml=830micron
Solder Ball (Eutectic)	Diameter	5 mil	5 mil
	Height	3.15mil = 80micron	3.15mil
	Pitch	10 mil	$10\sqrt{2}$ mil

B. Mesh

The analysis must be applied on a contiguous solid model in order to obtain the correct result, which means that all interfaces between different volumes must have the same mesh definition and share common nodes. Otherwise, each volume will be analyzed independently without taking into account the boundary influence from adjacent areas or volumes.

A contiguous solid meshed model generally can be realized by two methods. The easier method is to build the model from the bottom up. This means that all the geometry volumes are defined first, "Glued" together and then free meshed simultaneously in one meshing action. The key here is to merge the node and key points of the model along the way in order to keep the solid and finite element model contiguous.

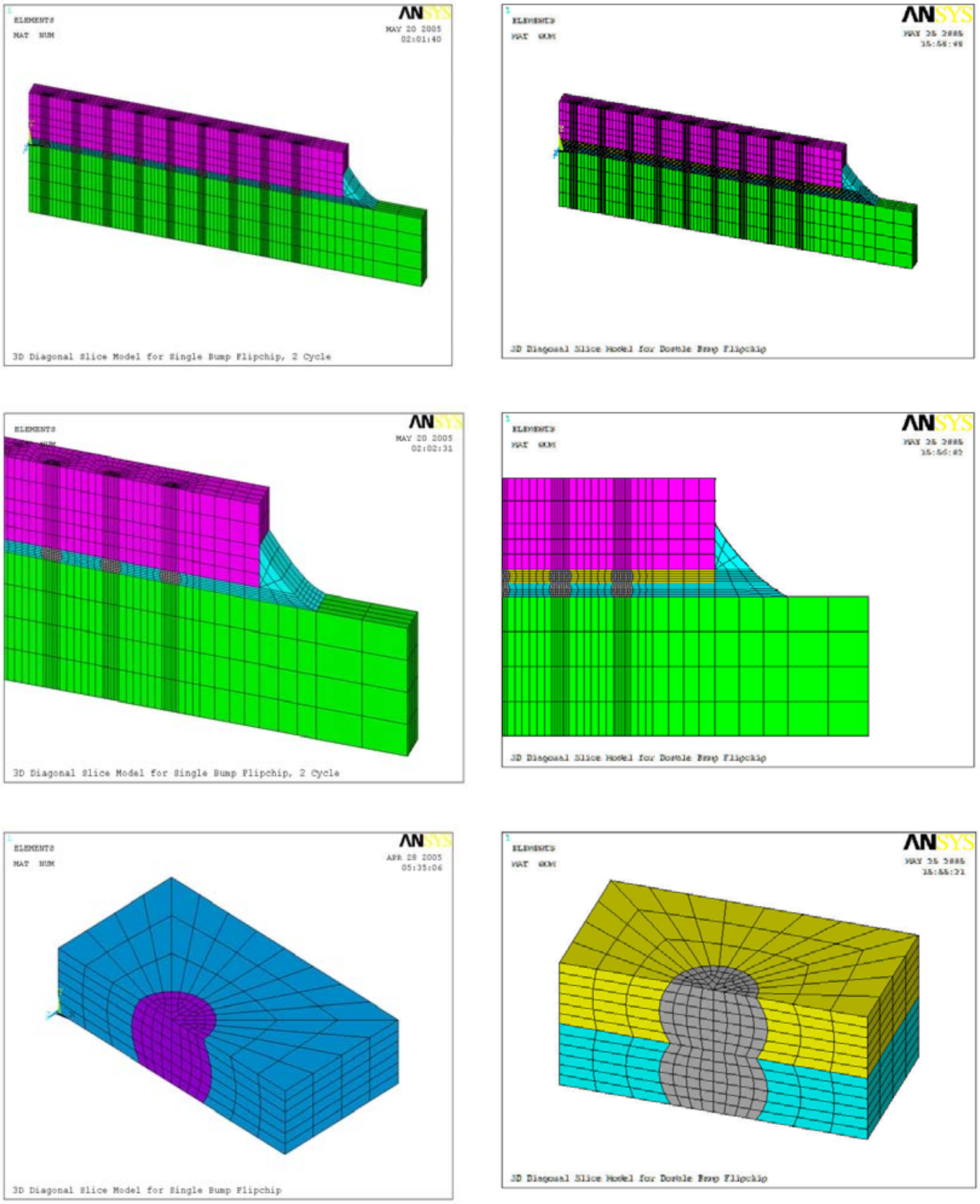
The other method is to build the model from top down. This requires the mesh to be created at the same time when the volumes were created by using the ANSYS "EXTRUDE" command or "VSWEEP" command. Both commands create the mesh from the adjacent area or volume which guarantees a coincident node and mesh at the interfaces of different volumes.

In a 3D model, the free mesh method can only be used to mesh volumes with a tetrahedral element shape, while a mapped mesh can only mesh volumes with hexahedral element shape. The preferred element shape is a hexahedral element, otherwise the model is too stiff. Therefore, creating a mapped meshable volume, or VSWEEP-able volume, is the only reasonable solution. The other advantages of choosing the second method over the first one are: (i) it will provide better accuracy in the results by eliminating any differences caused by different element orientations and sizes, especially when it comes to comparing the single bump and double bump models or comparing different solder balls in the same model; (ii) The complete control of mesh structure that is possible in the "top down" method allow it to generate a finer mesh at the locations of most interested and coarse mesh at the locations with less importance. This control of the "top down" method makes it possible to reduce the number of elements and nodes dramatically compared to the free mesh method.

Considering all the reasons mentioned above, the general steps needed to generate the model can be listed as follows:

- ✓ Create the cross-section area of solder ball geometry, and mapped mesh the area
- ✓ Create a single solder ball by rotating the meshed cross-sectional area of solder ball.
- ✓ Consider defining a "unit cell" containing one solder joint and the associated materials above, below or surrounding the joint (such as the chip, underfill, PCB)
- ✓ Once the unit cell is meshed, copy the unit cell multiple times to form a solder ball array.
- ✓ Add necessary additional geometry features, such as underfill fillet volume

(a) Single Bump flip chip Mesh (b)Double Bump Flip Chip Mesh



Single bump solder ball and surrounding underfill

Double bump solder ball and surrounding underfill with upper layer being capillary underfill and lower layer being fluxing underfill

Figure 5.10 Three-dimensional diagonal slice model with mapped finite-element mesh for (a) regular single bump flip chip assembly; (b) double bump flip chip assembly

Table 5.10 Elements and nodes statistics for two models

	Single bump flip chip model	Double bump flip chip model
Overall Node number	16053	22071
Overall Element number	13388	18740
DIE	4400 elements	4400 elements
PCB	3600 elements	3600 elements
Solder joints	288 elements/joint x 9 sets	576 elements/joint x 9 sets
Underfill	2688 elements	5376 elements
Underfill Fillet	108 elements	180 elements

Figure 5.10 shows the finished mesh for the single bump and double bump flip chip assemblies. The mesh for the solder balls and the surrounding underfill was constructed to be much finer than that of the die and the PCB component due to the interest of the solder joint fatigue study. As summarized in Table 5.10, the overall element number was 13388 for the single bump flip chip model and 18740 for the double bump flip chip model. Finer mesh can be achieved by adding more divisions, but meanwhile the computing time will also increases when the element number increases.

5.2.3 Boundary Conditions

The model imposes symmetric boundary conditions on the slice plane coinciding with the true symmetry plane. The other slice plane is neither a free surface nor a true symmetry plane, but was defined to be parallel to the symmetry plane by coupling all the nodes together on this plane. In summary, five boundary conditions were applied to the 3D-model, as shown in Figure 5.11.

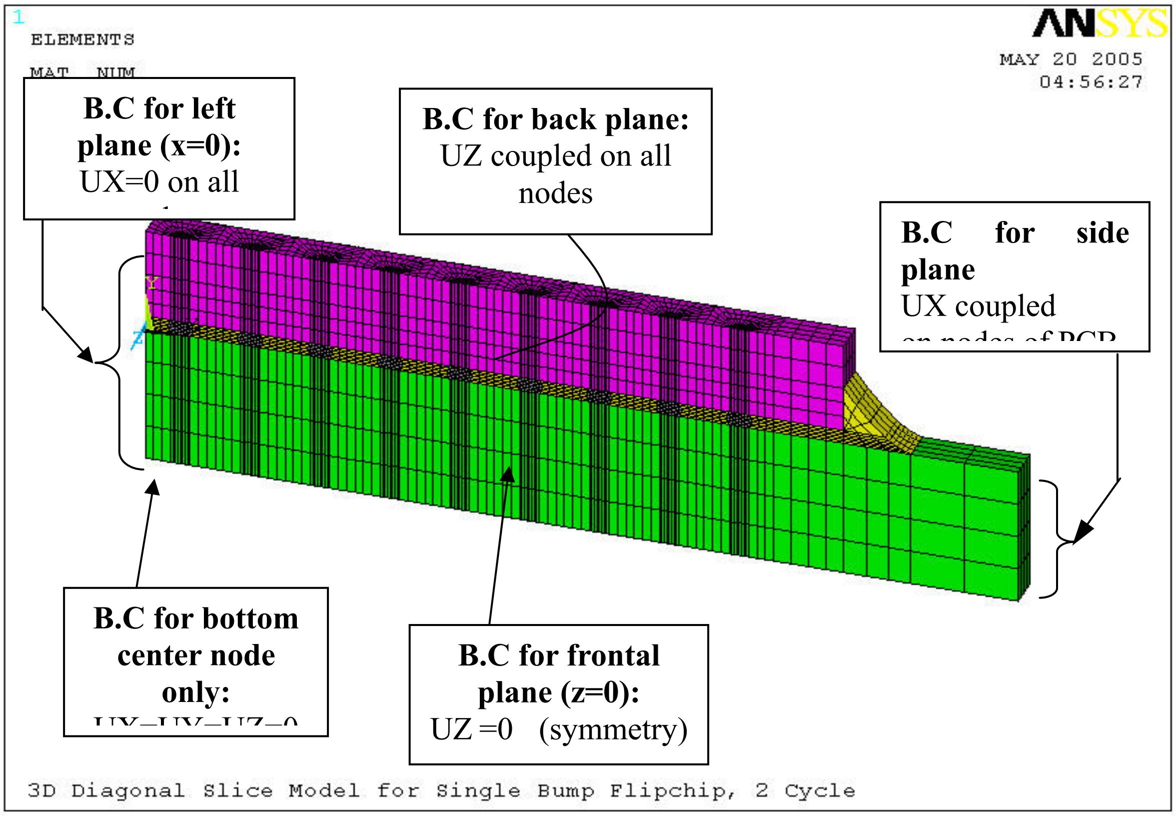


Figure 5.11 Boundary conditions applied to a typical diagonal slice model

5.2.4 Thermal Loading

The temperature profile in the thermal cycle used for the modeling consists of four stages as shown in Figure 5.13: temperature ramping to low extreme in 2 minutes; temperature dwelling at low extreme for 13 minutes; temperature ramping to high extreme in 2 minutes; temperature dwelling at high extreme for 13 minutes. The temperature range was from -40°C to 125°C. The zero strain reference temperature is set at the high temperature. The cycling began with the ramping down stages.

The temperature was assumed to increase linearly and the temperature loads were applied to the model elements uniformly at any given moment of the loading. The slice model was then executed for a steady-state structural analysis using the default frontal solver.

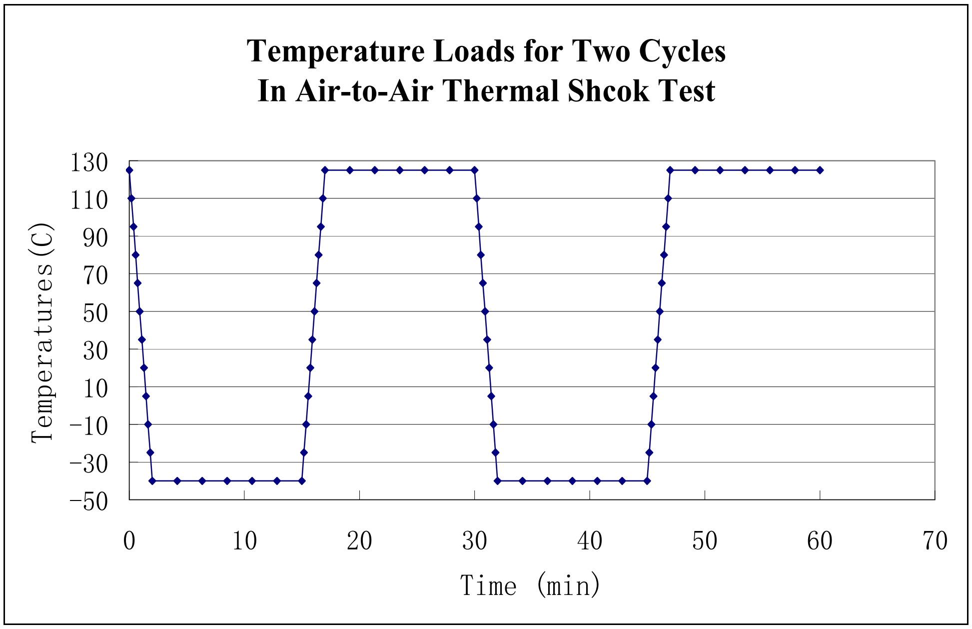


Figure 5.12 Temperature loads for two cycles in air-to-air thermal shock test

In this analysis, one substep was used for every 15 degree temperature change in a thermal ramp. Thus, a temperature ramp from -40 ℃ to 125 ℃ would be divided into 11 substeps for the purposes of the computation. In the case of non-converging computation due to large deformation, the substeps interval should be changed to a smaller value. For the temperature dwell stages, the number of substeps was decided by the ANSYS program automatically, which selected 6 substeps for each 15 minutes of dwell time. The following sequence of APDL commands were used to simulate the first thermal cycle:

```
!-------Cycle 1 Load 1, ramp from high temperature to low temperature----------

AUTOTS, OFF                    !turn off auto time step
NSUBST, rampstep               !set substeps for load
BFE, ALL, TEMP, , LowT         !apply temperature to nodes
KBC, 0                         !linearly ramp temperatures
TIME, lowramp                  !set time at end of this load step
Solve

!--------Cycle 1 Load 2:   dwell at low temperature-----------------

AUTOTS, ON                     !turn on auto time step
NSUBST, 10,100, 1              !set time substep
BFE, ALL, TEMP, , LowT         !apply temperature to nodes
KBC, 1                         !linearly ramp temperatures
TIME, lowramp+lowdwell         !set time at end of this load step
Solve

!-------Cycle 1 Load 3, ramp from low temperature to high temperature----------

AUTOTS, OFF                           !turn off auto time step
NSUBST, rampstep                      !set substeps for load
BFE, ALL, TEMP, , highT               !apply temperature to nodes
KBC, 0                                !linearly ramp temperatures
TIME, lowramp+lowdwell+highramp       !set time at end of this load step
Solve

!--------Cycle 1 Load 4:   dwell at high temperature-----------------
AUTOTS, ON                            !turn on auto time step
NSUBST, 10,100, 1                     !set time substep,
BFE, ALL, TEMP, , highT               !apply temperature to nodes
KBC, 1                                !linearly ramp temperatures
TIME, lowramp+lowdwell+highramp       !set time at end of this load step
Solve
```

All the words in lower case in the above APDL commands, including rampstep, lowT, highT, lowramp, lowdwell, highramp, highdwell were parameter variables defined previously according to the experimental environment setups. They were defined as program parameters for easy alternation of thermal cycling conditions. To ensure the calculation was accurate and achieved stable results, these four loads were repeated to complete the calculation for two consecutive thermal cycles.

5.2.5 Numerical Simulation

The simulations were carried out on both a Sun station server and a Windows PC. The computing times for simulations of a single bump flip chip model for two thermal cycles were typically 5-8 hours on both machines. The double bump flip chip model required more computing time since it had more elements than that of single bump flip chip model.

For materials other than solder, element type SOLID45 was chose. The SOLID45 element is defined by eight nodes and has three degrees of freedom at each node. It has plasticity, creep, swelling, stress stiffening, large deflection, and large strain capabilities.

For the solder material, element type VISCO107 was chosen for solder defined with the Anand model, while element type SOLID185 was used for solder defined with elastic-plastic-creep model. The VISCO107 element is defined by eight nodes and has three degrees of freedom at each node. The element is designed to solve both isochoric rate-independent and rate-dependent large strain plasticity problems.

The simulation results are introduced in the order of stress analysis, strain analysis and plastic work analysis. The outermost solder ball in the model was selected for further failure prediction considerations. Different solder fatigue life prediction methodologies were used to predict the life for both single bump flip chips and double bump flip chips.

5.2.6 Package Deformations

The dotted line in Figure 5.13 represents the original shape of the package before the thermal loading. Because the package was assumed to be stress free at the highest temperature (125℃), the temperature change from high to low temperature resulted in shrinkage in all components based on their individual own CTEs. The maximum deformation occurred at the low temperature dwell stages (-40℃). Since the silicon die had the lowest CTE compared to all the other materials (solder, PCB, underfill), the silicon die's shrinkage was much smaller than the shrinkage of the PCB, solder and underfill. Therefore a downward deformation of the whole package appeared after two thermal cycles, as shown in Figure 5.13, resulting in the warpage of package.

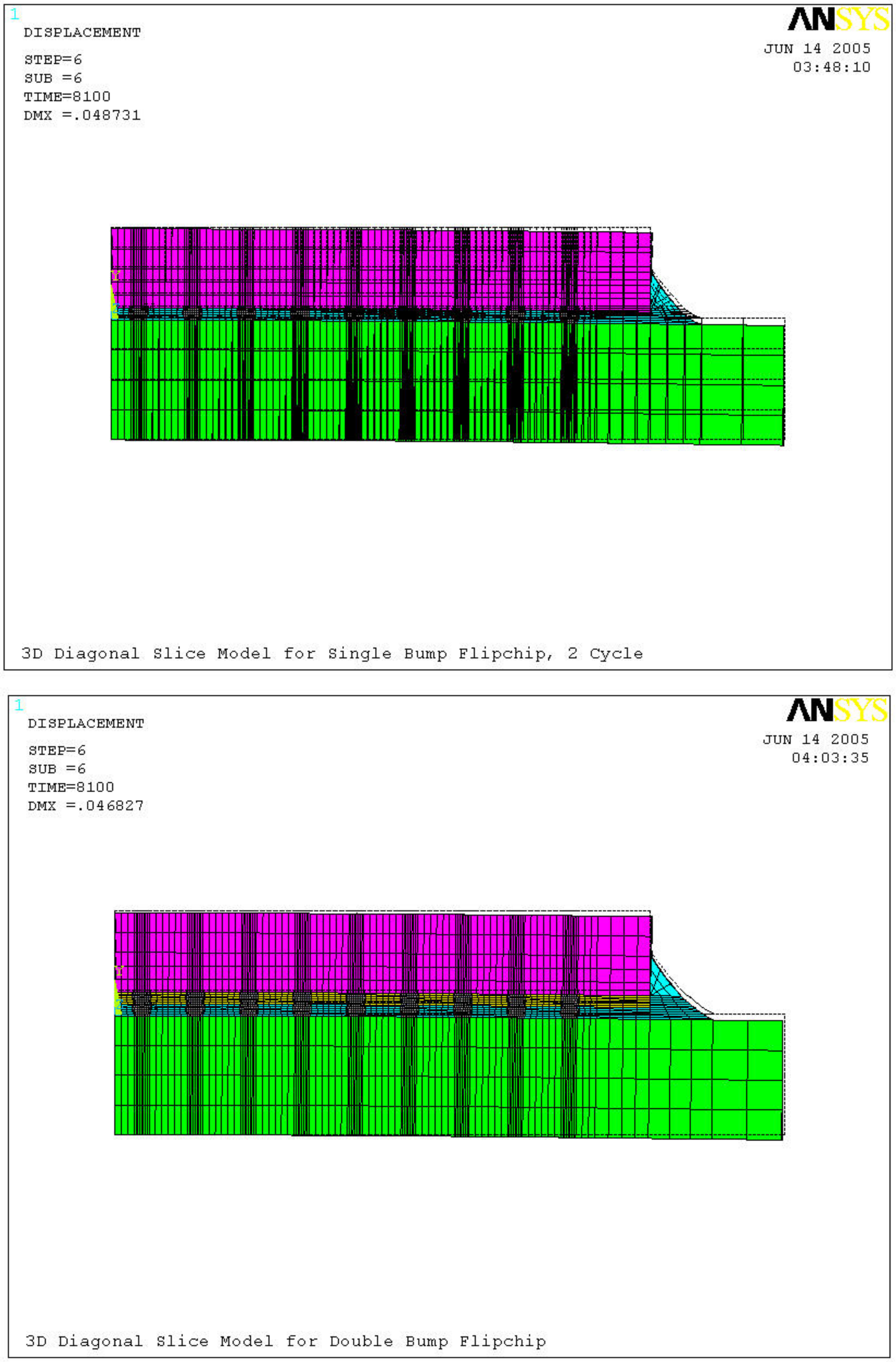


Figure 5.13 The deflection of the package due to the CTE mismatch: (a) single bump flip chip model; (b) double bump flip chip model

5.2.7 Stress Distribution Among the Solder Joints

At all moments of the thermal loading cycle, the highest Von Mises stress was found on the outmost (9th) solder joint for both the single bump and double bump models. The Von Mises (effective) stress distributions in the solder joints are shown in Figure 5.15. It can be seen that the maximum Von Mises stress in the single bump models appears at the 9th solder ball upper corner while it appears at the lower corner of the 9th bump for the double bump model. This large stress is due to the significant thermal expansion mismatch between the silicon die (2.8 ppm/℃), capillary flow underfill (22 ppm/℃), fluxing underfill (70 ppm/℃) and solder joint (24.5 ppm/℃). This large stress decreases rapidly to a much smaller value only a short distance away from the bottom of the chip towards the top of the chip.

Local stress levels were calculated at the different layers for comparison. For the individual solder bump, the single bump model had the top layer and bottom layer, and the double bump model had the top layer, middle 1 layer, middle 2 layer and bottom layer as shown in Figure 5.14. Each layer contains 48 elements and had a thickness of 13.33µm. All the calculations of the local stress, strain or plastic work were averaged over the elements among each layer.

Figure 5.16 shows how the local stress level at the sliced layers changed during the cycling for both single bump and double bump models. They both reveal similar behavior, which is high stress level when the package dwells at the low temperature and the stress level was returned to a low level when the package dwells at the high temperature. This is because the high temperature extreme was defined as the reference temperature, which means the package was defined to be stress free at the high temperature extreme. Comparing the stress level in the single bump model (Figure 5.16 (a)) to that of the double bump model (Figure 5.16(b)), it is clear that the solder joint in the single bump flip chip experiences higher levels of stress than that of a double bump flip chip.

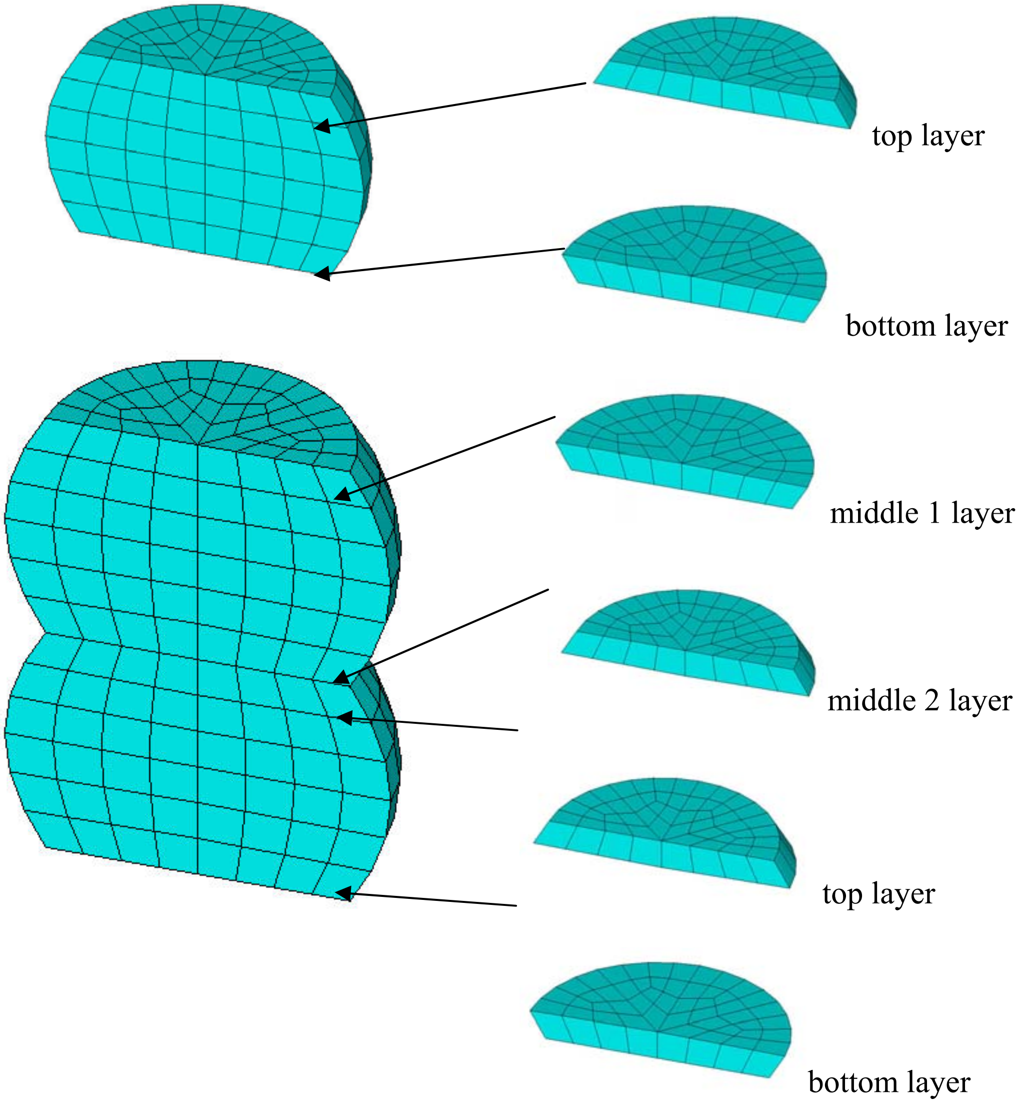
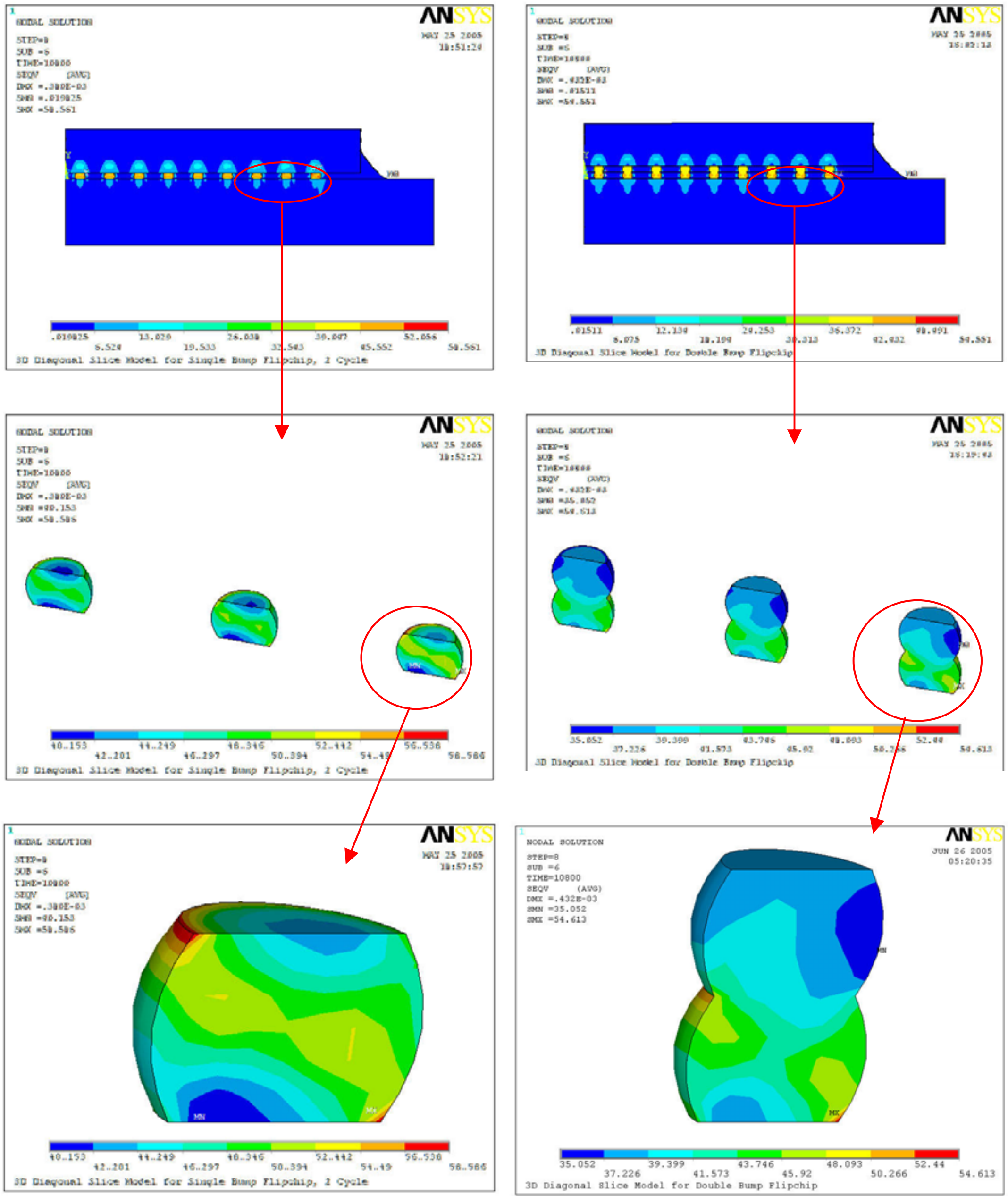


Figure 5.14 Sliced layers for calculation



(a)9th single bump

(b) 9th double bump

Figure 5.15 Von Mises stress (MPa) distribution in the solder joint for (a) single bump model and (b) double bump model

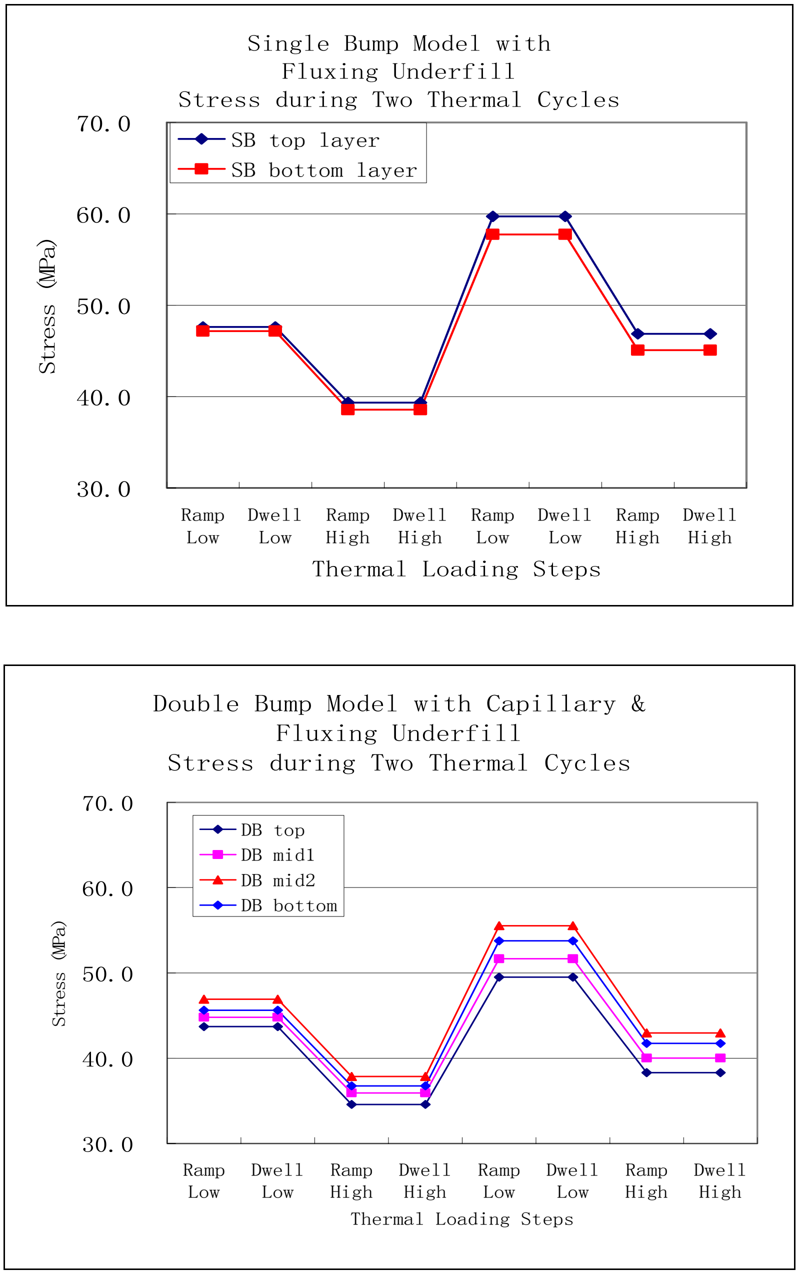


Figure 5.16 Von Mises stress changes during two thermal cycles

Referring to Figure 5.17, the values for Von Mises stress were almost the same for the center solder ball No.1 to solder ball No.5 for both the single bump model and double bump model. Then it increased when the distance to the neutral point increased. The 9th solder ball has the highest stress of 73.7 MPa (max. stress value) for the single bump model and 68.8 MPa (max. stress value) for the double bump model during the dwell stage at the lower temperature extreme.

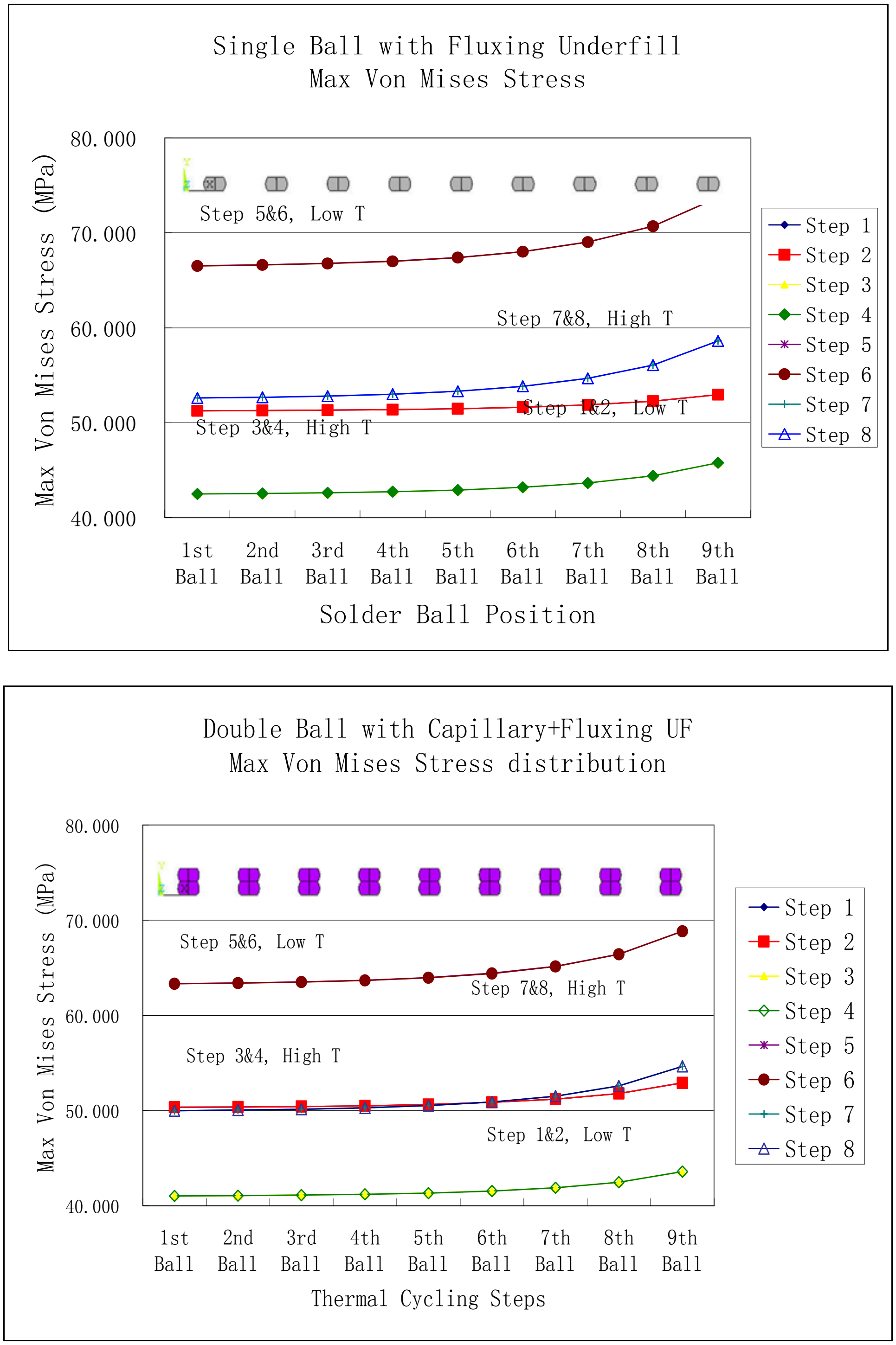


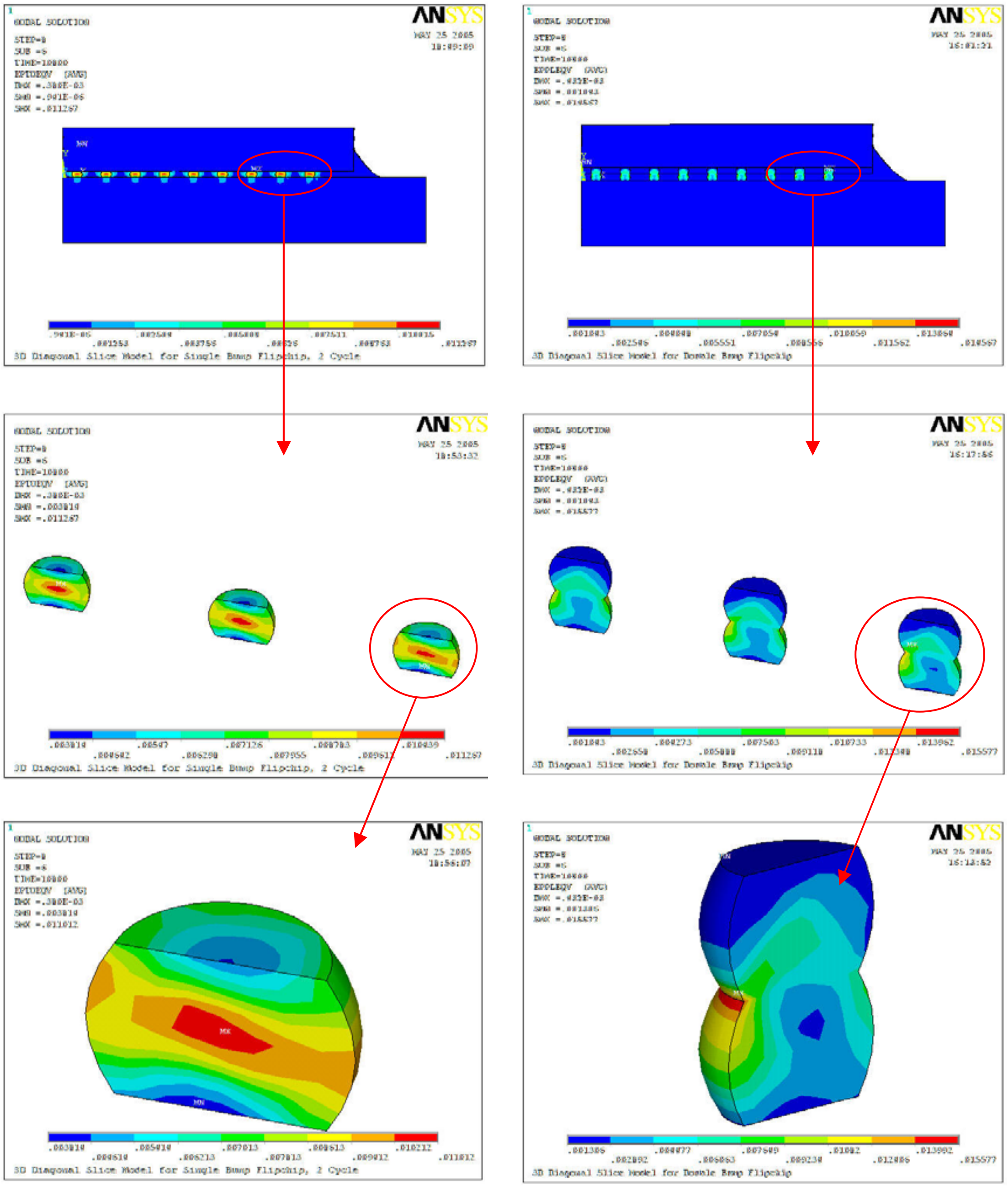
Figure 5.17 Maximum Von Mises stress value in nine solder balls of the (a) single bump model and (b) double bump model

5.2.8 Strain Distribution in the Solder Ball

The results shown in Figure 5.18(a) indicate the largest strain is in the center of the solder ball for the single bump solder ball structure, while Figure 5.18(b) indicates that the maximum strain appears at the upper ball and lower ball interface for the double bump solder ball structure. This is probably because of the different CTE of the upper layer underfill and lower level underfill.

Figure 5.19 shows a plot of the strain, plastic strain and plastic shear strain in the middle layer of the double bump model during the 2 complete thermal cycles. All three strain lines revealed the same trend during the thermal cycle, namely high at low temperature and low at reference (high) temperature. Comparing the values of the strain and plastic strain during all stages; most of the strain was in the form of plastic strain.

Comparing Figure 5.17 and Figure 5.20, the strain distribution is relative similar to the stress distribution among the 9 solder balls, the values for plastic strain stays the same for the balls located in the center area of the die (ball #1 to ball #5), and gradually increases when the DNP (distance to neutral point, i.e., distance to center of the die) increases (ball #6 to ball #9).



Von Mises Strain

Von Mises Strain

Figure 5.18 Strain distribution in single bump and double bump models

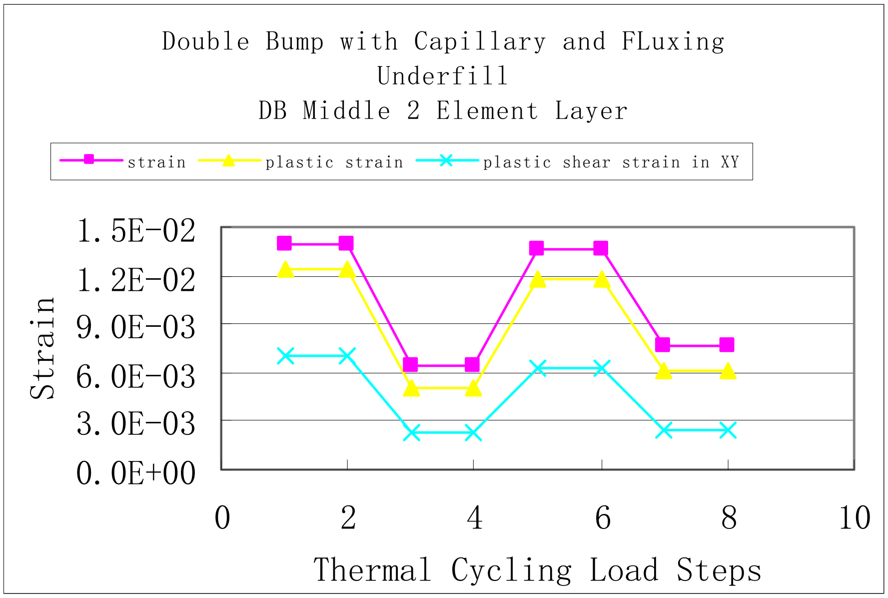


Figure 5.19 Von Mises strain changes during the 2 thermal cycles

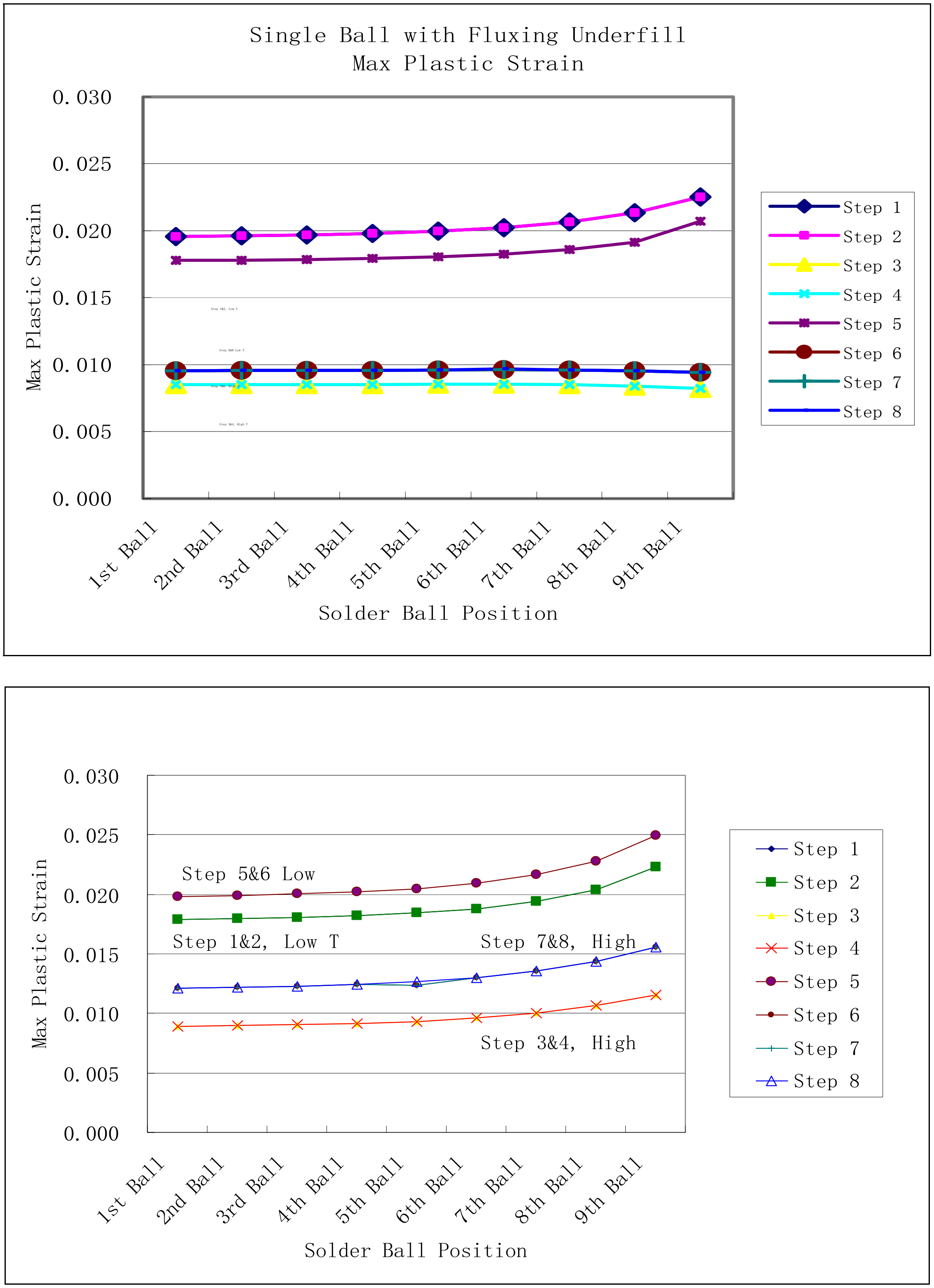


Figure 5.20 Maximum plastic strain values in nine solder balls of the single bump model and double bump model

5.2.9 Plastic Work Distribution in the Solder Ball

The plastic work distribution, as shown in Figure 5.21, was concentrated on the upper rim and lower rim of the solder joint interface. Therefore, the volume-averaged plastic work density was calculated for both the upper element and lower layer element layers, while two additional middle element layers were selected for calculation for the double bump model. The element layer thickness was chosen to be about 0.5mil.

The value of plastic work was averaged over the elements in the upper, bottom or middle layers of the solder ball using the following equation:

$$\Delta W_{avg} = \frac{\sum_{i=1}^{n} \Delta W_i \cdot V_i}{\sum_{i=1}^{n} V_i} \qquad (5\text{-}24)$$

W_{avg} were calculated at the end of every step of the thermal loading. To obtain the ΔW_{avg} per thermal cycle, W_{avg} at cycle 1 (step 4) was subtracted from cycle 2 (step 8).

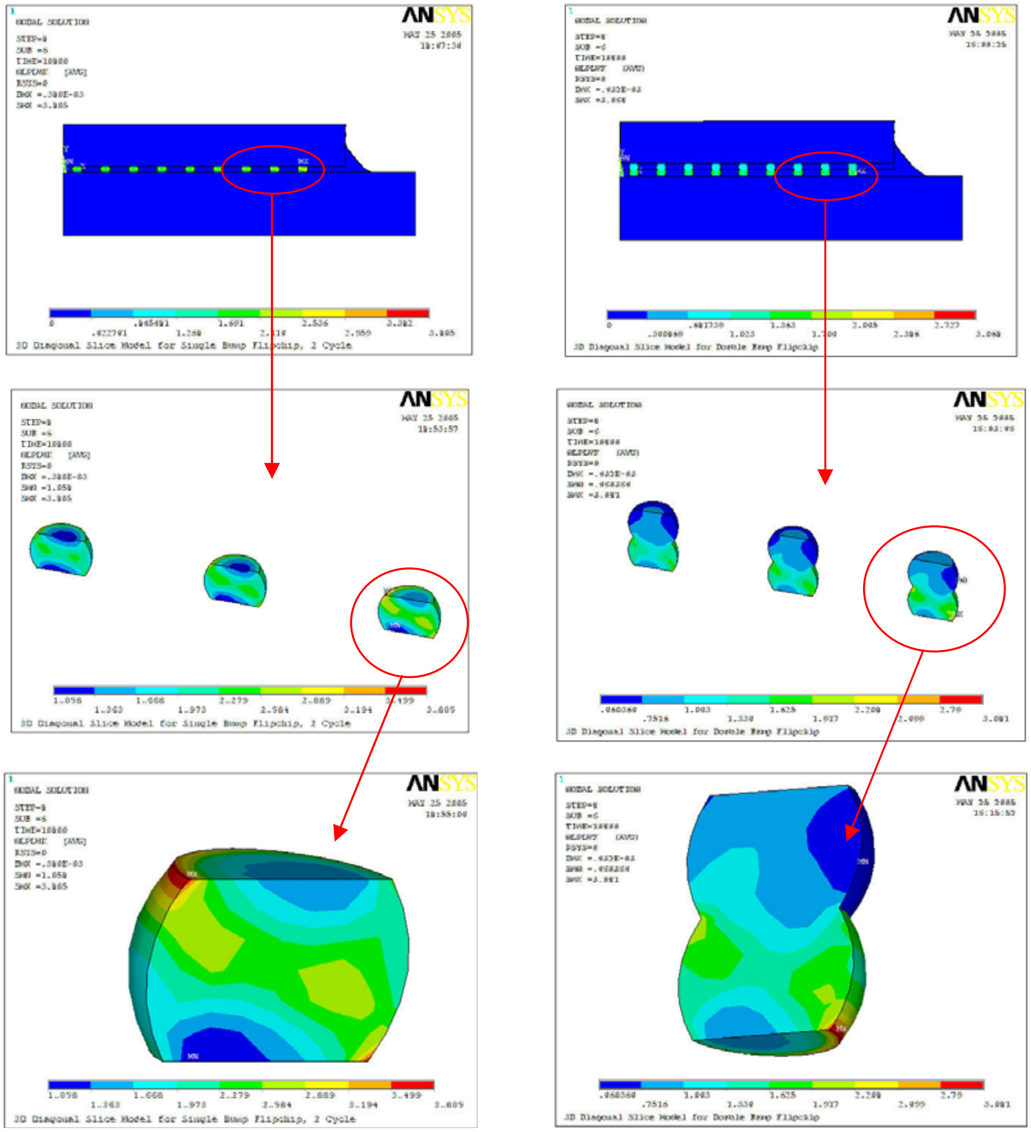


Figure 5.21 Accumulated plastic work/ volume distribution in the single bump and double bump models after two thermal cycles

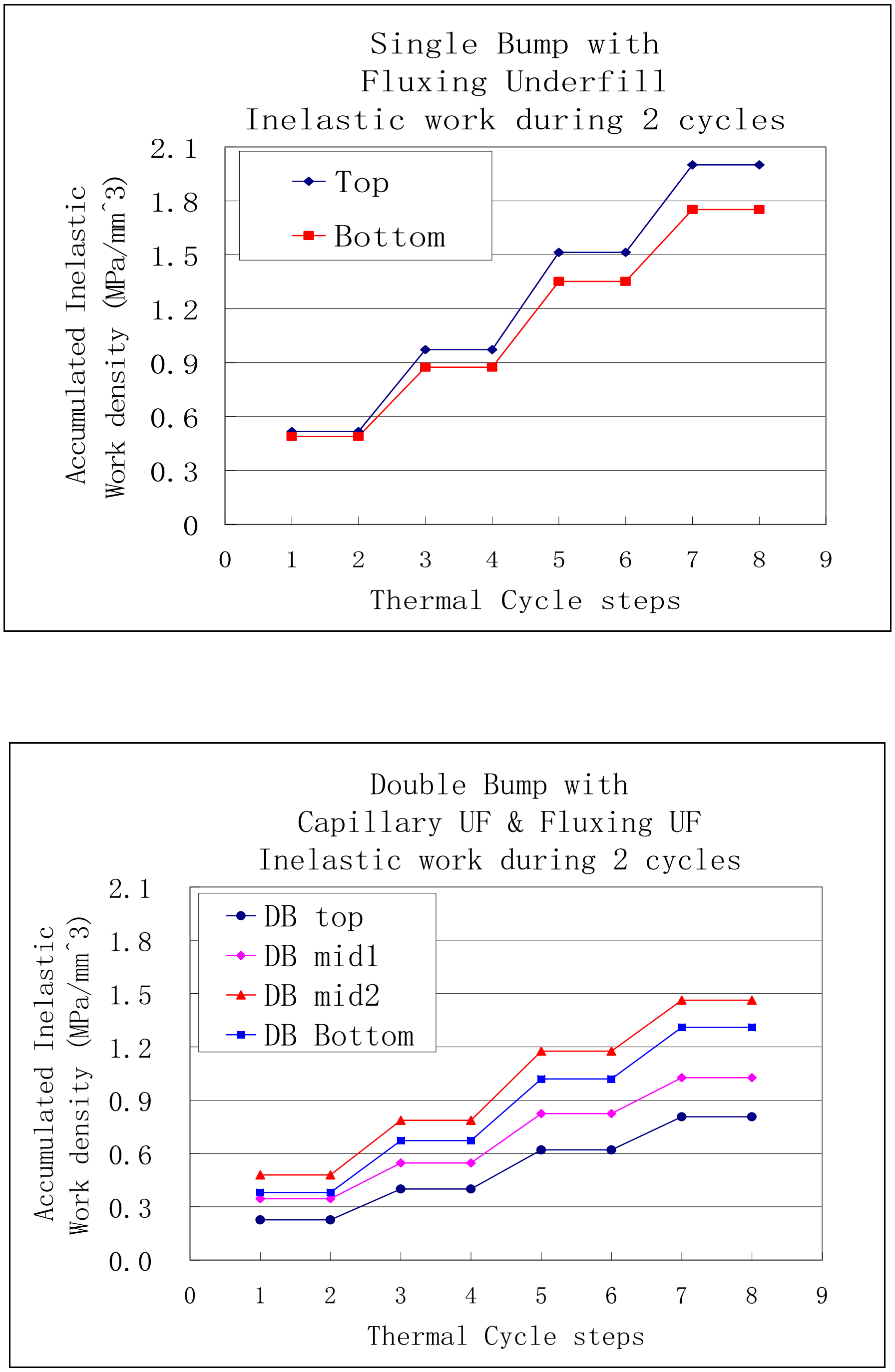


Figure 5.22 Accumulated Inelastic energy density for (a) top/bottom layer in single bump model; (b) top/mid/bottom layer in double bump model

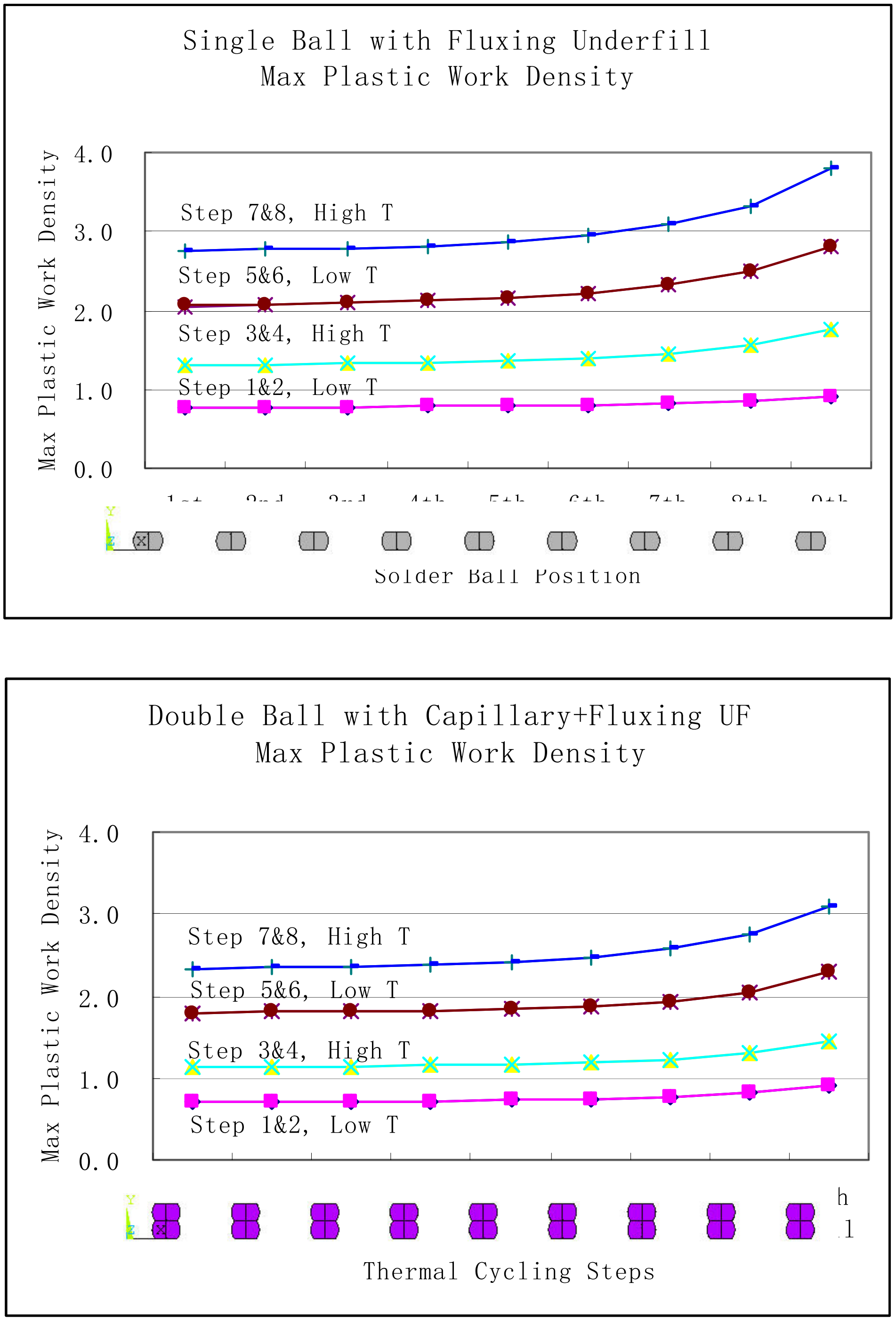


Figure 5.23 Maximum plastic work/volume value for nine solder balls in (a) single bump model and (b) double bump model

Again, as for the stress and strain distribution in the solder joints, the inelastic energy density distribution in the solder joints appeared to be almost the same for the first 5 solder joints, which were located closer to the center of the package (Figure 5.23). The further the distance between the solder joint and the center of package, the higher the inelastic energy density. The outmost (9th) solder joint appeared to accumulate the highest energy density in both the single and double bump models.

5.2.10 Thermal Fatigue Life Prediction of Solder Joints

In this section, several different thermal fatigue life prediction models of solder joint will be applied to the simulation results for both the single bump and double bump geometries. The predicted fatigue life for single bump and double bump flip chip models were compared and the improvement factor was calculated according to each method.

5.2.10.1 Calculation of the Failure Parameters

Based on analysis of the previous simulation results, the outmost solder joint, the 9th solder ball in this study, was used for quantitive analysis since it was the worst case solder joint. For each model, the interface element layers were calculated in order to obtain the failure parameters according to each prediction model. For both the single bump and double bump geometries, the upper element and bottom element layers were calculated, while two additional middle layers between the two stacked solder balls in the double bump model were also calculated.

According to the literature review of prediction models given in the previous section, stress range, inelastic energy density, strain range, plastic strain range and shear strain range are the possible failure parameters that are likely to be needed in order to predict solder joint life. These parameters were calculated and the results are listed in Table 5.11, Table 5.12, Table 5.13 and Table 5.14.

The results shown in Table 5.11 and Table 5.13 indicate that the top interface layer was the most vulnerable location for fatigue for the single bump flip chips, while it was the mid interface layer in the lower ball for the double bump flip chip model. Hence, the failure parameters calculated for the top layer should be used in the life time predictions for single bump flip chips, and the failure parameters calculated for the middle layer of the lower ball should be used for the life time prediction of double bump flip chips. These trends were the same for models using either the Elastic-Plastic-Creep constitutive model or the Anand constitutive model for solder. Anand models showed a much higher stress range and strain range, but lower inelastic energy density compared to the Elastic-Plastic-Creep models.

Table 5.11 Simulation results for single bump flip chip using elastic-plastic-creep constitutive model

	Stress range ($\Delta\sigma$) (MPa)	Inelastic energy density (ΔW_{avg}) (MPa/mm^3)	Von Mises Strain range ($\Delta\varepsilon$)	Plastic Strain range ($\Delta\gamma_p$)	Shear Strain range ($\Delta\gamma$)
Top layer	**12.84**	**1.028**	**8.287E-3**	**8.005E-3**	**8.807E-3**
Bottom layer	12.67	0.877	8.056E-3	7.874 E-3	8.159E-3

Table 5.12 Simulation results for the double bump flip chip using elastic-plastic-creep constitutive model

	Stress range ($\Delta\sigma$) (MPa)	Inelastic energy density (ΔW_{avg}) (MPa/mm^3)	Von Mises Strain range ($\Delta\varepsilon$)	Plastic Strain range ($\Delta\gamma_p$)	Shear Strain range ($\Delta\gamma$)
Upper Ball Top layer	11.19	0.406	4.719E-3	4.466E-3	2.804E-3
Upper Ball Bottom layer	11.66	0.479	3.891E-3	3.617E-3	3.819E-3
Lower Ball Top layer	**12.6**	**0.675**	**5.960E-3**	**5.670E-3**	**5.688E-3**
Lower Ball Bottom layer	12.0	0.638	5.662E-3	5.390E-3	5.413E-3

Table 5.13 Simulation results for the single bump flip chip using Anand constitutive model

	Stress range (Δσ) (MPa)	Inelastic energy density (ΔW_{avg}) (MPa/mm^3)	Von Mises Strain range (Δε)	Plastic Strain range ($\Delta\gamma_p$)	Shear Strain range (Δγ)
Top layer	**48.33**	**0.530**	**14.39E-3**	**13.412E-3**	**10.46E-3**
Bottom layer	46.95	0.466	13.94E-3	12.965E-3	10.28E-3

Table 5.14 Simulation results for double bump flip chip using Anand constitutive model.

	Stress range (Δσ) (MPa)	Inelastic energy density (ΔW_{avg}) (MPa/mm^3)	Von Mises Strain range (Δε)	Plastic Strain range ($\Delta\gamma_p$)	Shear Strain range (Δγ)
Upper Ball Top layer	42.46	0.219	7.758E-3	6.951E-3	4.034E-3
Upper Ball Bottom layer	42.43	0.297	10.52E-3	9.648E-3	5.803E-3
Lower Ball Top layer	**44.93**	**0.408**	**12.65E-3**	**11.71E-3**	**8.574E-3**
Lower Ball Bottom layer	44.64	0.351	10.58E-3	9.667E-3	8.565E-3

5.2.10.2 Fatigue Life Predictions

Five prediction models were selected from the literature review to predict the solder joint fatigue life for both single bump and double bump flip chips. The first two methods, Engelmaier's and Solomon's, are strain based approaches, and the other three approaches, Zahn's, Shi's and Darveaux's, are energy-based approaches. The governing equations and parameters are listed in the Table 5.15, and the solder joint fatigue lives predicted based on these models are listed in Table 5.16.

Table 5.15 Summary of prediction models and parameters used in this study

Category	Failure parameter	Equations	Reference
Strain -based	$\Delta\gamma$ shear strain range	$\overline{N}_f = \frac{1}{2}\cdot\left(\frac{\Delta\gamma}{2\varepsilon'_f}\right)^{\frac{1}{c}}$ ε'_f = 0.325 for eutectic solder C = -0.49 for AATS test (-125 to 40 C, 30min cycle)	Engelmaier [99], [100]
	$\Delta\varepsilon_{plastic}$ plastic strain range	$N_f = \theta(\Delta\gamma_p)^{\phi}$ θ =1.292 for eutectic solder φ =-1.96 for eutectic solder	Solomon [98]
Energy -based	ΔW_{avg} (MPa)	$N = C_1(\Delta W_{avg})^{C_2}$ C1 =1098.3 C2 =-0.8827	Zahn [105]
	$\Delta W_{avg,}$ (MPa) ΔStress (MPa)	$\left[N_f v^{(k-1)}\right]^m \left[\frac{\Delta W}{\Delta\sigma_f}\right] = C$ C=1.69 , m=0.7 (k-1)=-0.1 for v≥10^{-3} Hz v=1/1800 Hz	Shi [102]
	ΔW_{avg} (psi)	$N_0 = K_1\Delta W_{ave}^{K_2}$ $\frac{da}{dN} = K_3\Delta W_{ave}^{K_4}$ $N_f = N_0 + \frac{a}{da/dN}$ K1=22400, K2=-1.52 K3=2.76E-7, K4=0.98	Darveaux [92]

Table 5.16 Summary of prediction using different models.

Model References	failure parameter	Prediction for Elas-Plas-creep Model		Prediction for Anand Model	
		Predicted Charactertic N_f	Improve -ment	Predicted Charactertic N_f	Improve -ment
Engelmaier	$\Delta\gamma$ Shear strain range	Single: 2724 Double:6529	140%	Single: 1931 Double:2873	49%
Solomon,	$\Delta\gamma_p$ Plastic strain range	Single: 16632 Double: 25867	56%	Single: 6050 Double:7892	30%
Darveaux	ΔW_{avg}	Single: 61.7 Double: 97.5	58%	Single: 127 Double: 170	33%
Zahn's Power law	ΔW_{avg}	Single: 1070 Double:1494	40%	Single: 1923 Double:2423	26%
Shi's	ΔW_{avg} $\Delta\sigma$ (stress)	Single: 55.3 Double: 98.1	77%	Single: 947 Double:1240	31%

These simulation results were compared to the air-to-air thermal shock reliability test results under the test chamber profile plotted in Figure 5.24 as described in Chapter 4. The results were plotted using a Weibull distribution as shown in Figure 5.25, provided the characteristic life of single bump and double bump flip chip assemblies summarized in Table 5.17. Comparing the test results to the prediction results, the strain-based predictions generally over-estimated the fatigue life, while the energy-based predictions generally under-estimated the solder joint fatigue life. The Zahn's power law approach for elastic-plastic-creep constitutive model produced the best correlation, although Shi's approach for the Anand constitutive model also gave a close correlation but that was under-estimated by about 12%.

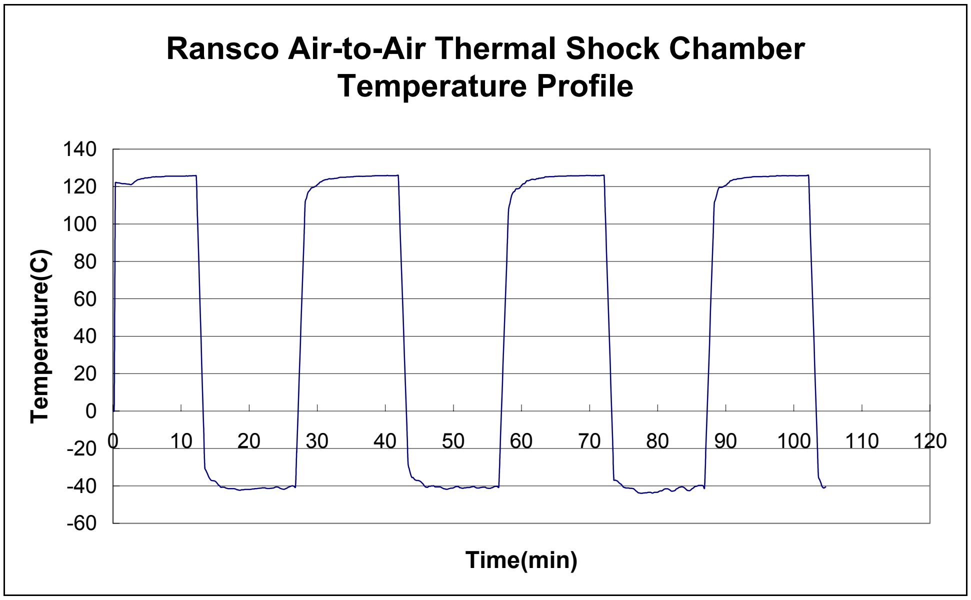


Figure 5.24 Air-to-air thermal shock chamber temperature profile (-40°C to +125°C)

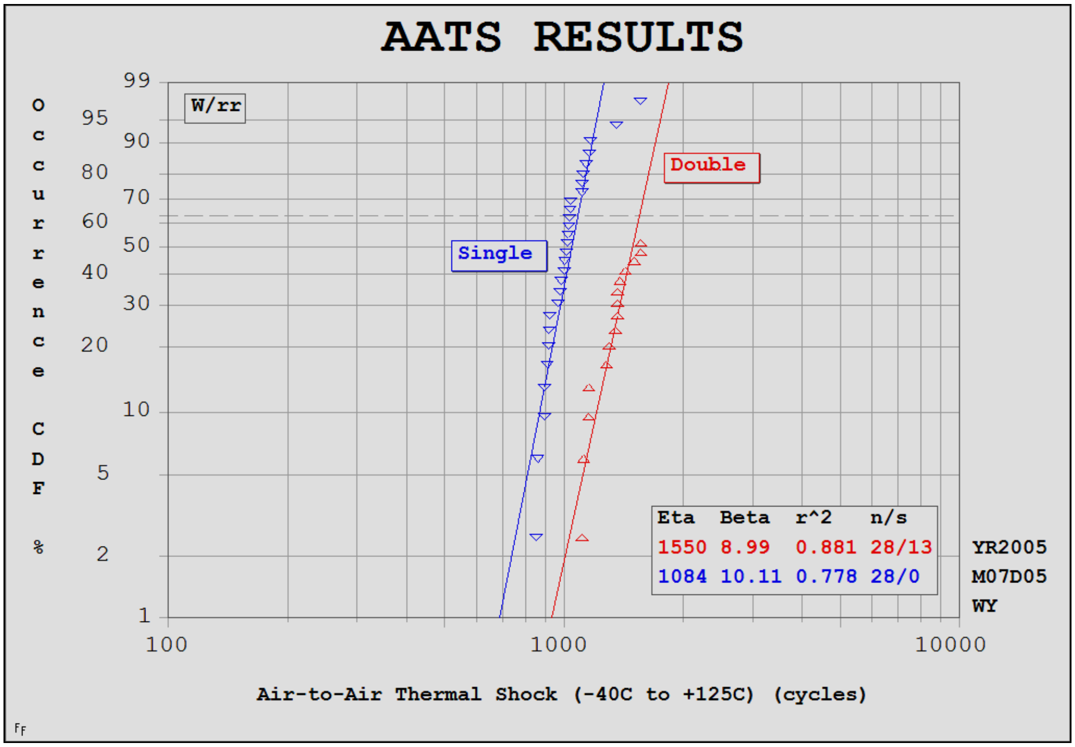


Figure 5.25 Weibull plot of the thermal shock test data for single bump flip chip and double bump flip chip assemblies.

Table 5.17 Experimental results of single and double bump flip chip from AATS test

	Single Bump	Double Bump	Improvement
Characteristic Life N_f	1084 cycles	1550 cycles	42.99%

All five fatigue life prediction models showed a reliability improvement when comparing the double bump package to single bump package, but different models suggested varying degrees of predicted solder joint life improvement. However, even the lowest improvement was 26% from single bump to double bump flip chip, obtained by Zahn's power law method with the Anand constitutive model. However, other models calculated a large fatigue life increase, in the range of 30% to 140%.

The deviation of these prediction results from the experimental data may come from several sources. First, most of these models were developed based on bulk solder material characterization or 2nd level BGA joint testing, most without underfill. When these are applied to flip chips, the volume effect and grain size effect may all introduce error into the prediction. These deviations were also reported in Yeo, et al.'s study [107] on different fatigue life model comparisons for flip chip applications. As in this study, Yeo predicted over-estimated lifetimes using Engelmaier and Solomon's strain-based models, and under-estimated lifetime by using Darveaux's and Shi's energy-based models.

In this modeling study, the main purpose was to demonstrate that double bump flip chips showed improved reliability compared to single bump flip chips and this conclusion can be directly draw from a comparison of the failure parameters before using any prediction calculations. However, when it comes to precise predictions of a fatigue life comparable to those obtained experimentally, more caution is needed when implementing the model and selecting a prediction model. Even though a close correlation with the model predictions was found in this study, further investigation needs to be undertaken to explore the sources of the deviation.

CHAPTER 6

CONCLUSIONS

6.1 Literature Review for Finite Element Modeling

The use of Surface Evolver software enables the prediction of solder joint geometries and was utilized in the double bump flip chip application. The predicted results show good agreement with experimentally measured results in terms of solder joint height and radius.

Predictive methodology for solder joint shape prediction has been used for stencil design for the bumping process of the double bump flip chip. Original solder ball "shrinkage" was identified and avoided by choosing different stencil designs. Six stencil aperture types with two aperture shape and three sizes each were investigated and square aperture shape was eliminated due to their undesirable double solder joint shape. Restoring force for the solder joint self-alignment was calculated for different offset % and was found to be increased when the offset% increased.

6.2 Double Bump Flip Chip Fabrication

A successful process has been developed for double bump flip chip fabrication. Details of the process were discussed for polishing, stencil printing, reflow, cleaning and wafer dicing. Polishing after coating the wafer with lower CTE underfill provides a flat, smooth surface and exposes the original solder ball top for second layer solder ball bumping. Other factors that affect the final double bump geometry, such as bumping site opening and stencil printing process parameters were determined. A cleaning step was found to be necessary to remove solder residues. Alteration of the dicing process for the coated wafer was made and compared to the regular bare wafer dicing.

6.3 Double Bump Flip Chip Assembly with No-flow Underfill

The assembly process with no-flow fluxing underfill was optimized for double bump flip chip application. Prebake of the PCB and the dies prior to assembly process was found to be necessary for no-flow fluxing underfill assembly. Effective prebake conditions were found to be 18hr at 125°C for PCB and 18hr at 90°C for dies. The underfill dispensing pattern also plays an important role in achieving void-free assemblies and the single dot pattern showed the best results among all the patterns tested. In order to investigate the influences of different assembly process parameters, such as underfill volume, placement force and dwell time, a DOE design was carried out for process optimization and underfill weight of 4mg, placement force of 9N and dwell time of 3 sec were found to be the best combination.

Both double bump and single bump flip chip assemblies were subjected to air-to-air thermal shock (-40 to 125°C with 15 min dwell time). The test results shows an improvement of 43% of characteristic life for the double bump flip chip assemblies compare to the single bump flip chip assemblies.

6.4 Finite Element Analysis

The present study demonstrates that finite element analysis with proper methodology applied offers a powerful tool to simulate, assess, and predict fatigue damage in solder joints. The simulated results shows a significant improved solder joint reliability. The double bump solder joint model shows less stress, less strain range and smaller accumulative inelastic work density.

Two strain-based and three energy-based predictions methodology were applied to the simulated results of both single bump and double bump flip chip models under thermal cyclic loads. A reasonably good match between the experimental data and the simulation results was found in two predictions.

6.5 Future Work

Although the current developed double bump flip chip shows improved reliability compare to the regular single bump flip chip, further optimization of the process can be achieved by replacing the eutectic solder ball with high melting temperature solder materials for the original solder ball layer in the double bump structure. Higher melting temperature of the original solder ball would assure the solder ball maintains its shape during the assembly reflow process, completely eliminating the "shrinkage" problem for any stencil design. The process window will be greatly enlarged due to this change and yield can be improved.

For the finite element analysis of the flip chip assemblies, considering the variation in results between the different prediction methodologies, further investigation should been taken to explore the source of the variation. More solder joint structure details are suggested to be included in the model for more precise simulation, such as solder mask and UBM (under bump metallurgy).

BIBLIOGRAPHY

[1] S. J. Adamson, "Review of CSP and Flip Chip Underfill Processes and When To Use the Right Dispensing Tools For Efficient Manufacturing," presented at GlobalTRONICS Technology Conference, Singapore, 2002.

[2] G. R. Blackwell, *The Electronic Packaging*. Morgan Hill, California: CRC Press, IEEE Press, 1999.

[3] Assembleon, "Flip Chip -- The Ultimate Package," 2004.

[4] T. Chen, J. Wang, and D. Lu, "Emerging challenges of underfill for flip chip application," presented at Electronic Components and Technology, 2004.

[5] J. Wang and T. Chen, "The Effect of Flow Properties on Filler Settling of Underfill in the Flip Chip Package," presented at Electronic Components and Technology Conference, 2004.

[6] B. Wun and J. Lau, "Characterization and Evaluation of the Underfill Encapsulants for Flip Chip Assembly," presented at Electronics Manufacturing Technology Symposium, La Jolla, CA, 1994.

[7] W.-B. Young and W.-L. Yang, "Underfill Viscous Flow between Parallel Plates and Solder Bumps," *IEEE Transactions on Component, Packaging, and Manufacturing Technology Part B*, vol. 25, pp. 695-700, 2002.

[8] S. Hanand and K. K. Wang, "Analysis of the Flow Encapsulant during Underfill Encapsulation of Flip Chip," *IEEE Transactions on Components, Packaging, and Manufacturing Technology Part B*, vol. 20, pp. 424-433, 1997.

[9] L. Gopalakrishnan, M. Ranjan, Y. Sha, K. Srihari, and C. Woychik, "Encapsulant materials for flip-chip attach," presented at 48th IEEE Electronic Components and Technology Conference, 1998.

[10] C.P.Wong, "Flip-Chips: Some History, a Tutorial and a Few New Perspectives," in *Chip Scale Review*, 2000.

[11] M. A. Previti, "No-Flow Underfill Reliability is Here--Finally!" presented at Technical Proceedings of APEX, Long Beach, CA, 2000.

[12] R. Pendse, Y. Choi, K. M. Kim, G. Narvaez, B. Jafari, and I. Singh, "Development of No Flow Underfill Technology for Next Generation Flip Chip Products," presented at 2005 Electronic Components and Technology Conference, 2005.

[13] C. Kim and D. F. Baldwin, "No-flow underfill process modeling and analysis for low cost, high throughput flip chip assembly," *IEEE Transactions on Electronics Packaging Manufacturing*, vol. 26, pp. 156, 2003.

[14] M. Colella and D. Baldwin, "Near Void-Free No-Flow Underfill Flip Chip On Board Assembly Technology Reliability Characterization," presented at IEEE/SEMI Int'l Electronics Manufacturing Technology Symposium, 2004.

[15] M. Colella and D. Baldwin, "Evaluation, optimization, and reliability of no-flow underfill process," pp. 216, 2002.

[16] T. L. Lazarakis and D. F. Baldwin, "Processing of no-flow fluxing underfill for flip chip assembly," presented at International Symposium on Advanced Packaging Materials. Proceedings. Stone Mountain, GA, 2002.

[17] L. Wang, H. Li, and C.P.Wong, "Reworkable No-Flow Underfills for Flip Chip Application," *IEEE Transactions on Electronics Packaging Manufacturing*, vol. 24, pp. 115, 2001.

[18] R. W. Johnson, Q. Wang, F. Ding, Z. Hou, L. Crane, G. Shi, and H. Tang, "Flip Chip Assembly with Wafer Applied Underfill," presented at IMAPS Flip Chip, Austin, TX, 2003.

[19] L. Nguyen, H. Nguyen, A. Negasi, Q. Tong, and S. H. Hong, "Wafer level underfill - processing and reliability," pp. 53, 2002.

[20] Q. Tong, B. Ma, E. ZHang, A. Savoca, L. Nguyen, C. Quentin, S. Luo, H. Li, L. Fan, and C. P. Wong, "Recent Advances on a Wafer-Level Flip Chip Packaging Process," presented at 50th Electronic Components and Technology Conference, 2000.

[21] N. Yala, J. Danvir, J. Qi, P. Kulkarni, and M. Chason, "Overview of Wafer-Applied Underfill Activities: How to Turn Flip Chip Drawbacks into Benefits," in *Chip Scale Review*, 2003.

[22] A. K. Chaudhuri, M. R. Walsh, F. Stepniak, S. Post, R. L. D. Zenner, M. A. Kropp, Scott B. Charles, R. B. Kinney, and F. B. Li, "Wafer Applied Underfill - Delivering Flip Chip to the Mainstream," presented at SAE World Congress, Detroit, Michigan, 2002.

[23] Z. Zhang, Y. Sun, L. Fan, R. Doraiswami, and C. P. Wong, "Development of Wafer Level Underfill Material and Process," presented at Electronics Packaing Technology COnference, 2003.

[24] L. Wang and C. P. Wong, "Recent advances in underfill technology for flip-chip, ball grid array, and chip scale package applications," presented at International Symposium on Electronic Materials and Packaging (EMAP 2000). Hong Kong, 2000.

[25] K. Gilleo and D. Blumel, "New Generation Underfills Power the 2nd Flip Chip Revolution," presented at Pan Pacific Microelectronics Symposium, 1998.

[26] IBM, "http://www-03.ibm.com/chips/technology/makechip/interconnect/4.html."

[27] K. Gilleo, "A Brief History of Flipped Chips." www.flipchips.com

[28] J. M. Hurley, T. Berfield, S. Ye, R. W. Johnson, R. Zhao, and G. Tian, "Kinetic Modeling of No-Flow Underfill Cure and Its Relationship to Solder Wetting and Voiding," presented at Electronic Components and Technology Conference, 2002.

[29] R. Zhao, R. W. Johnson, G. Jones, and E. Yaeger, "Processing of FLuxing Underfills for Flip Chip on Laminate Assembly," presented at Proceeding of 2002 APEX Conference, San Diego,CA, 2002.

[30] Z. Zhang and C. P. Wong, "Double-Layer No-Flow Underfill Materials and Process," *IEEE Transactions on Advanced Packaging*, vol. 26, pp. 199-205, 2003.

[31] S. H. Shi, T. Yamashita, and C. P.Wong, "Development of the wafer-level Compressive-Flow Underfill Encapsulant," presented at 49th Electronic Components and Technology Conderence, 1999.

[32] R. L. D. Zenner, B. S. Carpenter, A. Chaudhuri, F. Stepniak, S. Post, and J. Johnson, "Wafer-applied underfill film laminating," presented at 8th International Symposium on Advanced Packaging Materials, 2002.

[33] Y. Guo, W. Chen, and C. Lim, "Experimental Determination of Tehrmal Strains in Semiconductor Packaging Using Moire Interferometry," presented at ASME/JSME Joint Conference on Electronic Packaging, San Jose, California.

[34] J. Wang, "Flow Time Measurements for Underfill in Flip Chip Packaging," *IEEE Transactions on Components and Packaging Technologies*, vol. 28, 2005.

[35] C.-Y. Huang, K. Srihari, and P. Borgesen, "Optimization of the Substrate Preheat at Temperature for the Encapsulation of Flip Chip Devices," 2005.

[36] P. Kulkarni, R. W. Johnson, E. Yaeger, M. Konarski, A. Torres, P. Krug, and L. Crane, "Manufacturing of Flip Chip on Laminate," presented at Proceedings of the APEX Technical Program, Long Beach, CA, 2000.

[37] Y. Chia, H. Yam, S. H. Lim, K. S. Chian, S. Yi, and W. T. Chen, "An Optimization study of Underfill Dispensing Volume," *IEEE Transactions on Electronics Packaging Manufacturing, [see also Components, Packaging and Manufacturing Technology, Part C: Manufacturing, IEEE Transactions on]*, vol. 26, 2003.

[38] W.-B. Young, "The effect of solder bump pitch on the underfill flow," *IEEE Transactions on Advanced Packaging*, vol. 25, pp. 537-542, 2002.

[39] D. Davidson, G. L. Lehmann, and E. J. Cotts, "Effect of geometric surface features on void formation: application to the underfill process," presented at 52nd Electronic Components and Technology Conference, 2002.

[40] K. Slesinger, J. David, A. Mays, C. Fieselman, and M. Elkins, "Flip Chip Underfill: Void Detection and Correction," presented at SMTA International, 2002

[41] Z. Zhang and C. P. Wong, "Recent Advances in Flip-Chip Underfill: Materials Process, and Reliability," *IEEE Transactions on Advanced Packaging*, vol. 27, pp. 515-524, 2004.

[42] K. Banerji, "Vacuum infiltration of underfill material for flip chip devices," US Patent 5203076, 1993.

[43] K. K. Wang and S. Han, "Pressurized underfill encapsulation of integrated circuits," US Patent 6000924, 1997.

[44] P. Kondos, D. B. Peter Borgesen, and A. Prats, "Optimizing Flip Chip Substrate Layout for Assembly." www.uic.com, 2004

[45] R. Rendse, Y. N. Choi, K. M. Kim, G. Narvaez, B. Jafari, and I. Singh, "Development of No Flow Underfill Technology for Next Generation Flip Chip Products," presented at Electronic Components and Technology Conference, 2005.

[46] Y. Zhang, R. Zhao, D. K. Harris, and R. W. Johnson, "A Computational Study on Solder Bump Geometry, Normal, Restoring and Fillet Forces During Solder Reflow in the Presence of Liquefied Underfill," *IEEE Transactions on Electronics Packaging Manufacturing*, vol. 25, pp. 308, 2002.

[47] R. V. Burress, M. A. Capote, Y.-J. Lee, H. A. Lenos, and J. F. Zamora, "A Practical, flip-chip multi-layer preencapsulation technology for wafer-scale

underfill," presented at 51th Electronic Components Technology Conference, Orlando, FL, 2001.

[48] C.-M. Liu, "Solder Shape Design and Thermal Stress and strain Analysis of Flip Chip Packaging Using Hybrid Method," presented at Int'l Symp on Electronic Materials 44 & Packaging, 2000.

[49] R. Satoh, K. Arakawa, M. Harada, and K. Matsui, "Thermal Fatigue Life of Pb-Sn Alloy Interconnections," *IEEE Transactions on Components and Manufacturing Technologies*, vol. 14, 1991.

[50] S. F. Popelar, "A Parametric Study of Flip Chip Reliability Based on Solder Fatigue Modeling: Part II - Flip Chip on Organic," presented at Proceedings of the 1998 International Symposium on Microelectronics, San Diego, California, 1998.

[51] K.-N. Chiang and Y. T. Lin, "On Enhancing Eutectic Solder Joint Reliability Using A Second-Reflow-Process Approach," *IEEE Transactions on Advanced Packaging*, vol. 23, pp. 9-14, 2001.

[52] X. Liu and G.-Q. Lu, "Effects of Solder Joint Shape and Height on Thermal Fatigue Lifetime," *IEEE Transactions on Components and Packaging Technologies*, vol. 26, pp. 455, 2003.

[53] E. Madenci, S. Shkarayev, and R. Mahajan, "Potential Failure Sites in a Flip Chip Package With and Without Underfill," *ASME, Journal of Electronic Packaging*, vol. 120, pp. 336-341, 1998.

[54] S. Rzepka, M. A. Korhonen, E. Meusel, and C.-Y. Li, "The Effect of Underfill and Underfill Delamination on th Thermal Stress in Flip-Chip Solder Joints," *ASME, Journal of Electronic Packaging*, vol. 120, pp. 342-348, 1998.

[55] C. K. Gurumurthy, J. Jiao, L. G. Norris, C. Y. Hui, and E. J. Kramer, "A Thermal-Mechanical Approach for Fatigue Testing of Polymer Bimaterial Interfaces," *ASME Journal of Electronic Packaging*, vol. 120, pp. 328-335, 1998.

[56] S. Y. Kook, J. M. Snodgrass, A. Kirtikar, and R. H. Dauskardt, "Adhesion and Reliability of Polymer/Inorganic Interfaces," *ASME Journal of Electronic Packaging*, vol. 120, pp. 328-335, 1998.

[57] J. Wang, M. Lu, D. Zou, and S. Liu, "Investigation of Interfacial Fracture Behavior of a Flip-chip Package under a Constant Concentrated Load," *IEEE Transactions on Components ,Packaging, and Manufacturing Technology*, vol. 21, pp. 79-86, 1998.

[58] C. P. Wong and R. McBride, "Preencapsulation Cleaning Methosd and Control for Microelectronics Packaging," *IEEE Transactions on Components, Packaging, and Manufacturing Technology*, vol. 17, pp. 542-552, 1994.

[59] J. Lau, S.-W. R. Lee, and C. Chang, "Effects of Underfill Material Properties on the Reliability of Solder Bumped Flip Chip on Board with Imperfect Underfill Encapsulants," *IEEE Transactions on Components, Packaging, and Manufacturing Technology Part B*, vol. 23, pp. 323-332, 2000.

[60] Z. Cheng, "Lifetime of Solder Joint and Delamination in Flip Chip Assemblies," presented at International Conference on the Business of Electronic Product Reliability and Liability, 2004.

[61] N. Matsui, S. Sasaki, and T. Ohsaki, "VLSI Chip Interconnection Technology Using Stacked Solder Bumps," *IEEE Transactions on components, Hybrids, and manufacturing Technology*, vol. CHMT-12, pp. 566, 1987.

[62] F. Lovasco and M. A. Oien, "Double-bump surface-mount technology for very high IO interconnections," presented at 4th IEEE/CHMT European International for Electronic Manufacturing Technology, 1988.

[63] J. Jasper and J. Simon, "Wafer Level CSP with Solder Support Structure."

[64] J. Simon and H. Riechl, "Board Level Reliability of a Waferlevel CSP using Stacked Solder Spheres and a Solder Support Structure (S3)," presented at 50th Electronic Components and Technology Conference, 2000.

[65] A. Grieve, M. A. Capote, H. A. Lenos, and A. Soriano, "Low-Cost, Wafer Level Underfilling and Reliability Testing of Flip Chip Devices," presented at 54th Electronic Components and Technology Conference, 2004.

[66] K. Tatsumi, Y. Yamamoto, K. Iwata, E. Hashino, S. Ishikawa, and T. Kohno, "An Application of Micro-ball Wafer Bumping to Double Ball Bump for Flip Chip Interconnection," presented at Electronic Components and Technology Conference, 2005.

[67] M. Topper, J. Auersper, V. Glaw, K. Kaskoun, E. Prack, B. Keser, P. Coskina, D. Jager, D. Petter, and O. Ehrmann, "Fab Integrated Packaging (FIP): A New Concept for High Reliability Wafer-Level Chip Size Packaging," presented at Electronic Components and Technology Conference, 2000.

[68] B. Keser, B. Yeung, J. White, and T. Fang, "Encapsulated Double-Bump WL-CSP: Design and Reliability," presented at Electronic Components and Technology Conference, 2001.

[69] V. Patwardhan, H. Nguyen, L. Zhang, N. Kelkar, and L.Nguyen, "Constrained Collapse Solder Joint Formation for Wafer-Level-Chip-Scale Packages to Achieve Reliability Improvement," presented at 54th Electronic Components and Technology Conference, 2004.

[70] Q. Zhu, G. Wang, and L. Luo, "Optimization of Design and Manufacturing Parameters for Solder Joint Geometry and Self-Alignment in Flip-Chip Technology," presented at 5th International Conference on Solid-State and Integrated Circuir Technology, 1998.

[71] B. H. Yeung and T.-Y. T. Lee, "Evaluation and Optimization of Package Processing, Design, and Reliability through Solder Joint Profile Prediction," presented at 2001 Electronic Components and Technology Conference, 2001.

[72] Sidharth, R. Blish, and D. Natekar, "Solder joint shape and standoff height prediction and integration with FEA-based methodology for reliability evaluation," presented at 52nd Electronic Components and Technology Conference, 2002.

[73] K.-N. Chiang and C.-A. Yuan, "An Overview of Solder Bump Shape Prediction Algorithms with Validations," *IEEE Transactions on Advanced Packaging*, vol. 24, 2001.

[74] S. M. Heinrich, P. E. Liedtke, N. J. Nigro, A. F. Elkouh, and P. S.Lee, "Effect of chip and pad geometry on solder joint formation in SMT," *Journal of Electronic Packaging*, vol. 115, pp. 443-439, 1993.

[75] S. M. Heinrich, M. Schaefer, S. A. Schroeder, and P. S. Lee, "Prediction of solder joint geometries in array-type interconnects," *ASME Journal of Electronic Packaging*, vol. 118, pp. 114-121, 1996.

[76] I. R. Holub, J. M. Pitarresi, and T. J. Singler, "Effect of solder joint geometry on the predicted fatigue life of BGA solder joints," presented at the InterSociety Conference on Thermal Phenomena, pp. 187-194, 1996.

[77] M. J. Pfeifer, "Solder Bump Size and Shape Modeling and Experimental Validation," *IEEE Transactions on Components, Packaging, and Manufacturing Technology -- Part B*, vol. 20, pp. 452, 1997.

[78] K. Brakke, "The Surface Evolver Manual," 2.23 ed, 1992.

[79] P. M. Martino, G. M. Freedman, and L. M. Racz, "Predicting solder joint shape by computer modeling," presented at 44th Electronic Components and Technology Conference, 1994.

[80] D. C. Burr, "Printing Guidelines for BGA and CSP Assemblies," presented at Surface Mount International Conference and Exposition, 1999.

[81] Cookson Electronics Alpha Stencils, "Stencil Design Guidelines." http://www.alphametals.com/products/stencils/design_guide.html

[82] "Solder Bump Flip Chip," http://www.flipchips.com/tutorial02a.html

[83] P. Biocca, "Flux Chemistries and Thermal Profiling: Avoiding Soldering Defects in SMT Assembly," http://www.loctite.com/

[84] A. A. Primavera, "Influence of PCB Parameters on Chip Scale Package Assembly and Reliability," Universal Instruments Corporate technical paper

[85] John H. Lau, *Flip Chip Technologies*. New York: McGraw Hill, 1996.

[86] J. H. L. Pang and T. H. Low, "Modeling Thermal Cycling and Thermal Shock Tests for FCOB," International Society Conference on Thermal Phenomana, vol., pp. 987-992, 2002.

[87] G. Gustafsson, "Finite Element Modeling of BGA Packages for Life Prediction," presented at Electronic Components and Technology Conference, 2000.

[88] J. H. L. Pang, D. Y. R. Chong, and T. H. Low, "Thermal Cycling Analysis of Flip-Chip Solder Joint Reliability," IEEE Transactions on Components and Packaging Technologies, vol. 24, pp. 705-713, 2001.

[89] B. Rodgers, J. Punch, and J. Jarvis, "Finite Element Modeling of a BGA Package Subjected to a Thermal and Power Cycling," presented at ITherm Conference on Thermal Mechanics and Thermomechanical Phenomena in Electronic Systems, 2002.

[90] X. J. Zhao, G. Q. Zhang, and L. J. Ernst, "Solders Fatigue Prediction Using Interfacial Boundary Volume Criterion," Transactions of the ASME, vol. 125, pp. 582-589, 2003.

[91] S. Wiese, S. Jakschik, F. Feustel, and E. Meusel, "Fracture Behaviour of Flip Chip Solder Joints," presented at Electronic Components and Technology Conference, 2001.

[92] R. Darveaux, "Effect of Simulation Methodology on Solder Joint Crack Growth Correlation," presented at Proceedings of 50th Electronic Components and Technology Conference, 2000.

[93] "ANSYS 8.0 Manual," 2004.

[94] S. Wiese, S. Rzepka, and M. Ekkehard, "Time-Independent Plastic Behavior of Solders and its Effect on FEM Simulations for Electronic Packages," presented at 8th International Symposium and Exhibition on Advanced Packaging Materials, 2002.

[95] R. Darveaux, K. Banerji, and A. Mawer, "Reliability of Plastic Ball Grid Array Assembly," in *Ball Grid Array Technology*, J. Lau, Ed.: McGraw Hill, 1995, pp. 379-442.

[96] F. Feustel, S. Wiese, and E. Meusel, "Time-dependent material modeling for finite element analyses of flip chips," presented at Electronic Components and Technology Conference, 2000.

[97] S. S. Manson, *Thermal stress and Low-Cycle Fatigue*. New York: McGraw Hill, 1966.

[98] H. D. Solomon, "Fatigue of 60/40 Solder," IEEE Transactions on Components, Packaging and manufacturing Technology, vol. CHMT-9, pp. 423-433, 1986.

[99] W. Engelmaier, "Surface mount solder joint reliability controlling the thermal expansion," presented at Electronic Manufacturing Technology Symposium of Fourth IEEE/CHMT European International, 1988.

[100] W. Engelmaier, "Solder Joints in Electronics Design for Reliability," in *Keynote Address in Design and Reliability of Solder and Solder Interconnectons*, 1997, pp. 9-19.

[101] V. Gektin, A. Bar-Cohen, and J. Ames, "Coffin-Manson Fatigue Model of Underfilled Flip-Chips," IEEE Transactions on Components, Packaging and Manufacturing Technology, vol. 20, pp. 317-325, 1997.

[102] X. Q. Shi, H. L. J. Pang, W. Zhou, and Z. P. Wang, "A Modified Energy-Based Low Cylce Fatigue Model for Eutectic Solder Alloy," Scripta Materialia, vol. 41, pp. 289-296, 1999.

[103] J.-P. Clech, "Flip- Chip/ CSP Assembly Reliability and Solder Volume Effects," presented at Proceedings, Surface Mount International Conference, San Jose, CA, 1998.

[104] J.-P. Clech, "Solder Joint Reliability Issues," presented at CMAP Reliability Workshop, Toronto, Canada, 2000.

[105] B. A. Zahn, "Solder Joint Fatigue Life Model Methodology for 63Sn 337Pb and 95.5Sn4Ag0.5Cu Materials," presented at 2003 Electronic Components and Tehcnology Conference, 2003.

[106] J. Liu, R. Kraszewski, X. Lin, L. Wong, S. Goh, and J. Allen, "New Developments in Single Pass Reflow Encapsulant for Flip Chip Application," presented at Electronic Components and Technology Conference, 2001.

[107] A. Yeo, C. Lee, and J. H. L. Pang, "Flip Chip Solder Joint Fatigue Life Model Investigation," presented at Electronic Packaging Technology Conference, 2002.