MODELING, DESIGN AND DEMONSTRATION OF LARGE 2.5D GLASS BGA PACKAGES FOR BALANCED CHIP- AND BOARD-LEVEL RELIABILITY

A Thesis
Presented to
The Academic Faculty

by

Vidya Jayaram Mythly

In Partial Fulfillment
of the Requirements for the Degree
Master of Science in the
School of Material Science and Engineering

Georgia Institute of Technology

May 2017

MODELING, DESIGN AND DEMONSTRATION OF LARGE 2.5D GLASS BGA PACKAGES FOR BALANCED CHIP- AND BOARD-LEVEL RELIABILITY

Approved by:

Prof. Rao Tummala, Advisor
School of Material Science and Engineering
Georgia Institute of Technology

Prof. Suresh Sitaraman
School of Mechanical Engineering
Georgia Institute of Technology

Dr. Vanessa Smet
School of Electrical and Computer Engineering
Georgia Institute of Technology

Date Approved: April 20, 2017

To my Amma and Appa

ACKNOWLEDGEMENTS

By submitting this thesis, my long and delightful journey of being a student is officially over. This journey would have not been possible without having the guidance, friendship, help, and support of many wonderful people to whom I will always be grateful.

I would like to express my sincere gratitude to my advisor, Prof. Rao Tummala, for giving me the opportunity to join his team and work under his supervision. His passion and enthusiasm for research, his distinct personality and leadership skills have constantly inspired me throughout my studies as a Master's student. I have always felt I am a fortunate graduate student to have him as my advisor, not only because he is a great professor who provides enormous number of research opportunities for his students, but also because he is an extraordinary human being and a great teacher.

Next, I would like to thank my mentor, Dr. Vanessa Smet. Her encouragements, patience, and continuous support of my work have made my 2 years of working with her an exceptional experience for me and I am deeply grateful. She has stood by me through both my professional and personal difficulties, and I always say with confidence that I can go to war with Vanessa by my side.I could not have asked for a better mentor and a friend.

I am also very grateful to Dr. Raj Pulugurtha and Dr. Venky Sundaram for sharing with me their knowledge and guiding me throughout the course of my research. I would like to thank Prof. Suresh Sitaraman for being on my thesis defense committee and for providing insightful comments about my work.

Being part of an active research group of motivated and knowledgeable members is significantly important to the success of every graduate student. I feel very fortunate to have been part of such a team at the 3D Systems Packaging Research Center with hard-working students who selflessly helped me out with different aspects of my research. I would specially like to mention Dr. Scott McCann for his expertise and relentless help on ANSYS modeling - without which my progress on that front would be impossible. His

training sessions on tools for warpage and reliability analysis helped me a long way. Many thanks to my colleagues from the Interconnections & Assembly team: Bhupender Singh for mentoring me, Nathan for his guidance on assembly with Kulicke and Soffa bonder and underfill processes, Siddharth Ravichandran for assisting me with fabrication and assembly processes, Ninad Shahane and Kashyap Mohan for their constant encouragement and moral support. I would also like to thank Zihan Wu, Jialing Tong, Bruce Chao, Karan Bhangaonkar, Shreya Dwarakanath, Chandrasekharan Nair, Omkar Gupte and all other fellow GRAs. My past two years at PRC have been filled with a lot of good memories because of these wonderful people. I will definitely miss our group meetings, conference trips, coffee breaks and I will always be grateful for their friendship. I am grateful to the visiting engineers - Satomi Kawamoto for help with underfill processes, Hiroyuki Matsuura and Yuya Suzuki for their help on fabrication. I would like to specially thank Chris White and Jason Bishop for their assistance with lab equipment. I would also like to thank Karen May, Brian McGlade, Patricia Allen and Kimberly Purvis for their great support of my administrative needs.

I am indebted to my parents, Mythly and Jayaram, for their unconditional love and support throughout my life. My father has always been an example to follow my mother has been a continuous source of love and moral support. They taught me to never give up and have continuously provided me not only great hopes and strong motivations but everything I need in the world to build myself a better future. Many thanks to my wonderful grandparents for their love over the years. This thesis would have not been possible without the support of my friends and family.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

Transistor scaling, driven by Moores Law, has enabled the integration of billions of transistors on a single integrated chip (IC); thereby enabling rapid miniaturization of microprocessor devices such as smartphones, servers and personal computers. However, silicon integration following Moore's law is now reaching its limits due to increasing design complexity and cost, bringing the need for a new "System Scaling" approachfor further miniaturization and performance improvements. The System-on-Package (SOP) approach, pioneered by Georgia Tech PRC, relies on co-integration of multiple electronic functions on a package substrate, as opposed to on-chip. Packaging, therefore, becomes key in enabling higher functional densities. An example of this new approach to system design is the recent trend of "split dies" where large devices is broken down into smaller devices at finer I/O pitches that are interconnected on a substrate using high-density wiring. These advanced package architectures such as 2.5D interposer packages now rely on packaging to improve performance and miniaturize the system as a whole.

Silicon interposers are particularly attractive in such split-die applications due to their outstanding lithographic capability enabling high-density, high-speed die-to-die interconnections. Such 2.5D interposers tend to be fairly large with body sizes exceeding 30 mm x 40 mm, bringing unprecedented board-level reliability challenges due to large mismatch in coefficients of thermal expansion (CTE) between silicon and mother boards. These challenges are typically addressed by introducing an additional organic BGA package between interposer and board to accommodate for the CTE mismatch and decrease in pitch. However, this degrades electrical performance with longer interconnection lengths, and adds to the overall cost.

Glass has emerged as an alternative substrate technology to overcome the shortcomings of silicon. Glass has been demonstrated to have superior electrical properties than silicon with lower losses and can accommodate high-density wiring owing to micron-scale

lithographic design rules. Further, glass can be tailored for a wide CTE range of 3.3 to 9.8 ppm/K. This unique property brings design flexibility to address board-level reliability challenges and directly assemble large glass interposer packages to boards without the need for an intermediate organic package.

The primary objective of this research is to model, design and demonstrate a large, 2.5D glass BGA package with 1) direct SMT-to-board interconnection; and 2) balanced chip- and- board-level reliability. The ultimate goal is to provide guidelines for the design of 2.5D glass BGA packages, optimizing the glass CTE to mitigate warpage and achieve system-level reliability, and subsequently the assembly process and sequence.

Finite-element models were built to assess the reliability of 2.5D glass packages with direct SMT assembly to the board. The methodology for achieving balanced chip- and-board-level reliability was validated through focused modeling and experimental results for a single-chip package. Board-level reliability was recognized as the most critical challenge and enhanced byusing innovative doped solder materials such as Indium's Mn-doped SACm™ alloy and strain-relief mechanisms to give more design flexibility. Failure distribution analysis and optical characterization was performed to evaluate thermal cycling reliability. A process design approach was demonstrated for mitigating warpage induced by thermocompression bonding on ultra-thin, low- and- high-CTE substrates at I/O pitches below 50 μm. By selecting optimum thermal profiles for mitigating chip-level assembly warpage, board-level assembly is enabled at larger package sizes, and system-level reliability is thereby enhanced.

CHAPTER 1
INTRODUCTION

Modern consumer electronics industry is being driven by size, cost and performance. With Moore's Law reaching its physical limits, the need for advanced microelectronics packaging is being recognized. The recent split die trend is an example of this new System Scaling approach to system design, in which a large die is divided in multiple smaller dies at finer pitch. Functionality is then reconstituted through high-density fine-pitch wiring on the substrate. Silicon interposers have gained strong momentum in such 2.5D package architecture as they, unlike organic substrates, can support the required interconnect density. However, such interposers tend to be thin and large, averaging 30 mm 40 mm in body size, which brings unprecedented yield and reliability challenges at board level due to the large mismatch in coefficient of thermal expansion (CTE) between silicon and motherboard. To address these challenges, an additional organic BGA package is typically introduced between interposer and board, which degrades electrical performance and adds to the system's cost. A two-level hierarchy with direct surface-mount (SMT) assembly of the interposer on the motherboard is, therefore, desirable to meet the performance and miniaturization needs of future high-performance systems. Georgia Tech has recently pioneered glass as an alternative substrate technology overcoming the shortcomings of silicon. Glass has superior electrical properties with low losses, is capable of 5 μm lithographic design rules giving high-density multi-layered wiring, and has unique mechanical properties such as high modulus and tailorable CTE in the 3.2-9.8 ppm/K range, giving design flexibility to balance chip- and board-level reliability. Thus, 2.5D glass interposer packages where the interposer also acts as a package can comprehensively address the aforementioned challenges.

1.1 Evolution in electronics packaging from transistor scaling to system scaling

Transistor scaling, driven by Moore's Law, has enabled the integration of billions of transistors on a single integrated chip (IC); thereby reducing cost and enhancing the performance in addition to rapid miniaturization of microprocessor devices such as smartphones, servers and personal computers. It gave rise to on-chip integration also known as system-on-chip (SOC). This technology aims at combining multiple heterogeneous functions such as processor, memory, wireless and graphics by integrating the required components on a single chip. Although SOC aims at providing highest performance at a compact system level, it faces cost, design, fabrication and integration challenges. These challenges act as drivers for considering other ways to integrate at system level such as system-in-package (SIP) and multi-chip modules (MCM). To achieve the expected performance with the SOC approach, larger multi-functional dies are required. However, such large and complex dies face major yield limitations at wafer scale. MCMs were sought as an approach to functionally integrate smaller dies interconnected horizontally in addition to providing design flexibility for each functional component. They evolved from HTCCs that included multilayer ceramics interconnected with multiple wiring layers to LTCCs with dielectric layer build ups and sputtered or electroplated copper conductors with better electrical conductivity. In parallel to MCMs, the SIP approach was also pursued. SIPs are defined as three dimensional or vertical stacking of ICs that may or may not be similar in functionality. This approach leads to miniaturization if the size and thickness of stacked components are reduced. In the process of integrating all system components and power sources on the system board, this technology faces processing challenges at nanoscale, and integration thus becomes difficult. In addition, SIP only aims at silicon integration that accounts for barely 10% of the overall system. However, due to their benefits, a wide range of SIP-based modules has been implemented in high-volume production, with the following classification: (a) SIP by wire bonding, (b) SIP by flip chip and wire bonding, (c) SIP by flip chip-on-chip, (d) 3D

integration by through-silicon-via technology and (e) SIP by package stacking. Although SOC, MCM and SIP seek to ultimately increase functionality while following Moore's Law, they fail to achieve miniaturization at system level. The performance of ICs depends on transistor density which is now being limited by constraints on Moore's Law. The gap between transistor scaling (based on the gate length of a transistor) and system scaling (based on off-chip interconnection pitch) illustrated in Figure 1.1 indicates the need for a new paradigm for system integration and miniaturization.

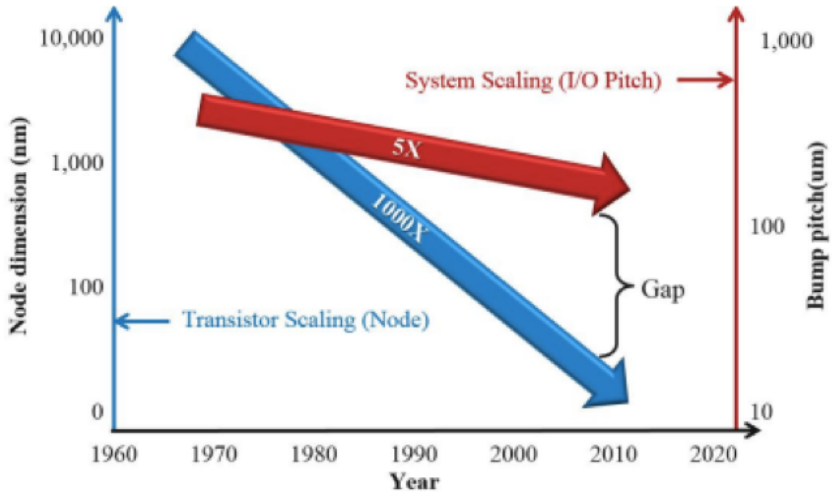


Figure 1.1: Gap between transistor and system scaling (Courtesy Dr. S. Iyer, IBM).

Georgia Tech's 3D Systems Packaging Research Center (3D-PRC) has been pioneering System-on-Package (SOP) to address the fundamental challenges of existing technologies [1]. System on package (SOP) is a novel concept based on miniaturization of the device, package and system board into a single package that includes all system functions. The two main criteria governing SOP are size reduction and increased functionality. SOP reduces the size of 80-90% of the non-IC part of the system through ultra-high wiring densities with less that 5 μm lines and spaces, use of RDLs and embedded ultrathin film components. Additionally, it also reduces cost and provides technical advantages in digital, wireless and optoelectronic-based systems. SOP provides a balance between IC and package integration that overcomes the shortcomings of the technologies stated above. Advanced packaging so-

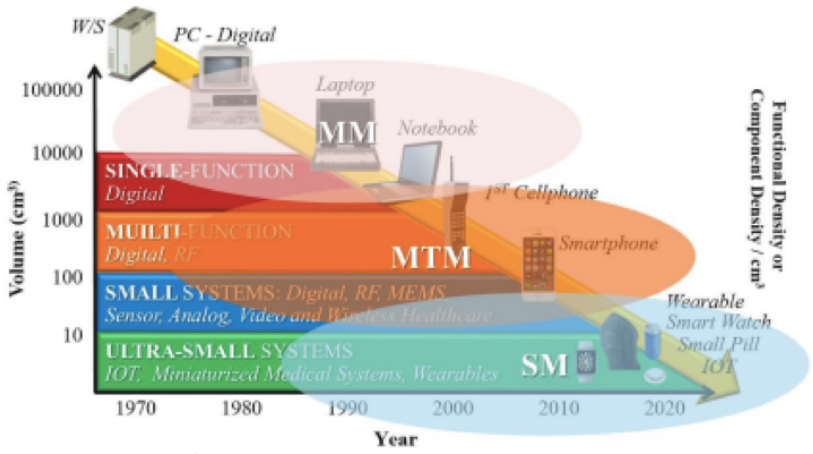


Figure 1.2: Increasing functional densities for next-generation ultra-small systems: system scaling by System Moore (SM) for heterogeneous integration beyond More of Moore (MM) and More than Moore (MTM) [1].

lutions, realizing the SOP concept, are therefore necessary to achieve ultra-miniaturization and increased functionality, with tomorrows smartphones having the same functions as that of todays supercomputers, tomorrows smart-watches with the same performance as that of todays smartphones and so on (Figure 1.2).

1.2 Emergence of 2.5D technology

Smart systems of today require high logic-to-memory bandwidth with high signal switching speeds. The system bandwidth for high-performance applications is expected to double every year [2] and estimated to increase to 512 GB/s to 1 TB/s in the near future. Such systems continue to follow the trend of ultra-miniaturization. This introduces additional demands to achieve higher bandwidths at reduced form factor, besides the need to reduce cost and power consumption. To meet the demands, new technologies were introduced with ultra-fine pitch interconnections between logic and memory devices through 3D stacking with vertical through-silicon-vias (TSV) interconnections or horizontal 2.5D interconnections. The 3D vertical stacking of logic and memory chips, interconnected by TSVs is shown in Figure 1.3 [3]. This approach stacks memory chips on top of the logic die at the bottom and requires TSVs in the logic dies. This creates several challenges such as

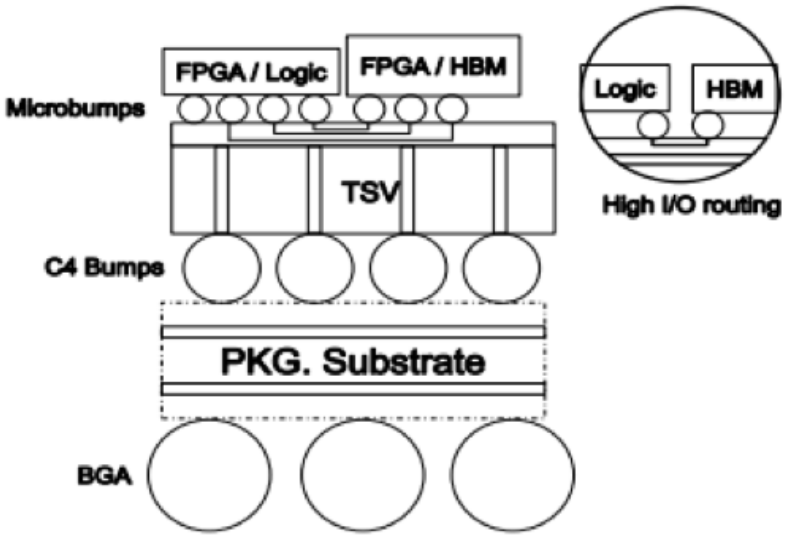


Figure 1.3: 2.5D approach with lateral I/O connections using passive Si interposers.

complex co-design between logic and memory dies, large thermomechanical stresses on active devices from TSVs, high electrical losses and yield loss due to lack of testability of known-good-dies (KGD).

To address these challenges, a new approach of high-density horizontal interconnections was introduced between logic and memory dies. It follows a similar concept as that of multichip modules (MCMs), but with higher I/O densities, known as 2.5D technology. To maintain high functional densities, chip-level interconnections of logic and memory dies are moving towards finer pitches of less than 35 μm. This, in turn, imposes finer die-to-die interconnect pitches below 10 μm. Silicon-based 2.5D interposers were subsequently considered due to their high-density sub-5 μm wiring capability. AMDs Fury and Xilinxs FPGAs are some of the commercial implementations of the silicon interposer technology. However, the dielectric losses associated with silicon is a limiting factor. Also, an additional organic layer was introduced in 2.5D packages, to address the challenge of board-level thermomechanical reliability. Nevertheless, this resulted in additional concerns of increased parasitics, larger thicknesses and higher costs. An ideal solution would be to have a package that can be directly mounted on to the board with a 2-level hierarchy, with

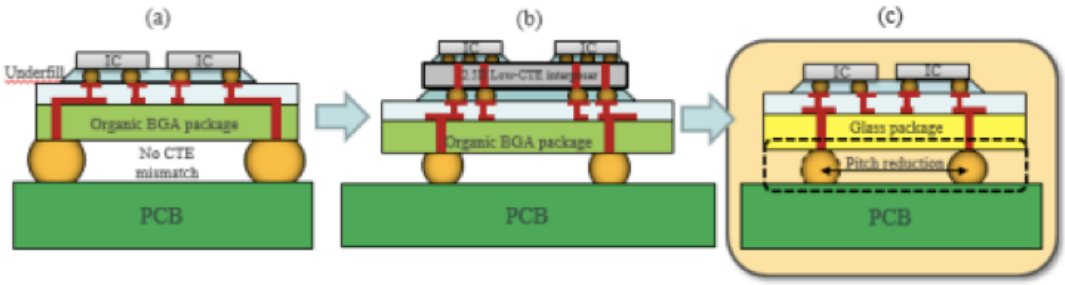


Figure 1.4: Traditional package approach (a) IC organic BGA PCB, current package (b) IC low-CTE interposer organic BGA PCB, and new package approach (c) IC large and thin glass package PCB.

lowered costs and increased bandwidths (Figure 1.4).

Glass substrates have emerged as a promising alternative to silicon interposers for 2.5D applications owing to their superior electrical properties including low loss tangent, low dielectric constant. Glass also exhibits superior thermal and dimensional stability and low surface roughness giving micron-scale lithographic capability. Glass also exhibits superior thermal and dimensional stability and low surface roughness. Additionally, glass can be processed at a large panel scale, resulting in higher throughput and lower cost. Further, glass has tailorable CTE (3.8-9.8 ppm/K) and high modulus, hence can mitigate warpage introduced by thickness reduction. The tailorable CTE also provides design flexibility to optimize chip- and board-level reliabilities. With glass, the desired 2-level hierarchy can therefore become possible, bridging the technology gap between the high-performance needs of the market at low form factor and the challenges associated with achieving them.

1.3 Research objectives

The primary objective of this work is to model, design and demonstrate a large, 2.5D glass BGA package with 1) direct SMT-to-board; 2) minimum warpage; and 3) balanced chip-and- board-level reliability. This work also aims at providing guidelines for the design of chip-level assembly processes for minimum warpage with considerations of CTE and package thickness. The metrics required to achieve the desired goals along with related

Table 1.1: Research objectives beyond prior art, technical challenges and associated research tasks

Parameters	Objectives	Prior Art	Technical Challenges	Research Tasks
Interposer size	Glass package: ~30mm x 40mm	Si interposer: 30mm x 38mm	1. Balanced chip- and-board-level reliability 2. Aggravated plastic strains in BGAs 3. Warpage mitigation in assembly	1. Modeling and design of glass package for chip- and-board-level reliability 2. Design and demonstration of board-level interconnections for strain relief and reliability 3. Modeling and design of assembly processes for controlled warpage
Interposer technology	Low- & high-CTE glass: 3.3-9.8 ppm/°C	Si: 2.5 ppm/°C		
Chip-level assembly	Optimized TCB for min warpage	3D ICs with TSV		
Board-level assembly	Direct SMT of glass package to PCB	3-level hierarchy with Si interposer, organic BGA and PCB		

prior art, main technical challenges and their associated research tasks are summarized in Table 1.1.

1.4 Technical challenges

The gap between transistor scaling and system scaling led to System Moore as the frontier for system scaling, beyond Moore's law. Emerging high-performance computing systems have been aggressively driving advances in packaging technologies to meet their escalating performance and miniaturization needs. Three main challenges have been identified in realizing a 2.5D glass BGA package with direct SMT assembly to the board:

1. Balanced chip- and board-level reliability:

   Due to the large CTE mismatch between Si and organic boards, chip- and board-level reliability are driving conflicting requirements on the mechanical properties of glass,

in particular its CTE. While lower CTEs of glass are preferred to achieve reliable chip-level interconnections at pitches below 35m, higher CTEs are critical for board-level reliability. The glass CTE, therefore, needs to be optimized to mitigate chip- and board-level reliability.

2. Aggravated plastic strains in BGAs:

   The large package size of 30 mm x 40 mm and reduced thickness of the glass substrate in the 100 - 300 μm range result in severe plastic strains in BGA interconnections, with board-level reliability identified as the most critical challenge in this package configuration. For the past 2 decades, solder alloys have been used as the primary interconnection technology for SMT assembly of packages onto printed circuit boards (PCBs). These interconnections are required to meet thermomechanical reliability requirements defined by JEDEC standards. Thermomechanical reliability is dependent on the plastic strain experienced in the solder joints. These cyclic strains are due to the mismatch in the coefficients of thermal expansion (CTE) between the package and the board and the applied temperature gradient, as given by the equation below [4]:

$$\Delta\gamma\alpha\frac{L_{DNP}(\alpha_{PWB}-\alpha_{pack})\Delta T}{h} \tag{1.1}$$

   where L_{DNP} is the distance to the neutral point (DNP), between solder joint and the center of the package; h is the solder height; ΔT is the temperature change during each loading cycle; and α_{pack} and α_{PWB} are the CTEs of the package and PCB, respectively, as illustrated in Figure 1.5.

   Accumulation of cyclic strains results in crack initiation and propagation in the solder joints leading to fatigue failures. The use of underfills to lessen plastic deformation in the joints, while widely adopted at chip level, is limited at board level by the need for reworkability. Innovations in materials and stress-relief mechanisms are, therefore,

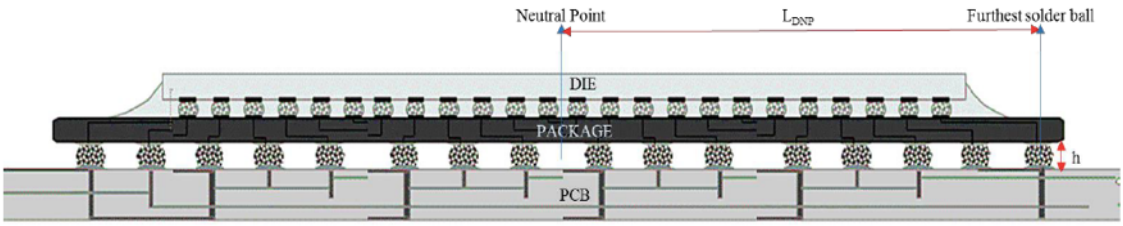


Figure 1.5: Die-package-board interconnections with large LDNP and low stand-off height.

required to address this challenge with minimum system-level impact.

3. Warpage mitigation in assembly:

   Warpage control is critical for board-level assembly yield and system-level reliability. Warpage is aggravated as the package size increases while its thickness is reduced, as is the trend in consumer electronics. Warpage occurs as a result of differential thermal expansion between mechanically coupled materials. While package substrates typically feature some residual warpage build through different fabrication processes with thermal steps such as dielectric curing or electroplating, chip-level assembly contributes for most of the package warpage experienced in board-assembly. In conventional reflow, the assembled package is isothermally heated and cooled with a uniform temperature established through the structure. Package warpage originates during the cool-down phase as solder solidifies and creates mechanical coupling between die and substrate, typically around 150 °C for standard lead-free SAC alloys. Consequently, package warpage induced in reflow is primarily governed by the CTE mismatch between die and glass substrate. This warpage can lead to defects in the solder joints like head-in-pillow defects, stretched solders, collapsed center balls etc. resulting in early failures (Figure 1.6). On the other hand, in emerging thermocompression bonding, heat is applied from the die side only, while the substrate is maintained at a constant stage temperature. A thermal gradient is therefore established in the package assembly giving independent control over the thermal expansion of die and substrate. Consequently, careful design of thermocompression thermal profiles

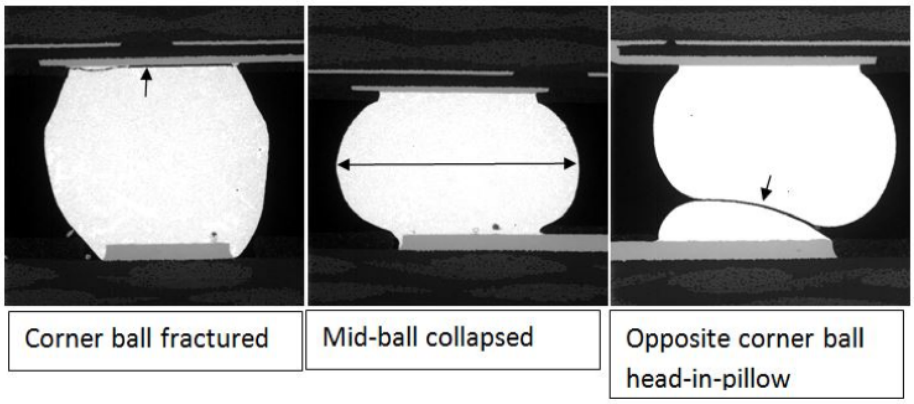


Figure 1.6: Typical defects in solder balls due to assembly-induced package warpage (Courtesy of SEM Lab, Inc.)

can enable minimum warpage when die and substrate expands by the same amount. The conditions giving minimum warpage, therefore, are defined as a function of the CTE of the package. As fundamental understanding of thermal gradients in thermo-compression bonding is still relatively limited, with no existing standards for process design, warpage control in assembly remains a grand challenge that is addressed in this work.

1.5 Unique approach addressing technical challenges

The technical challenges in achieving balanced chip- and- board-level reliability and mitigating package warpage as described in the above section need to be addressed to realize the desired objectives. Direct SMT assembly of a large, 2.5D glass BGA package to the board with system-level reliability is achieved through the following innovations: 1) optimization of the CTE of glass for balanced chip- and board-level reliability; 2) advanced polymer collars and doped solder alloys to further improve thermal cycling performance at board level; and 3) optimization of chip-level assembly processes for warpage mitigation with modeling and empirical correlation (Figure 1.7).

Glass has emerged as a promising solution to enable, for the first time, direct SMT assembly to the printed wiring board of a large 2.5D interposer package, as opposed to

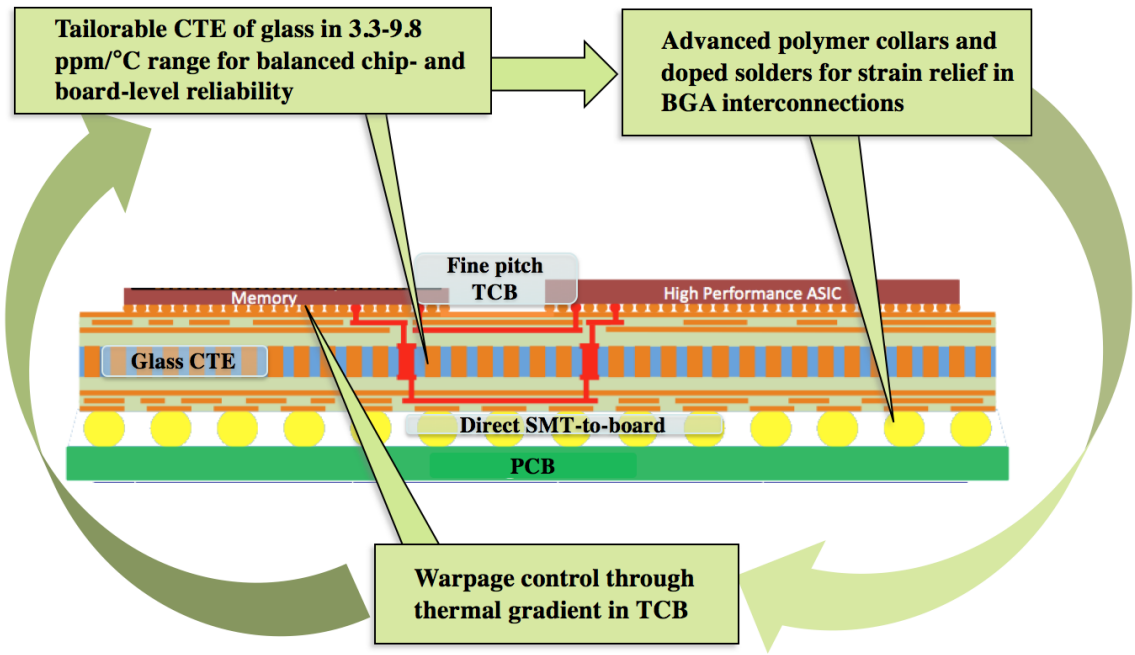


Figure 1.7: Unique approach for balanced- chip and- board-level reliability of a large, 2.5D glass package with direct SMT-to-board.

the conventional Si interposer organic BGA PWB 3-level hierarchy. Benefitting from the unique properties of glass such as tailorable CTE, high-density wiring capability, high modules and mechanical stability, this additional organic BGA package layer can be removed. A parametric thermomechanical model of the 2.5D package was built to identify the optimal glass substrate CTE and thickness. Additionally, the glass CTE greatly impacts package warpage and, subsequently, yield of the SMT assembly and board-level reliability. Further, package warpage after chip-level assembly defines the stress build up in the interconnections as well as the flatness of the package prior to assembly on board. Excessive package warpage can result in yield loss in SMT assembly. Thus, the ideal CTE range for minimum package warpage and system-level reliability is extracted from the models to provide guidelines for design of the glass package to meet the JEDEC reliability and warpage standards.

Even though the board-level interconnection pitch is targeted to be coarser (>500 μm) for 2.5D applications in comparison to single-chip packages, the increase in package size

and decrease in chip-level interconnection pitch, driving towards lower glass CTEs, will adversely impact the board-level fatigue life. It is thus critical to independently improve board-level reliability to meet the reliability requirements. Board-level reliability will be enhanced further by introduction of innovative strain relief mechanisms to improve the thermal cycling performances with minimum change to current processes. These strain relief mechanisms include advanced interconnection materials like doped solders and circumferential polymer collars. Georgia Tech PRC, along with its industry partner Namics Corporation Inc. have recently demonstrated the benefits of circumferential polymer collars. These filler-free epoxies with high-CTE and low modulus act as partial underfills, increasing fatigue life by up to 30%. Since solder properties are greatly dependent on their alloy composition and microstructure, they can be tuned to achieve desired properties without affecting its processability. Recently, solders doped with Mn (SACm™) developed by Indium Corporation have enabled superior thermal cycling performance by constraining the intermetallics (IMCs) growth and stabilizing the microstructure. However, SACm™ alloy is limited in its applicability only to pitches above 500 μm since it is available only in paste form. Bhupender et al. have previously initiated single-chip glass BGA packages with SACm™ alloys as package-to-board interconnections at 400 μm pitches. This work will focus on extending board-level reliability through the aforementioned strain relief mechanisms and performing thermal cycling studies of standard solders in comparison to the innovative novel doped SACm™ solder alloy.

Package warpage after chip-level assembly affects the board-level assembly yield. Hence, it is essential to mitigate warpage to improve yield and system-level reliability. There is little control of package warpage with the traditional mass reflow as opposed to TCB, which enables better process control. Thus, the package warpage can be minimized to obtain a zero net warpage point through optimization of the TCB conditions based on the ideal substrate CTE.

The proposed unique approach realizes the research objectives with parametric model-

ing for providing design guidelines for balanced reliability, improving board-level reliability through novel strain relief mechanisms and optimizing the TCB conditions to minimize assembly warpage at chip-level to prevent yield loss and enhance reliability.

1.6 Research tasks and thesis organization

This thesis document is divided into six main chapters. Chapter 1 introduces the transistor and system scaling trends, objectives, challenges and unique approach to address the technical challenges. Chapter 2 reviews literature of state-of-the-art technologies used to address the aforementioned challenges. The target objectives are achieved through the unique approach proposed above with associated research tasks, organized as the next 3 chapters, respectively addressing the respective technical challenges.

Chapter 3 introduces the design of a novel 2.5D glass BGA package architecture directly mounted on to the board with SMT-compatibility to obtain balanced reliability. This is achieved through finite element modeling for warpage mitigation and enhanced system-level reliability using a parametric approach. This work investigates the effect of glass CTE on the chip- and- board-level reliability thereby extracting the optimum CTE to pass reliability standards.

Chapter 4 evaluates the thermomechanical reliability of a large, single-chip glass BGA package with direct SMT-mount-on-board through a) finite element analysis of fatigue life at chip and board levels as well as warpage mitigation at chip level as a function of glass CTE; b) validating the models with focused reliability studies of low- and high-CTE glass package assemblies; and c) using innovative doped solder materials such as Indium's Mn-doped SACm™ alloy and strain-relief mechanisms to extend board-level reliability to larger body sizes and give more design flexibility.

Chapter 5 investigates the effect of package warpage due to chip-level assembly and minimizes it through a) accurate thermomechanical finite element modeling of TCB process, focusing on the effect of the thermal profiles on warpage and long-term reliability, b)

using these models to estimate substrate warpage at package level, c) validating the models with focused experimental results, and d) mitigating assembly warpage based on these models to achieve superior system-level reliability.

Chapter 6 summarizes the overall research, aligning the results with the research objectives. It finally concludes on the inferences and suggestions for future work, towards achieving balanced reliability.

CHAPTER 2
LITERATURE REVIEW

This chapter provides a background on the recent advancements in IC packaging leading to multi-chip packaging technologies for high-performance computing applications with fine pitch and high I/O densities. The materials and processes needed to achieve such high performance present several challenges along the way. This chapter discusses the innovative solutions proposed by researchers to address the challenges related to the mechanical design of 2.5D packaging, with warpage and reliability considerations.

2.1 Recent advances in multi-chip packaging

Emerging high-performance systems drive the need for high-bandwidth between application processors and memory ICs. The bandwidth demand is predicted to increase exponentially in the next decade [2, 5]. These systems continue to shrink in size forcing the need for such bandwidths to be achieved at smaller form factors, at lower costs with least power consumption. Advances in transistor scaling and miniaturization has led to continuous improvements in electronics systems resulting in further advances in SOC packaging. SOC aims at combining multiple functions on to a single chip. However, SOC poses cost and fabrication limitations in integrating certain IC function's such as DRAMs, RF and MEMS. Further, increase in functional densities are driving the need for larger die sizes at finer pitches. To meet these demands of high bandwidth, 3D architectures with vertical staking of logic and memory dies interconnected by through-silicon-vias (TSVs) were considered. TSVs in such large logic dies requires complex co-design between logic and memory dies and results in high thermomechanical stresses and testability issues without the known-good-dies (KGDs) in the stack. The above cost, thermal and fabrication challenges associated with 3D packaging created a need to explore alternate approaches to

enable high logic-to-memory bandwidth.

To overcome the above challenges due to TSVs in logic dies, another approach was developed involving split-dies with side-by-side interconnections between the logic and 3D memory stacked dies. This 2.5D packaging technology can integrate multi-functional chips at high I/O densities for large, thin package sizes. Logic-to-memory interconnections are achieved through ultra-fine re-distribution layers (RDL) on silicon interposers. It eliminates the need for TSVs in the logic die and employs TSVs only in the passive silicon interposer to connect to the organic BGA package. It also provides design flexibility, scalability, testability and thermal management. However, this 2.5D package present several challenges associated with the mechanical and electrical design for balancing reliability at chip- and- board-level. Xilinx has demonstrated successful 3-level hierarchical 2.5D integration with two 28 nm FPGAs and one transceiver die placed side by side on a 20 mm x 25 mm silicon interposer with minimal heat flux issues [6]. The TSV interposer is connected to a 35 mm x 35 mm large, high-CTE ceramic substrate through optimized thermocompression bonding process to mitigate the warpage (Figure 2.1) and the dies are stacked horizontally through traditional mass reflow process. The importance of assembly process conditions to mitigate warpage and the reliability challenges associated with high-CTE ceramic substrates were evaluated.

SK Hynix's HBM as shown in Figure 2.2 also illustrate a 3-level hierarchy with four HBM stacks and a GPU assembled on a 30 mm x 38 mm Si interposer [7]. The additional organic package layer in both of the above systems adds to the system cost apart from inducing a large CTE mismatch at package level. This CTE mismatch leads to warpage and yield issues adversely impacting the reliability of the system.

Glass as a substrate platform is extensively being researched due to its tunable CTE, dimensional stability and high-density wiring capabilities. Using glass, a 2-level hierarchy can be built to overcome the challenges of the additional package layer. Georgia Tech's PRC successfully demonstrated a 2-level 2.5D glass BGA package that can be directly

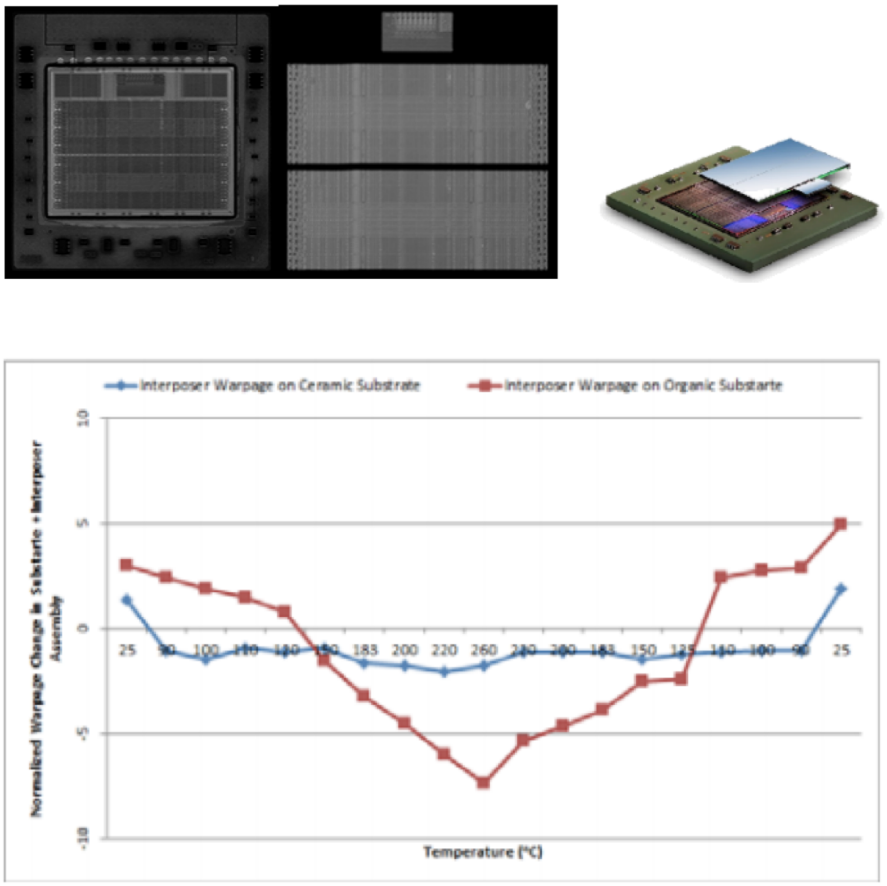


Figure 2.1: 3-level hierarchical 2.5D integration with two 28 nm FPGAs and one transceiver die placed side by side on a 20 mm x 25 mm silicon interposer assembled on high-CTE ceramic substrate (Xilinx).

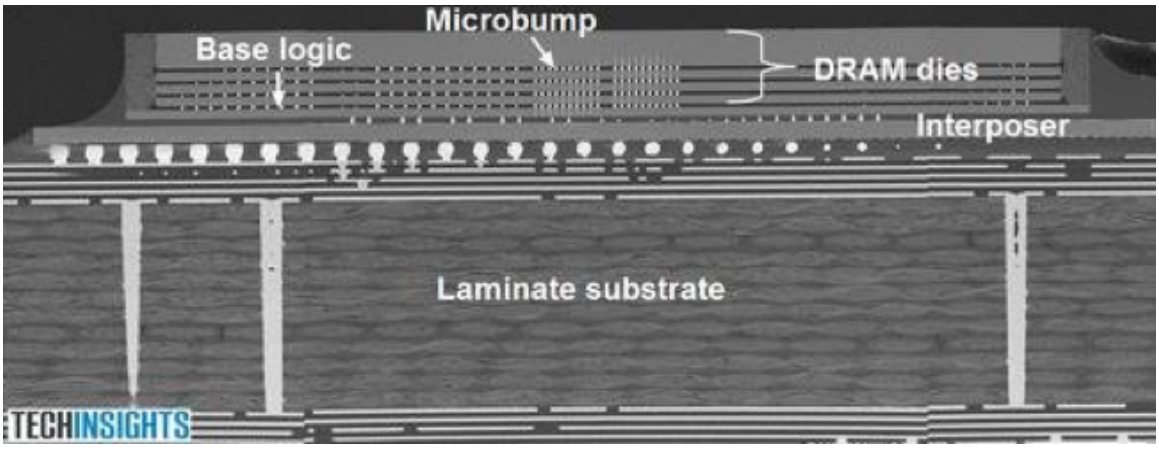


Figure 2.2: High-bandwidth memory stacks with a center GPU on a silicon interposer assembled on an organic laminate (SK Hynix).

SMT-mounted on the board (Figure 2.3).

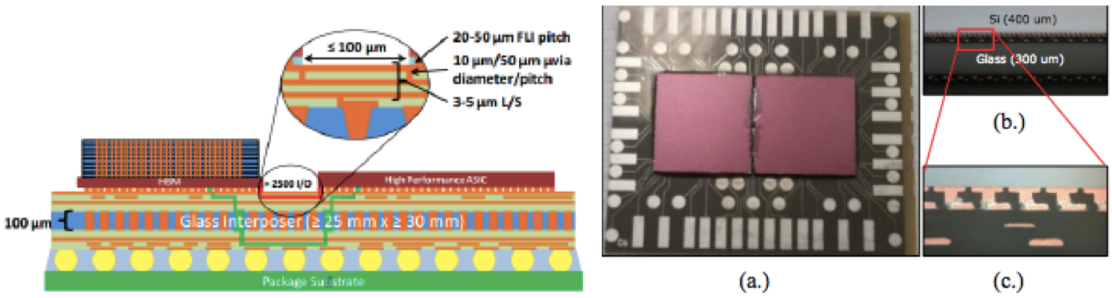


Figure 2.3: 2.5D FLI assembly at 100 μm die-to-die spacing (a.) top view of 25 mm x 30 mm six-metal-layer glass interposer (b) cross-section (c) FLI detailed cross-section.

Glass with low- and- high-CTE are readily available today in large panel sizes or roll-to-roll form. However, in order to balance chip- and- board-level reliability as well as mitigate warpage, the glass CTE needs to be tuned. Low-CTE glass creates a large CTE mismatch at board-level, thus degrading its reliability. On the other hand, high-CTE glass induces high warpage at chip-level, subsequently resulting in yield loss and chip-level reliability concerns. TSMC performed a finite element analysis of a glass interposer for a high-performance flip-chip BGA (HP-fcBGA), varying the glass CTE to obtain a middle CTE for least stress build-up [8]. They evaluated the effect of glass CTE on different stresses experienced by the BGAs. Based on their analysis, a middle CTE of 8.3 ppm/K generated least stress because of the CTE mismatch between the glass and die and PCB were close, which relieved the thermomechanical stresses accordingly. Further, IBM recognized the importance of minimizing stress based on the effect of interposer material, CTE, size for mechanical design of the package [9]. Similar to TSMC, IBM also conducted a finite element analysis with a parametric approach, evaluating the glass substrate CTE on warpage and the Von Mises stress at the solder joint. Glass CTE of 6 ppm/K was found to have least Z-displacement and the von Mises stress reduced by 18% when the CTE was increased from 3.8 ppm/K to 6 ppm/K (Figure 2.4 and Figure 2.5).

The packaging industry has been innovating several materials, fabrication and intercon-

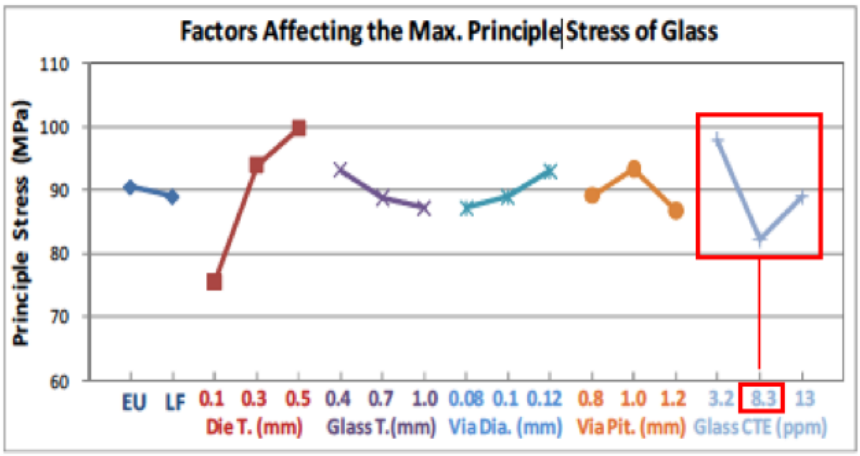


Figure 2.4: Parametric effects on the maximum principle stress of glass.

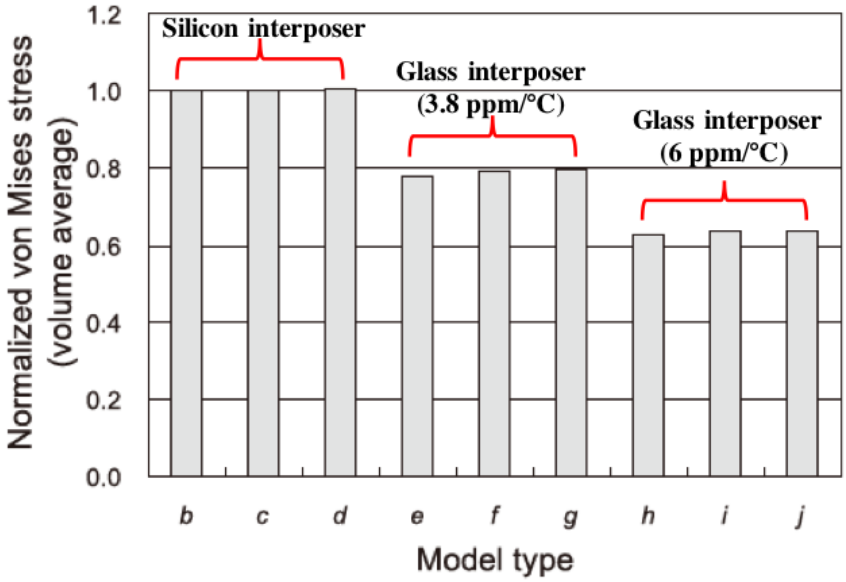


Figure 2.5: Von Mises stress in volume average at the corner solder ball.

nection techniques to address the warpage and reliability challenges held by 2.5D packaging that will be discussed in detail in the following sections.

In addition to the 2.5D technology, researchers have also been looking at other approaches for high-density interconnections between heterogenous dies on a single package including (a) Intel's Embedded Multi-die Interconnect Bridge (EMIB) [10], (b) fan-out wafer-level packaging platform (FOWLP) [11], (c) Amkor's Silicon-Less Integrated Module (SLIM) [12] and (d) Silicon Wafer Integrated Fan-out Technology (SWIFT) [13]. These

different technologies are as shown in Figure 2.6.

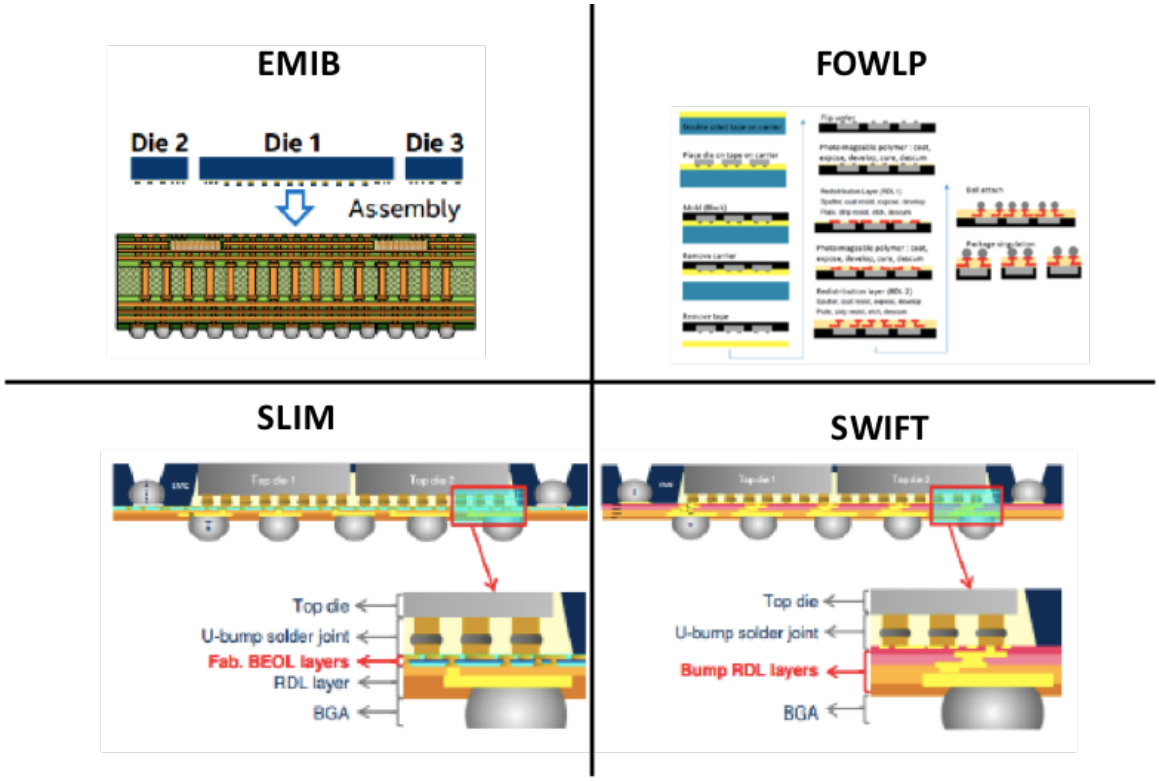


Figure 2.6: Intel's EMIB assembly (top left), Process flow to achieve FOWLP - Courtesy of Beth Keser (top right), Amkor Technology Inc's SLIM (bottom left) and SWIFT (bottom right).

2.2 Material developments in interconnections for board-level reliability

In emerging high-performance systems, the split-die trend is driving high-density interconnections at sub-5 μm pitch at large body sizes. Thus, improving board-level reliability without adversely impacting chip-level reliability has become a major concern. Additionally, due to reduction in pitch at board-level to below 400 μm, higher strains and warpage is experienced in BGAs, further degrading reliability. Advances in materials development used to enhance board-level reliability are discussed in this section.

2.2.1 Advances in surface finishes and doped solders

Standard solders used in BGAs are called SAC alloys that comprise of tin, silver and copper that are used to form metallurgical bonds between the copper pad on the package and the board. These copper pads are coated with a thin layer of surface finish in order to provide least surface tension for the solder to wet the pads effectively. The surface finish also takes part in intermetallics (IMCs) formation and forms the interface between the pad and bulk solder. After reflow, the solder melts and partially dissolves in the surface finish, forming IMCs due to diffusion phenomena. These IMCs are brittle in nature and their composition, thickness and microstructure greatly affect the strength and reliability of the joint. Pad surface finishes play an important role in determining the reliability of the solder joint. Traditional surface finishes include hot air solder leveling (HASL), organic solderability preservative (OSP), immersion Sn (ImSn) and immersion Ag (ImAg) [14]. Due to their limitations like non-uniformity, cost, solderability, shelf-life, surface oxidation, etc., advanced in surface finishes have been made. The most popular amongst these are electroless nickel immersion gold (ENIG) and electroless nickel electroless palladium gold (ENEPIG). Literature studies have been made on the effect of ENIG and ENEPIG on reliability based on their IMC formation. Even though they address the challenges posed by traditional surface finishes, they impede pitch scalability with sub-10 μm gaps between traces. Another novel surface finish technology developed by Atotech GmBH known as electroless Pd autocatalytic Au (EPAG) enables high-density wiring at finer pitches. The thickness of IMC depends on the reflow time and number of reflow cycles [15]. The IMCs grow further during thermal ageing and the presence of these brittle IMCs act as crack initiation points leading to interfacial failure [16]. Control of IMC growth and consequent control of microstructure is critical in enhancing board-level reliability.

Apart from advances in surface finish, development in solders by refining their microstructure have also been researched. As mentioned in Chapter 1, with current standard solders, higher drop performances can be achieved only at the cost of thermomechanical

fatigue life. Doping these standard SAC alloys with Nickel (Ni), Manganese (Mn), Cerium (Ce), Bismuth (Bi), Titanium (Ti), Germanium (Ge) and Yttrium (Y) have been studied [17]. Due to their low solubility and low melting points, they refine the solder microstructure upon reflow. During solidification, they make the grains finer and limit the growth of IMCs, improving the tensile and creep resistance. Among the various dopants, it was found that Mn and Ti doped SAC alloy reduced the undercooling required for the initiation of crystal nucleation and also hindered the formation of detrimental intermetallics like Ag3Sn and suppressed the thickness of Cu6Sn5. Microstructures of SAC305, SAC105 and SAC105 doped with different compositions of Mn and Ti are observed in Figure 2.7 [18]. With different variations of dopants, Liu and Lee (Indium Corporation) found that SAC105 + 0.13% Mn displayed the best composition for balanced thermal cycling and drop test reliability.

The drop test and thermal cycling results are shown in Figure 2.8 and Figure 2.9 respectively. Indium corporation commercialized SACm™ based on their experimental findings on the Mn dopant as an alternative to SAC105, bridging the gap between drop and thermal fatigue performance [19]. The superior thermal cycling reliability of SACm™ was attributed to the refinement of microstructure, suppression of IMC growth and improved joint strength. Currently, SACm™ is only available in paste form which limits its applicability to pitches above 500 μm.

2.2.2 Advances in underfill materials at board-level

Underfills are generally used at chip-level to enhance the reliability by absorbing and redistributing the stresses away from the solder joints. Reworkability is an important factor in board-level assembly. However, underfills prevent reworkability and are not widely used for board-level interconnections. Thermally reworkable underfills are being developed to improve reliability to the flip-chip package while allowing the chip to be easily removed at elevated temperatures. Wong et al used two approaches being development of thermally-

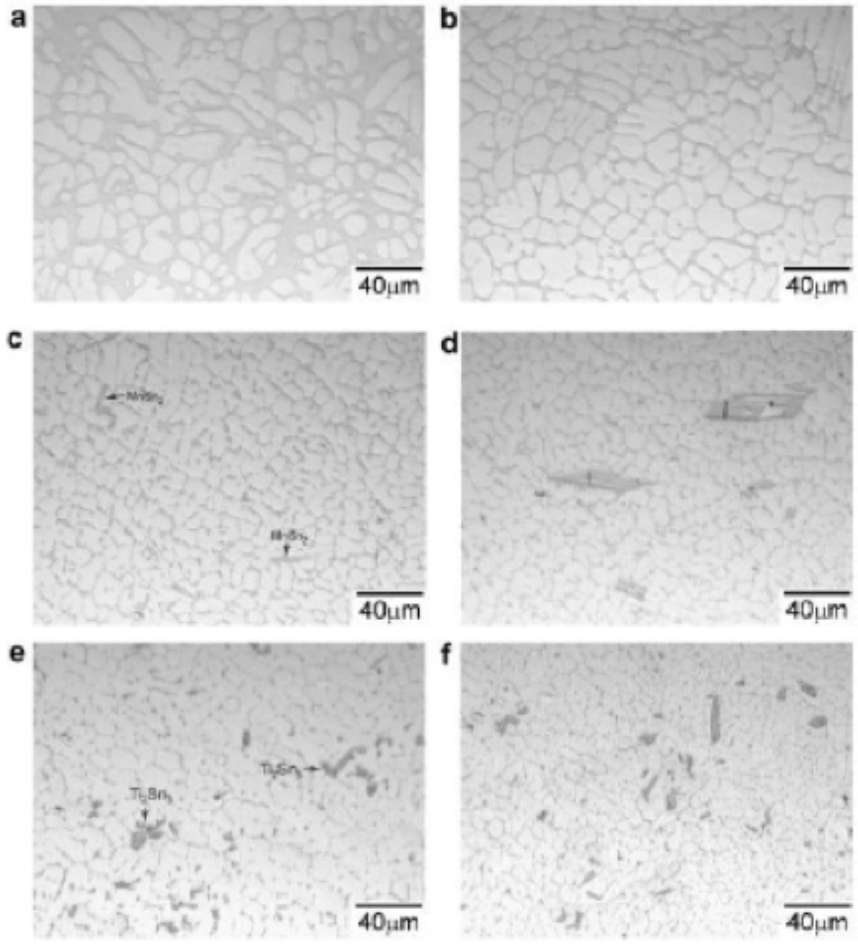


Figure 2.7: Microstructures of (a) SAC305, (b) SAC105, (c) SAC105 + 0.15Mn, (d) SAC105 + 0.5Mn, (e) SAC105 + 0.15Ti and (f) SAC105 + 0.5Ti.

cleavable-block-containing epoxies and additive-modified epoxies. They observed that by combining the two approaches, both die removal and underfill removal were achieved [20]. Peng et al studied the effects of snap cure underfill and reworkable underfill. They showed that reworked underfill had least percentage of failures when compared to snap cure and no underfill packages as shown in Figure 2.10 [21]. However, reworkable underfill takes longer to cure and causes reliability issues. Thus, their reliability concerns need to be addressed by reducing curing time, increasing molecular weight and extending their application at wafer-level.

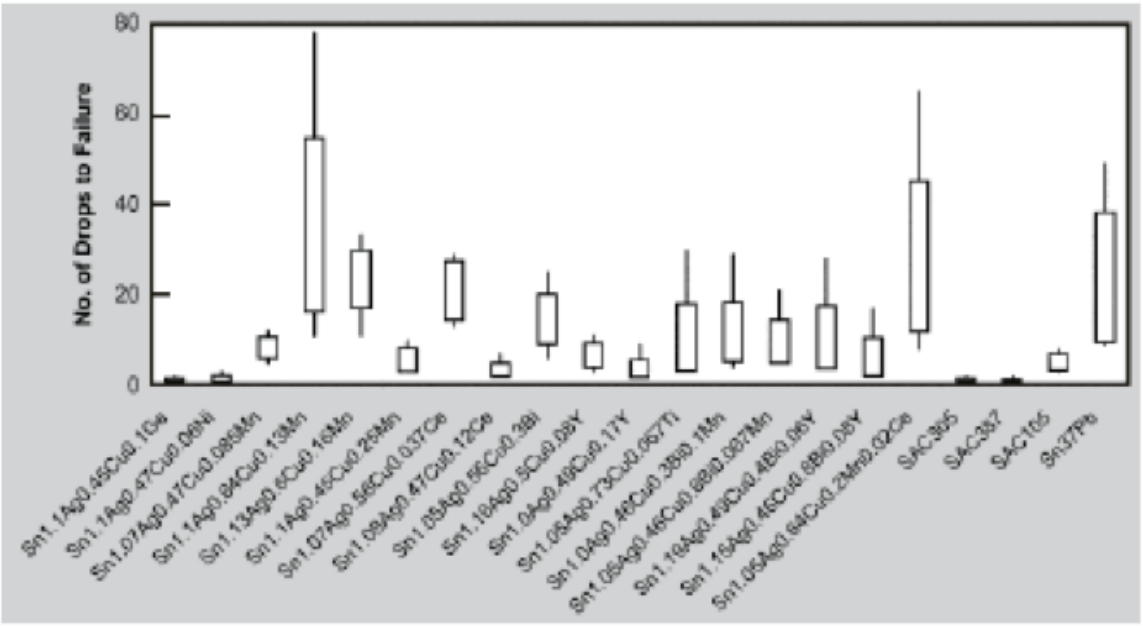


Figure 2.8: Drop test reliability of SAC solder with dopants.

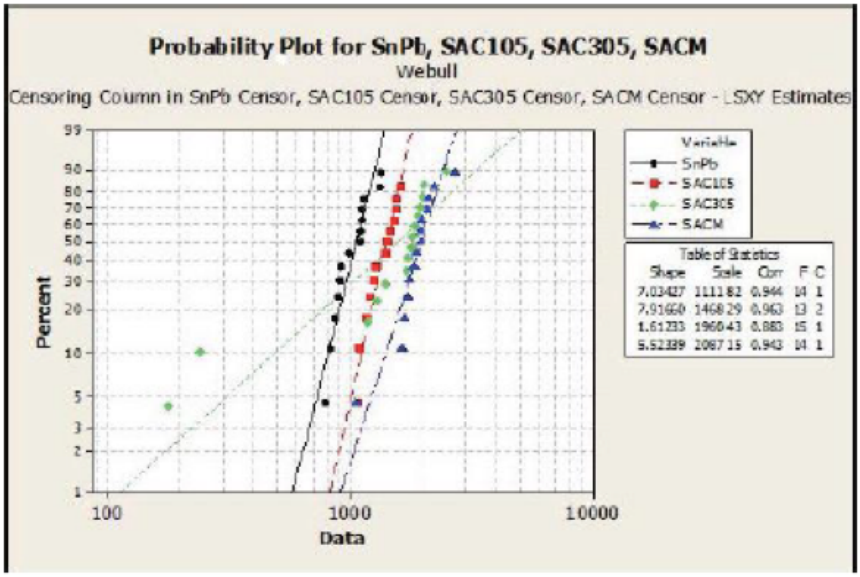


Figure 2.9: Thermal cycling results for SnPb, SAC105, SAC305 and SACm™.

2.2.3 Compliant interconnections

Standard solders and interconnection techniques are reaching their limits to achieve advanced computing and high-performance application needs. A variety of compliant interconnections have been investigated to mitigate the stresses induced by such high-functional

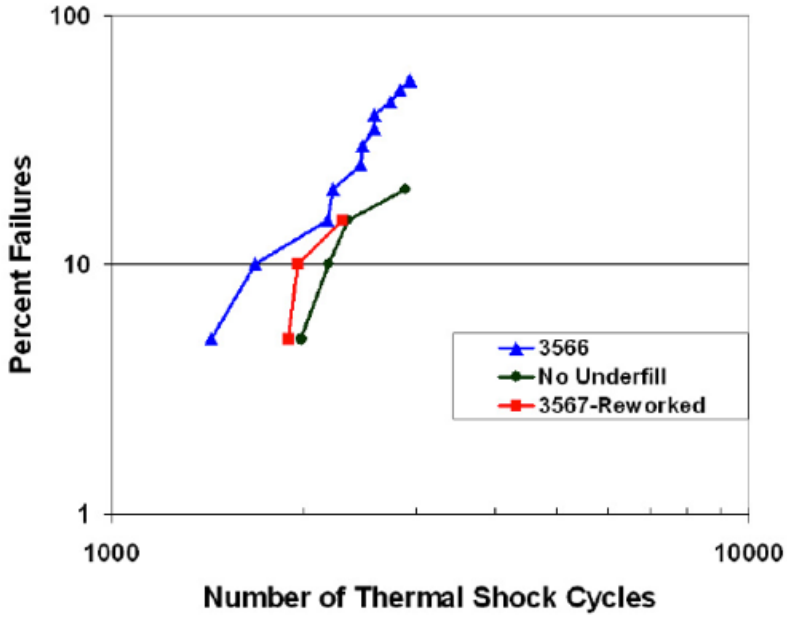


Figure 2.10: Percent failures for different underfill types.

requirements. Compliant solder balls have been reported consisting of a polymer core instead of the bulk metal to reduce the effective elastic modulus of the joint (Figure 2.11). This also increases the CTE and redistributes the stress uniformly. A 2x - 3x improvement in thermal cycling was observed as compared to SAC305 [22]. The complex fabrication and cost limitations prevent this approach from being adopted commercially.

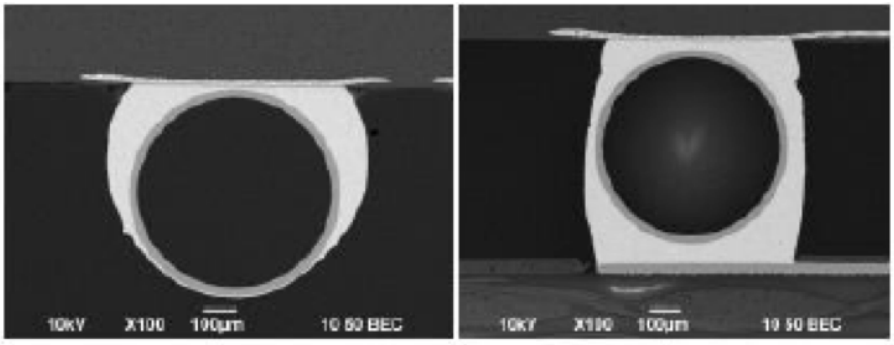


Figure 2.11: Cross-section of polymer core solder ball after balling (left) and assembly (right).

Another compliant interconnection technique developed by [23] involves differential displacement that mechanically decouples the die and substrate. It accommodates the CTE mismatch between the die and substrate during thermal cycling by easily deforming in

the x, y and z directions to provide stress relief in the interconnections as shown in the Figure 2.12.

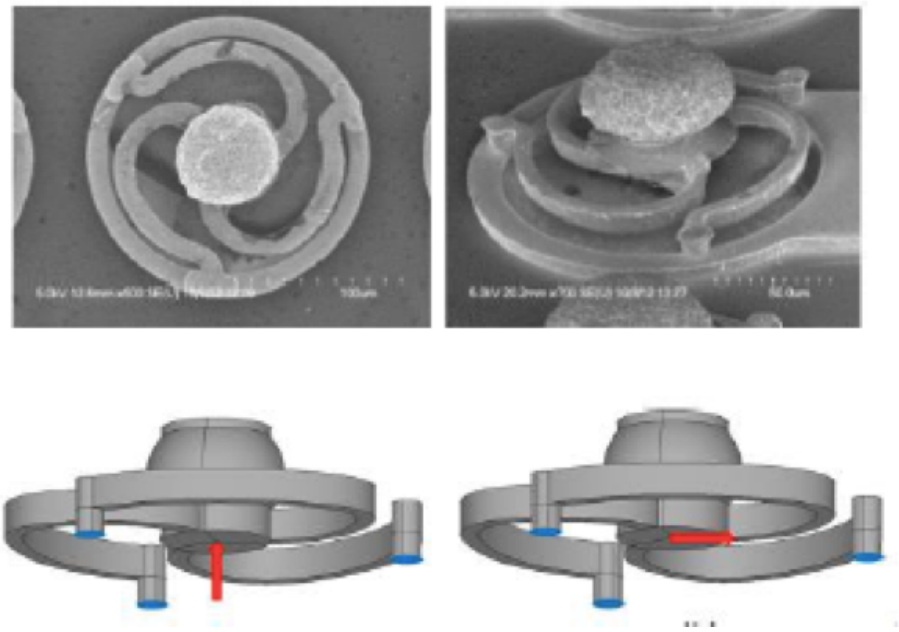


Figure 2.12: Multi-path complaint structures - SEM image (top), simulations (bottom).

Other metal-based compliant interconnections have also been reported including the micro-spring [24], stress-engineered [25] interconnections, double-ball wafer-level packaging (WLP) [26], bed of nails [27] etc.

Previous work on board-level compliant interconnections carried out at Georgia Tech's Packaging Research Center includes compliant micro-wire arrays and circumferential polymer collars. Copper micro-wire arrays eliminated the need for board-level underfill. They were pre-fabricated in ultra-thin carriers and then assembled as a stress-relief interlayer between the package and board [28] (Figure 2.13).

Addition of low-modulus dielectric build up layers on either side of the glass substrate further enhance the thermomechanical reliability by increasing the effective CTE of the substrate, reducing the CTE mismatch to the board. Circumferential polymer collars were used as to relieve stress by spin-coating on to the BGAs after balling (Figure 2.14). They act as partial underfills by redistributing the stress while maintaining reworkability [29].

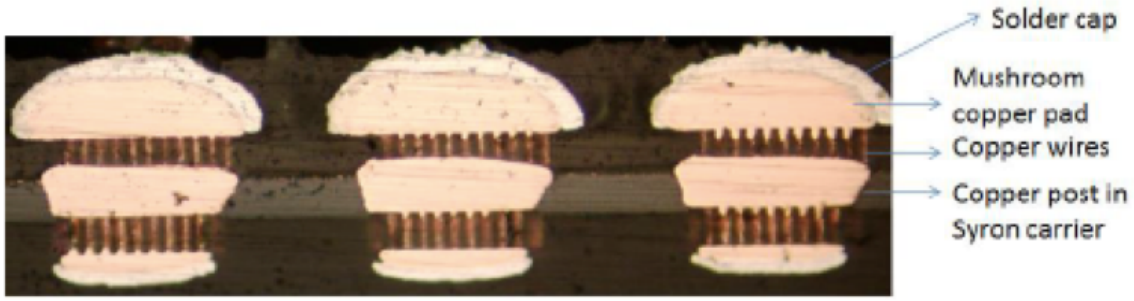


Figure 2.13: Cross section of MWA interconnections.

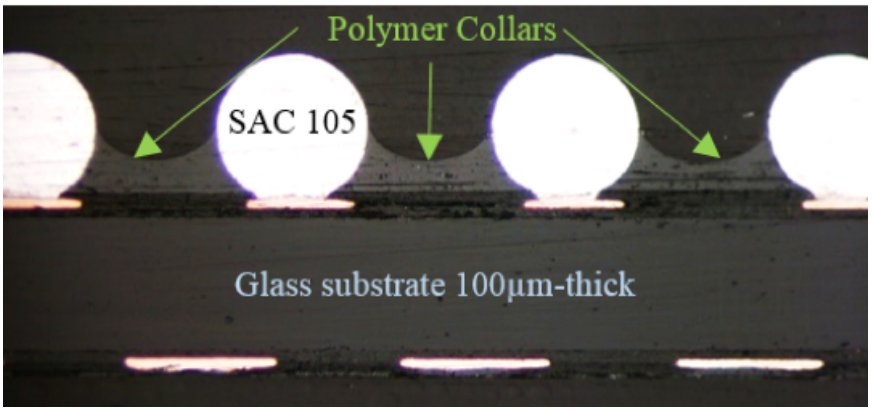


Figure 2.14: Cross-section of an 18.4 mm interposer with uniform collar formation achieved after spin-coating profile optimization.

2.3 Warpage mitigation in assembly

The trend to meet low cost, high density, speed and bandwidth at low power for emerging smart mobile and high performance systems is driving package sizes towards large, ultra-thin architectures. High-CTE organic packages are currently being used as the substrate core. Traditional organic laminates such as FR-4 or BT, with build-up epoxy dielectric based re-distribution layers (RDL) are the main substrates used for fine pitch BGA packages [30, 31]. But the drawbacks associated with it are limiting its use for high performance needs. The demand for thinner packages for ultra-thin substrates leads to substrate warpage before and after assembly. A major substrate parameter affecting warpage of the package is the substrate coefficient of thermal expansion (CTE). This CTE mismatch between the die and substrate causes warpage. High-CTE organic packages are often thick with a core

thickness ranging from 500 μm to 1 mm and are not sufficiently rigid to meet the JEDEC warpage standards. Therefore, it can be assumed that by lowering the CTE mismatch, package warpage can be reduced. Low-CTE silicon core substrates have become increasingly popular due to their high elastic modulus, high stiffness at high temperatures and their ability to be manufactured at high-wiring densities. They are shown to display good performance and high interconnection reliability to achieve excellent mounting capability of semiconductor devices [32]. It has been shown that low CTE core substrates have lower warpage than high CTE core substrates [33]. However, these low CTE substrates transfer the CTE mismatch to board inducing large stresses at the BGAs, interfacial delamination and adversely affecting the solder joint reliability. To overcome this issue, interposers are being developed. For interposers with ultra-high I/O density and 2 μm lines and spaces RDL, silicon wafers fabricated in the back end of line (BEOL) have been demonstrated [34, 35, 36]. These interposers are then mounted on to an organic substrate and connected to the board. This involves an additional assembly step and aggravates warpage, making it critical to optimize process conditions to minimize assembly warpage. Mass reflow is the most common process used in assembly at chip level. The other process that are being developed is the thermocompression bonding (TCB) process. Mass reflow technology cannot be controlled as efficiently as TCB, since the latter can be optimized on the basis of various parameters like force, ramp rate, temperature profile (heating and cooling rates) [37]. Mass reflow is generally used with capillary underfill (CUF) while TCB can be used with CUF or pre-applied underfill materials. Satomi et al. have demonstrated that the warpage induced by traditional mass reflow is higher than that through an optimized TCB process for an organic substrate package as seen in Figure 2.15 [38].

Researchers have also studied ways to minimize warpage during the reflow process. IBM has successfully demonstrated warpage mitigation in mass reflow through the use of a vacuum fixture with a carrier on which the package is placed. The carrier has holes as depicted in Figure 2.16 through which the vacuum permeates to minimize concave warpage.

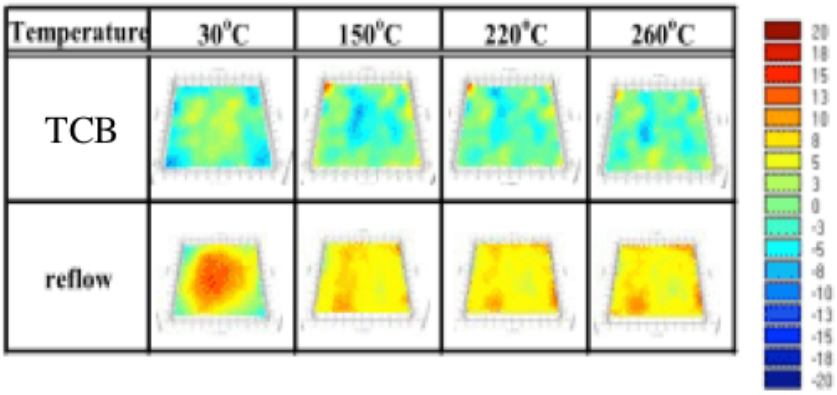


Figure 2.15: Comparison of warpage during thermocompression bonding and mass reflow.

Atmospheric pressure is applied from the top that pushes down the top surface uniformly to prevent convex warpage. Cross sections in Figure 2.16 show the uniformity in the C4 bump height throughout the package with minimum warpage. However, as we move towards 2.5D packages, the chip-level pitches reduce drastically, making it harder to assemble via mass reflow process.

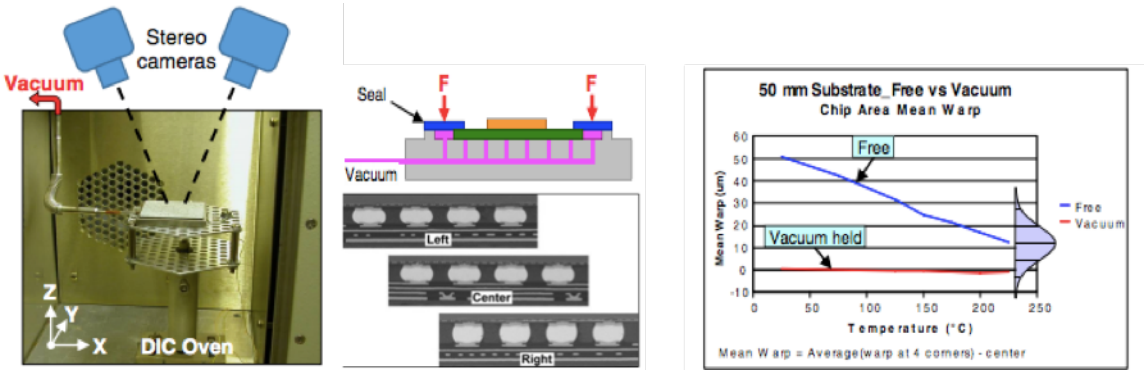


Figure 2.16: Vacuum fixture with carrier during mass reflow (left), uniform ball height cross-section (middle), warpage with and without vacuum fixture (right).

Thermocompression bonding process is typically used at such fine pitches. Even TCB is shown to mitigate warpage better as compared to reflow, it is still extremely critical to optimize the conditions based on the package type and pitch to achieve yield and system-level reliability. Liang Wang et al. have investigated the effect of TCB conditions on warpage [39] and demonstrated that the stage temperature and subsequent peak temperature has a

strong effect on assembly-induced warpage. For a 2.5D package with silicon interposer, it was shown that a lower stage temperature resulted in higher warpage and cracked solder joints (Figure 2.17). On the other hand, higher stage temperatures warped lesser than 12 μm and formed good solder joints. Precise and dynamic control of thermal gradient in TCB is critical in achieving reliable solder joints.

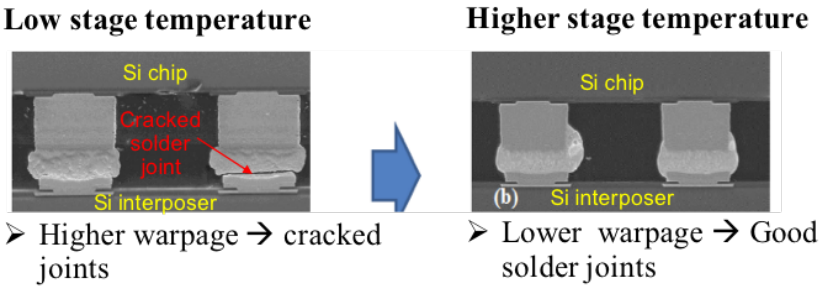


Figure 2.17: Interposer warpage as a function of TCB stage temperature.

Warpage induced during chip-level assembly subsequently affects yield and reliability at board-level. Studies have shown that the sum of warpage and variation in ball diameter account for almost all the non-planarities [40]. It is thus necessary to quantify the total warpage based on the package size, BGA pitch and BGA height in order to determine the successful assembly at board-level. While employing 2.5D and 3D packages with large size silicon interposers, the warpage caused by the CTE mismatch between the package and organic substrate is one of the most significant problems. To prevent cracks and solder bump failures, the bonding sequence is critical in addition to the TCB process conditions. Invensas has investigated different assembly sequences with chip-first, chip-last and the use of a permanent carrier to mitigate warpage (Figure 2.18 and Figure 2.19). Kei Murayama et al. [41] have also studied the effect of assembly sequence on warpage with the chip-first and chip-last process and highlighted the importance of defining the sequence to achieve superior reliability.

Bhupender et al. studied the chip-first assembly warpage for a single-chip glass BGA package after chip-level and consequent board-level assembly at low- and- high-CTE glass

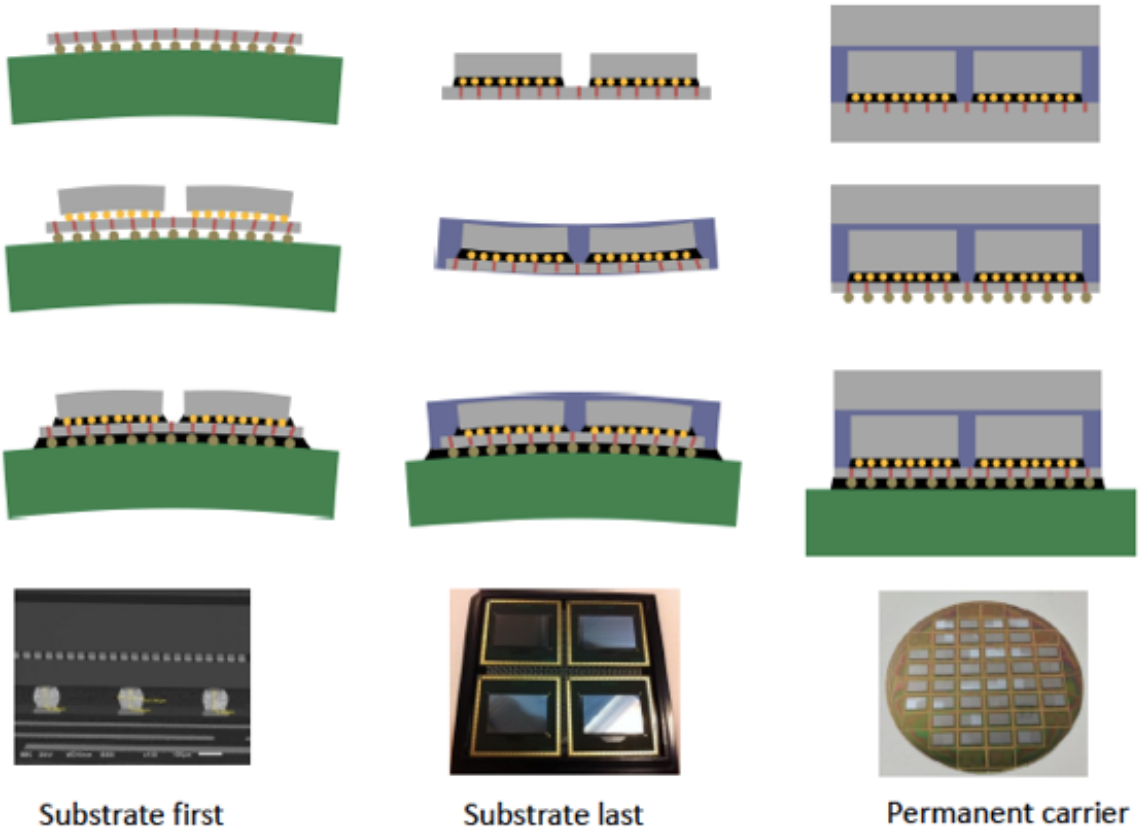


Figure 2.18: Different assembly sequences for warpage mitigation.

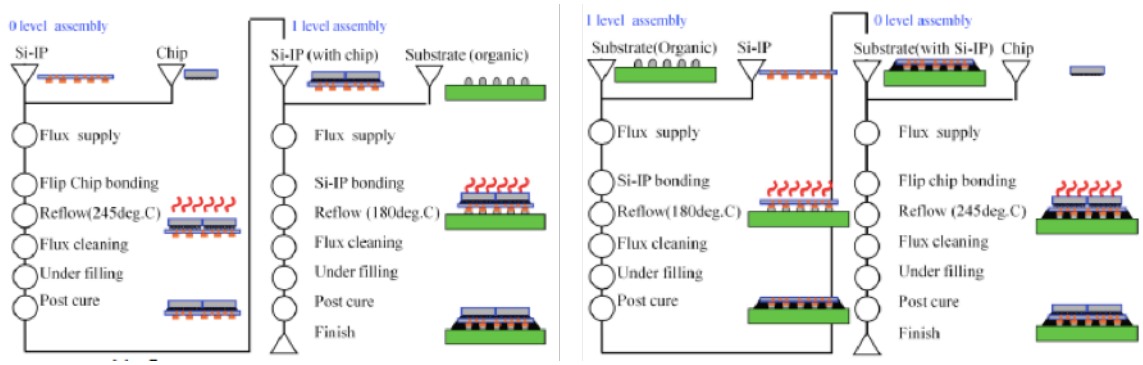


Figure 2.19: Assembly sequence process flow with chip-first and chip-last.

substrates [42]. The warpage after chip-level assembly for low-CTE glass was the least; but showed highest warpage when assembled to the board. High-CTE glass, on the other hand, had maximum warpage after chip-level assembly and resulted in yield loss when assembled on to the board. Although, the high-CTE glass samples that yielded, warped the least (Figure 2.20). It is thus necessary to define the assembly process and sequence based

on the substrate CTE to prevent yield loss and enhance reliability.

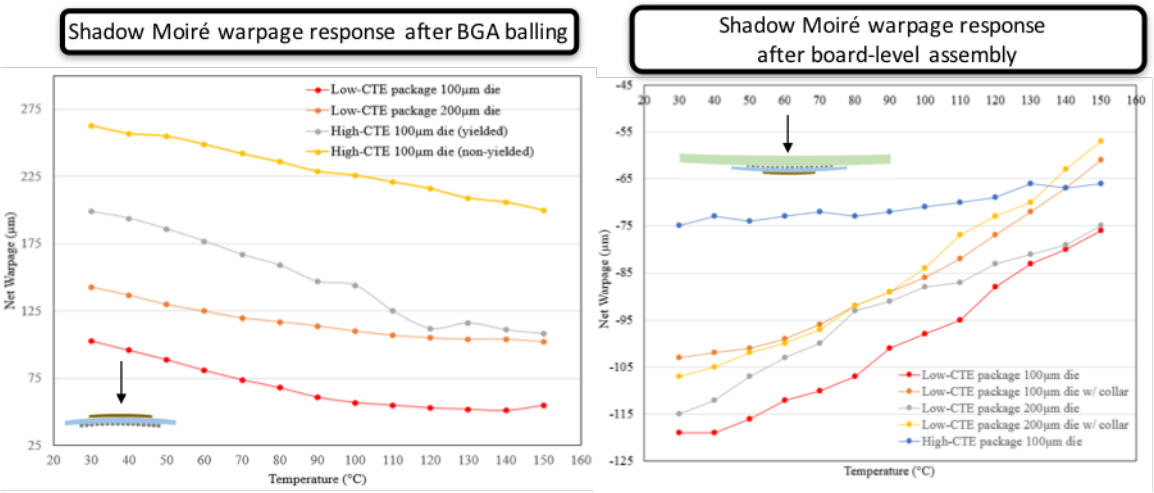


Figure 2.20: Shadow Moire warpage response after BGA balling (left) and after package-to-board assembly (right).

In addition to controlling the assembly process and sequence, studies have been made on the effect of solder melt temperature on warpage. Solders with lower melting points require lower reflow temperatures and time, thereby reducing the temperature gradient while cooling from the stress-free temperature. Since warpage depends on the CTE and temperature gradient, warpage can be significantly reduced by lowering the gradient. Shinko [42] has demonstrated a low-temperature solder, Sn57wt%Bi, and evaluated the warpage and reliability behavior compared to the standard SAC305 solder alloy. Their results indicated that the Sn57Bi solder warped significantly lesser while maintaining the thermomechanical reliability as that of the standard SAC305 solder (Figure 2.21).

Other factors like fabrication defects, variation in raw substrates and underfill properties (Tg, modulus, CTE) from different suppliers also need to be factored in while evaluating and minimizing warpage [30].

2.4 Summary

This chapter discussed in detail the advances in IC packaging from single-chip to multi-chip packaging. Demonstration of a reliable 2.5D package in a 3-level hierarchy with minimum

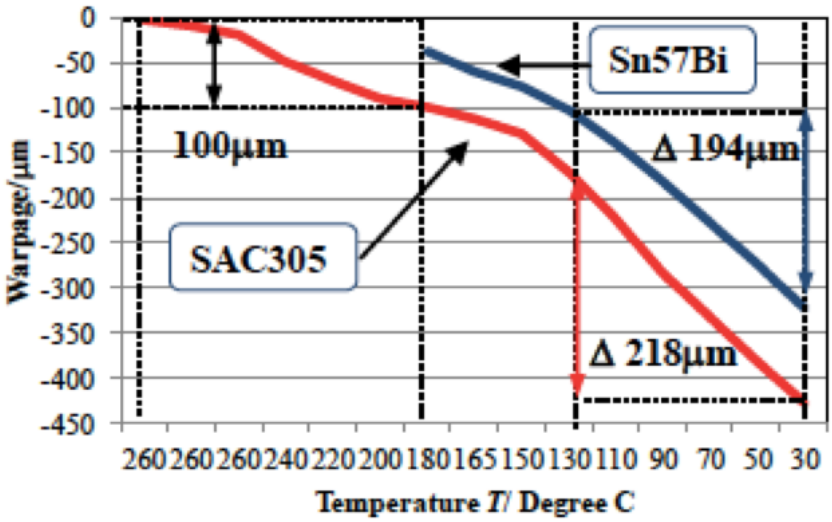


Figure 2.21: Warpage response of standard SAC305 vs low-melting Sn57Bi solder

warpage and enhanced board-level reliability are identified as key challenges. Various solutions in literature proposed to address these challenges have been extensively reviewed. Critical challenges related to a) balancing reliability at chip- and- board-level, b) achieving board-level interconnections with simultaneous TCT and drop-test reliability, and c) assembly process design to minimize warpage need to be addressed to achieve system-level reliability.

CHAPTER 3
MODELING AND DESIGN OF A 2.5D GLASS PACKAGE FOR BALANCED CHIP- AND BOARD-LEVEL RELIABILITY

This chapter details the test vehicle dies and substrate design that will be used to model a 2.5D glass BGA package for analysis of thermomechanical reliability. A design methodology will be developed to perform a parametric analysis of fatigue life and warpage of the package as a function of the substrate CTE. The model will ultimately provide assembly design rules for achieving balanced reliability at minimum warpage.

3.1 Test vehicle design of a 2.5D glass BGA package

3.1.1 Daisy chain test dies

The 2.5D assembly test vehicle is a GPU emulator consisting of two types of dies: one center logic emulator and four stacked memory emulators as given in Figure 3.1 (Courtesy of Brett Sawyer). The logic emulator, 19 mm x 20 mm in body size and 740 μm in thickness is designed with an exterior staggered array of 35m pitch and an interior array at 100 μm pitch (Figure 3.2 top). Copper pillar interconnections with a bump count of 24880 I/Os were designed with a diameter of 20 μm, 20 μm copper height, a Ni barrier layer of 2 μm and a SnAg solder cap of 10 μm. The stacked memory die is a high-bandwidth memory emulator, 5.5 mm x 7.7 mm in size, with a thickness of 740 μm and a bump count of 4943 I/Os arranged in a staggered array at a pitch of 55 μm (Figure 3.2 bottom). The copper pillar interconnections with Ni barrier and SnAg solder cap has the same height as that of the logic emulator with a diameter of 25 μm.

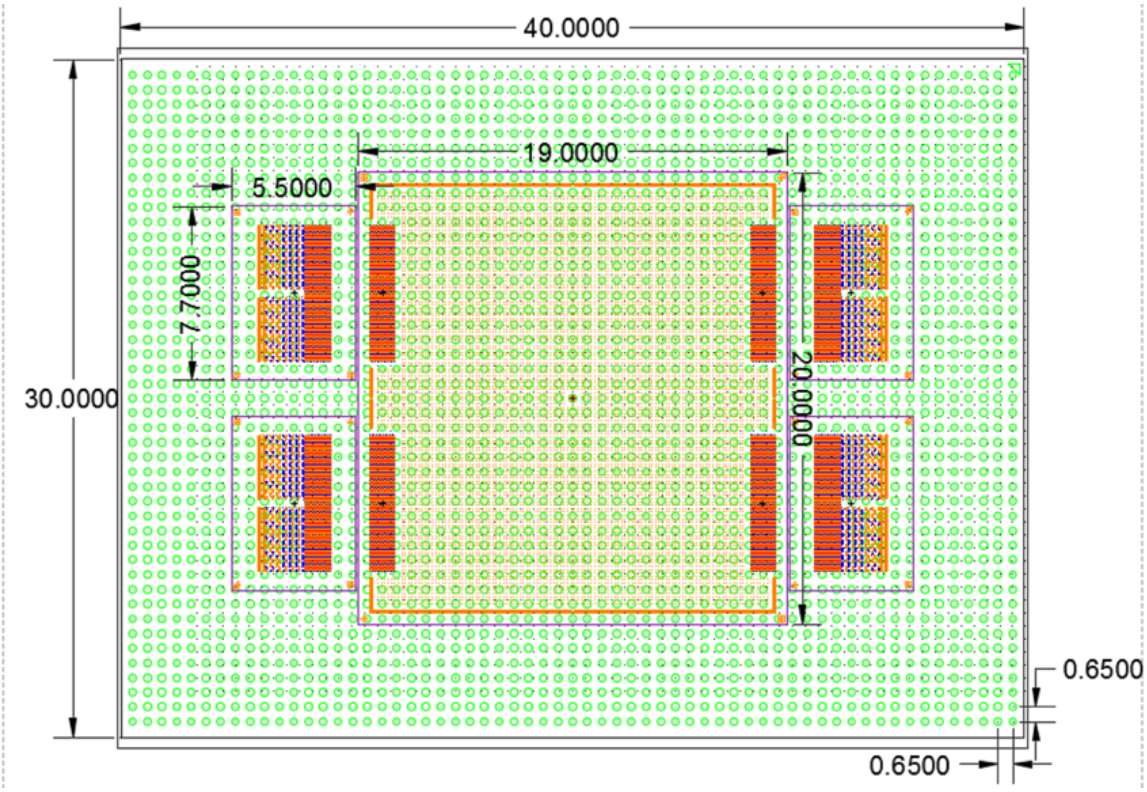


Figure 3.1: Assembly test vehicle layout of 2.5D glass BGA package with center logic emulator and 4 stacked HBMs on the periphery.

3.1.2 Test vehicle substrate stack up

The glass substrate with a body size of 40 mm x 30 mm is proposed with a varying core thickness of 50 μm to 300 μm in electrical and mechanical design. The substrate has the die footprint on the top side and an area array of daisy-chain BGA interconnections with 2745 I/Os at 650 μm pitch and 500 μm diameter on the bottom side. This proposed I/O count is similar to that of a 2.5D silicon interposer with 50 mm x 55 mm in size at 1 mm BGA pitch. The substrate design and schematic of the stack up is shown in detail in Figure 3.3.

3.2 Design methodology

One of the main parameters affecting the fatigue life of the package is the substrate CTE. Taking advantage of the tunable CTE of the glass substrate, a range of CTEs between 3.3 ppm/C to 9.8ppm/C, a design methodology (Figure 3.4) is developed to parametrically

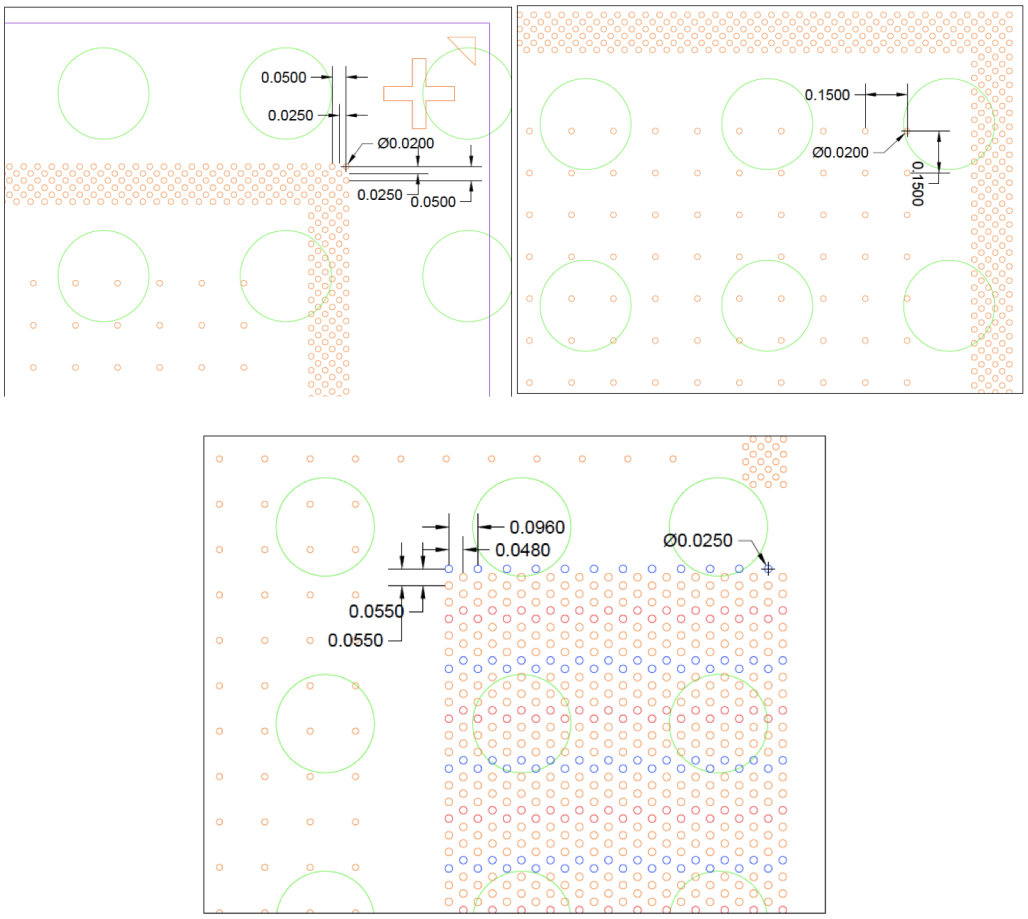


Figure 3.2: Logic emulator with exterior (top left) and interior (top right) array design parameters; high-bandwidth memory emulator die design (bottom).

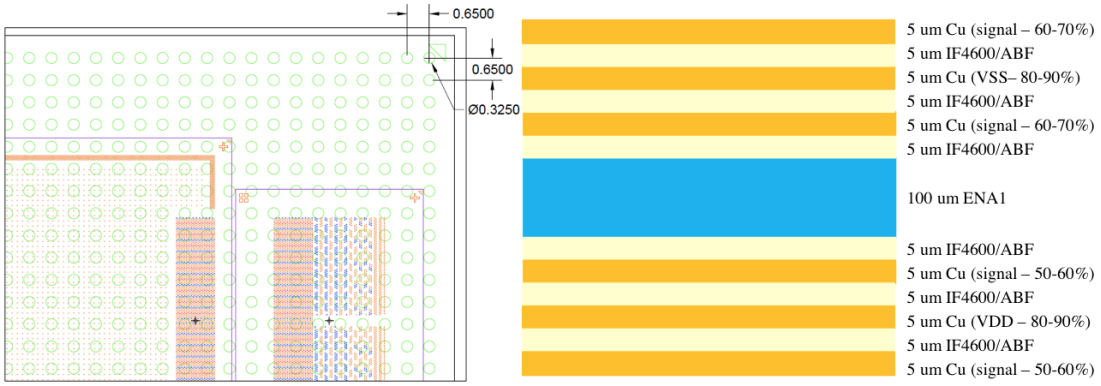


Figure 3.3: Substrate design (left) and substrate stack-up (right).

evaluate the thermomechanical reliability and warpage of the system. The effect of substrate CTE was considered for the chip-first assembly sequence which is most widely used in industry. The CTE mismatch induced at chip- and- board-level depending on the substrate CTE determines the amount of stress experienced at the solder joints. Additionally, higher CTE mismatch leads to increased warpage, leading to yield issues. Therefore, it is critical to find the ideal CTE to balance reliability and mitigate warpage. This is done through thermomechanical modeling of a 2.5D glass package to obtain a) the fatigue life at chip- and- board-level as a function of CTE in order to extract the optimum CTE for balanced reliability, b) room temperature warpage after chip-level assembly as a function of CTE, and c) maximum permissible warpage by JEITA standards to prevent open solder joints for the subsequent board-level assembly. This methodology will ultimately provide a range of CTEs for design of the assembly process with minimum warpage and enhanced system-level reliability.

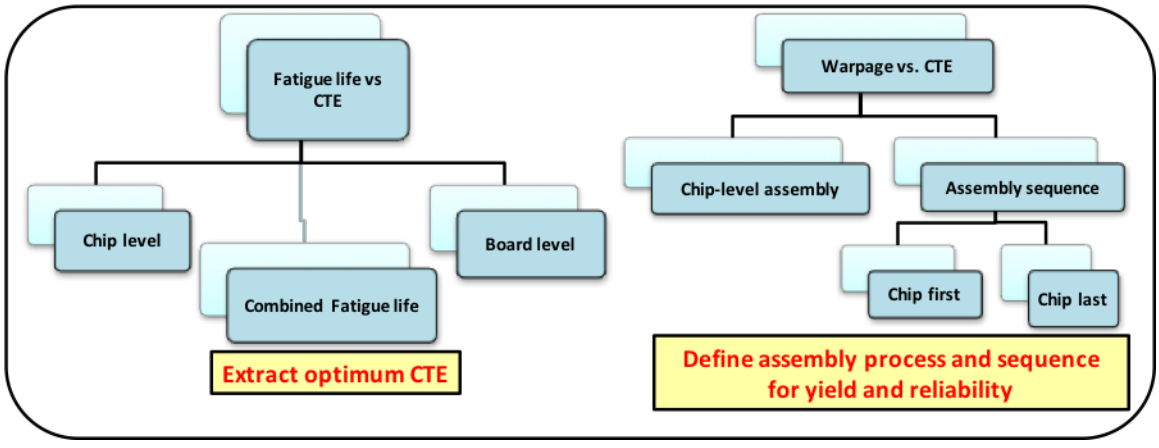


Figure 3.4: Assembly design methodology for balanced reliability of 2.5D packages.

3.3 Finite-element modeling for system-level reliability and minimum warpage of a 2.5D glass BGA package

3.3.1 Geometric model

A basic 2-dimensional geometry for a 2.5D glass package based on the test vehicle design was built with chip- and- board-level assemblies. The thickness of dies, substrate and interconnections was kept the same as that given in the TV design. The board was assumed to be 5mm-thick based on typical PCB thicknesses used in 2.5D packages [43]. The design of the package in section 3.1 consists of a large, center die and four, smaller memory (HBM) dies on the periphery. Since the basic model is built to only validate the methodology for assembly process design, the 2.5D structure was modified to a simple 2D package. Two models were built: one with the large center die and the other with an approximated size considering the size effect of the smaller dies. The rationale behind the first model is that the center die is larger and will have the highest impact on chip-level reliability; it is therefore sufficient to consider only the center GPU die to get an approximate value of chip-level fatigue life. The second model focuses on evaluating board-level reliability as well as substrate- and- board-side warpage. Since the four memory dies contribute to increasing the size of the package, it will have a significant impact on the warpage. Thus, it is combined with the center GPU die to form a larger chip with the same area to yield a more accurate value of substrate warpage. Additionally, in this case, the actual substrate size was considered to obtain precise results for board-level thermomechanical reliability.

Both the models were cut along the diagonal with symmetry boundary conditions at the center. The node at the left-bottom was fixed in the y-direction to prevent rigid body motion. The mesh was refined at the solder and thinner layers and coarser as we move away from the critical regions. An example geometry and mesh of the simplified model is as shown in Figure 3.5 and Figure 3.6 respectively.

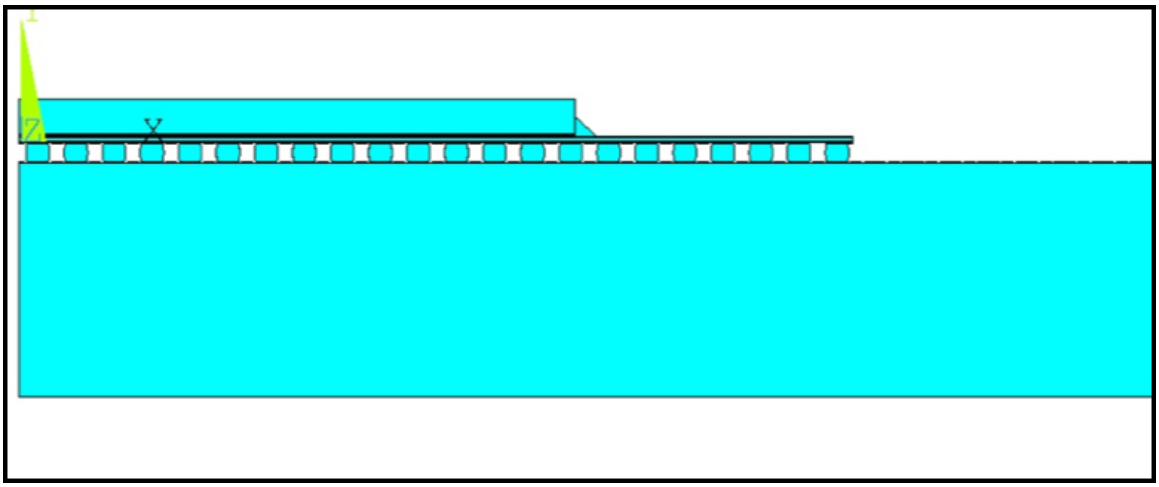


Figure 3.5: Example geometry in modeling.

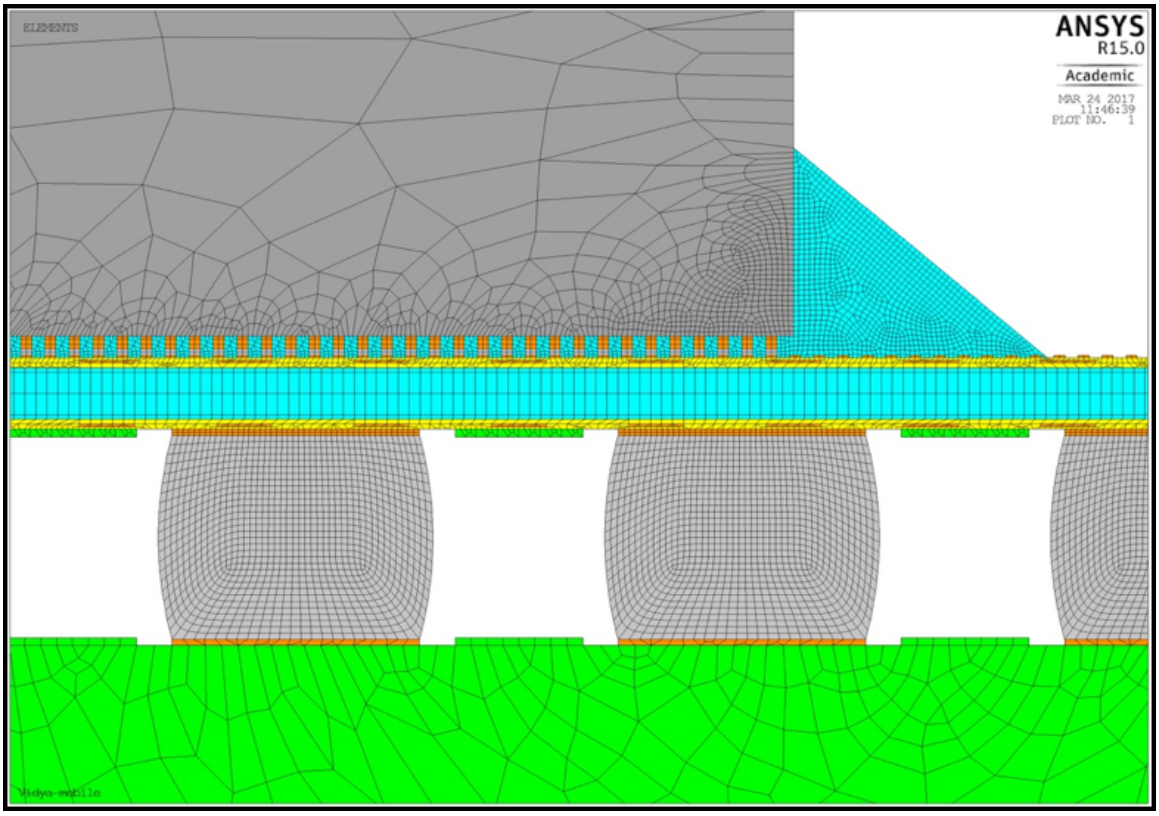


Figure 3.6: Example mesh in modeling refined at solder and thin layers.

3.3.2 Material modes and stress free temperatures

Definition of material models are critical in extract accurate data from the model. The fabrication occurs at different temperature ranges during chip- and- board-level assem-

Table 3.1: Physical properties of materials in modeling.

Material	Modulus (GPa)	CTE (ppm/°C)	Poission's Ratio
Silicon die	169	2.8	0.28
Low-CTE glass	77	3.8	0.22
High-CTE glass	74	9.8	0.23
Solder mask	13.45	4.5	0.30
Copper	121	17.3	0.30
FR-4 (PCB)	24	17	0.30

bly. During assembly, the solder melts and subsequently cools, establishing mechanical coupling during solidification to form interconnections. The model geometry consists of several different materials in the package. Each material has its own coefficient of thermal expansion (CTE). The mismatch in this CTE between the die and substrate at chip-level as well as substrate and board at board-level causes warpage and plastic strains in the solder joints upon temperature excursions. These plastics strains accumulate during thermal cycling and lead to crack initiation and propagation, resulting in fatigue failure. It is thus important to adequately represent materials to understand the mechanical behavior of the package. All materials except the metals were considered elastic and temperature independent. The physical properties of the materials are listed in Table 3.1. Copper was modeled with a bilinear elastic-plastic law with isotropic kinematic hardening (tangent modulus of 1034 MPa and yield stress of 172.4 MPa). The solder was considered viscoplastic. Anand's model capturing time-independent plasticity and solder creep was used for modeling of the SAC105 solder. Anand's model parameters for SAC105 are shown in Table 3.2.

Stress-free temperatures are assigned to the materials to mimic the fabrication and assembly processes. Stress-free temperature is defined as the temperature at which the components of an assembly do not exhibit deformations from their pre-assembled shapes [44]. Typically, in SMT assemblies, the stress-free temperatures are assumed to be close to the freeing point of the solder. This is because the coupling between the package and board is initiated when the solder starts solidifying. However, this does not apply in the presence of underfills or other adhesives. Underfills are epoxy-based polymers that are characterized

Table 3.2: Anand's model parameters for SAC105.

Anand's Parameter	**Units**	**Value**
A	1/s	5200
Q/R	K	10150
ξ	-	6
m	-	0.18
s	MPa	30
N	-	0.008
h_0	MPa	34000
A	-	1.62
s_0	MPa	23

by their glass transition temperatures (Tg). Tg is defined as the temperature at which the polymer changes from hard, glassy state to a soft, rubbery state. In assemblies that involve underfills, the warpage behavior is predominantly affected by the forces applied by the underfill material. The Tg of the underfill is therefore a better representation of the stress-free temperature [45, 46].

3.3.3 Thermal loading conditions for thermomechanical and warpage simulations

Coupled thermal-structural finite-element models were constructed to predict the fatigue life of the solder joints and provide assembly process guidelines for balanced reliability. The modeled structure was initially subjected to a temperature drop from 260 °C to 25 °C, simulating the cooling of SMT reflow process. JEDEC standards (JESD22-A106B) were followed that define five thermal cycles between -40 °C and 125 °C for fatigue life evaluation. The ramp-up and ramp-down temperatures as well as dwell time at the temperature were fifteen minutes each emulating real-time thermal cycling. The model was solved and the equivalent plastic strain range in the corner joint was extracted after cycling and used as damage metrics to predict the fatigue life of the interconnection for chip- and board-level assemblies. Subsequently, the warpage was extracted by calculating the maximum displacement in the y-direction. The predicted values were extracted for a range of substrate CTEs to provide guidelines for balanced reliability.

3.3.4 Effect of coefficient of thermal expansion of substrate on fatigue life of solders

As introduced in Chapter 1, the substrate CTE is known to have a major impact on board-level reliability. Finite element modeling was used to investigate the effect of substrate CTE on fatigue life and warpage to prevent yield loss and enhance board-level reliability. Fatigue models can be categorized based on damage metrics that include: 1) stress-based, 2) plastic strain-based, 3) creep based, 4) energy-based, and 5) damage accumulation based. Generally, fatigue experienced through CTE mismatch in the solder joints is related to plastic strain from the models. Several models like Coffin-Manson, Engelmaier-Wild, Solomon etc. have been proposed to predict the fatigue life of the package based on the plastic strain experienced by the solder during thermal cycling. The plastic strain range was extracted from the outermost solder joint at chip- and- board-level. In this work, the fatigue life of the solder joints was estimated using the conventional Coffin-Manson equation [47] as well as Engelmaier-Wild fatigue model. The Coffin-Manson equation below is an empirical fit to determine the fatigue life below 10,000 thermal cycles.

$$N_f = (\frac{2 * \Theta}{\Delta\gamma_p})^{(\frac{1}{a})} \tag{3.1}$$

where N_f is the number of cycles to failure, a is the fatigue strength exponent that is assumed to be 0.5413, is the fatigue ductility coefficient assumed to be 0.2516 and is the plastic strain range, obtained using the NL, EPEQ' command over the fifth thermal cycle. Generally, thermomechanical models used to estimate the fatigue life of the solder joints is related to the calculation of inelastic strain energy density per cycle. The inelastic strain energy density is the average plastic work in the volume of solder joints. The fatigue life estimation based on Darveaux's theory should be meshed to have two to three layers of elements. Due to the difference in volume between the BGA at board-level and a solder bump with Cu pillar at chip-level, the plastic strain energy calculation needs to be modified. Therefore, the plastic strain of the solder bump at chip-level is multiplied by a size factor

Table 3.3: Maximum plastic strain range values at chip- and- board-level solder joints

Package CTE (ppm/°C)	**Plastic strain range**	
	Chip-level	**Board-level**
3.3	4.27 x 10^{-3}	1.30 x 10^{-2}
5	4.55 x 10^{-3}	1.16 x 10^{-2}
6	4.96 x 10^{-3}	1.10 x 10^{-2}
7	4.90 x 10^{-3}	1.04 x 10^{-2}
8.3	5.15 x 10^{-3}	9.78 x 10^{-3}
9.8	6.52 x 10^{-3}	9.14 x 10^{-3}

to match the volume of the BGA strain. This size factor is the ratio between the applied height of the solder bump and the height of the BGA used in the model. The plastic strain values for chip- and- board-level solder joints for the range of substrate CTEs (3.3 ppm/°C 9.8 ppm/°C) are given in Table 3.3.

The Engelmaier model is a modification of the Coffin-Manson model and is typically fit for board-level fatigue life calculations that have larger solder volumes. The Engelmaier model takes into account the cyclic frequency and temperature effects. In this work, the Engelmaier-Wild model was used to calculate the board-level BGA fatigue life [48, 49]. The equation used to predict the number of cycles to failure is given below:

$$N_f = 0.5 \times (\frac{2\Delta\epsilon_f}{\Delta\epsilon_p})^{(\frac{1}{c})} \tag{3.2}$$

$$\frac{1}{c} = c_0 + c_1 T_{SJ} + c_2 \times ln(1 + \frac{t_0}{t_d}) \tag{3.3}$$

where $\Delta\epsilon_f$ is the fatigue ductility coefficient, $\Delta\epsilon_p$ is the strain range amplitude, c is the fatigue strength exponent, c_0,c_1,c_2, t_0 are solder-specific constants. T_{SJ} is the mean cyclic thermal solder-joint temperature (42.5 °C), and t_d is the half-cycle dwell time in minutes. These material constants are reported in literature as given in Table 3.4 [50].

From these constants, the value of was calculated to be 0.4516. The plastic strain distribution of the outermost solder joints in the Cu pillar as well as BGA interconnections

Table 3.4: Material constants in the Engelmaier-Wild model

Solder	ϵ_f	c_0	c_1	c_2	t_0
SAC105	0.225	0.48	9.30E-04	-1.92E-04	500

Table 3.5: Predicted number of cycles to thermomechanical failure

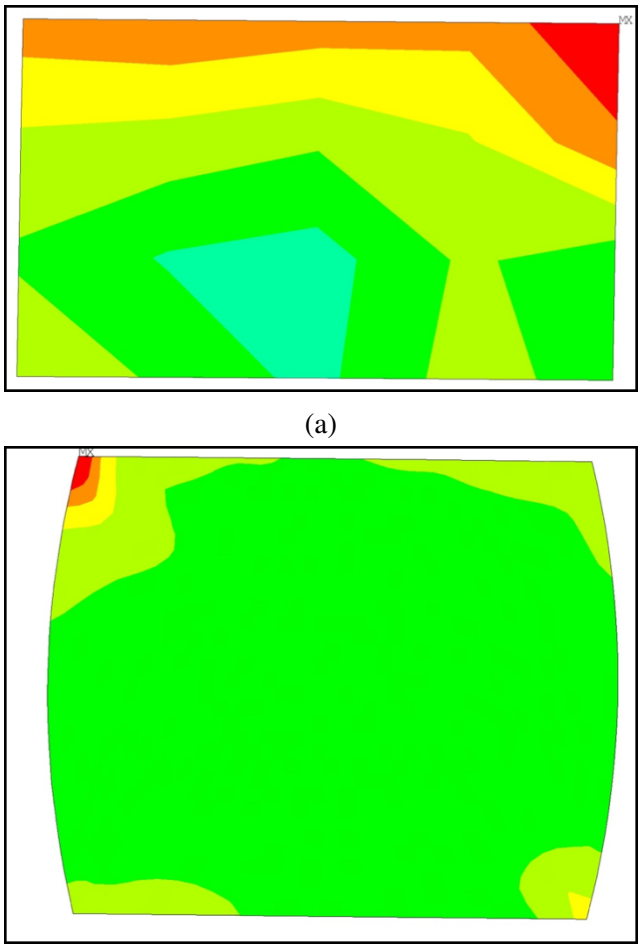
Package CTE (ppm/°C)	**Chip-level Fatigue Life**	**Board-level Fatigue Life**	
	Coffin-Manson	**Coffin-Manson**	**Engelmaier-Wild**
3.3	6702	850	1267
5	5962	1045	1624
6	5283	1164	1848
7	4865	1287	2084
8.3	4232	1450	2403
9.8	3063	1644	2794

are shown in Figure 3.7 and the number of cycles to failure are tabulated in Table 3.5.

JEDEC reliability standards require the package to pass a minimum of 1000 cycles. Model predictions show that the chip-level fatigue life exceeds the JEDEC standards by a large margin, regardless of substrate CTE. The low-CTE packages at board-level do not pass these standards according to Coffin-Manson predictions and barely pass a 1000 cycles with the Engelmaier-Wild predictions. As expected, the board-level fatigue life increases with increase in CTE. There is no cross-over between the chip- and board-level reliability as a function of substrate CTE. The plot of fatigue life as against CTE is shown in Figure 3.8. Based on the model predictions, it can be seen that moving towards a higher CTE is a solution to enhance board-level reliability without adversely affecting chip-level reliability. However, the large CTE mismatch induced at chip-level can lead to warpage and yield failures that need to be accounted for. If low-CTE substrates are still required for the package, compliant or alternative interconnection materials need to be developed. The plot in Figure 3.8 can thus be used to provide conclusions on balancing system-level reliability.

3.3.5 Effect of coefficient of thermal expansion of substrate on warpage behavior

The fatigue life model predictions indicate the need to move towards higher CTE for larger package sizes. However, higher CTEs induce a large CTE mismatch between the silicon



(a)

(b)

Figure 3.7: Plastic strain distribution in the outermost solder joint at (a) chip-level and (b) board-level.

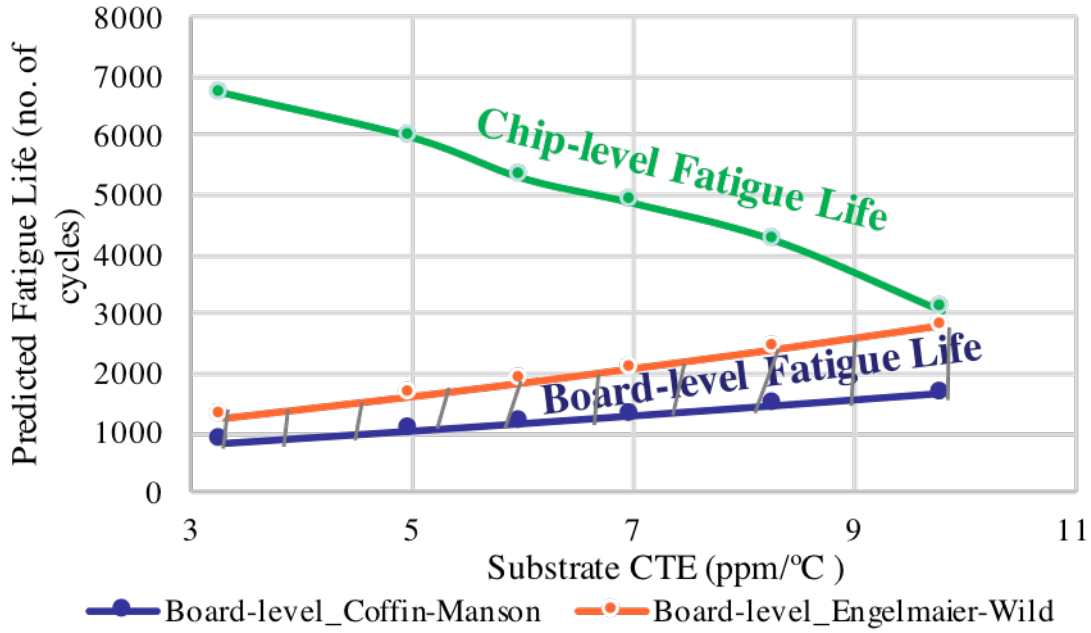


Figure 3.8: Model-predicted chip- and- board-level fatigue life for a 2.5D glass BGA package.

die and substrate. This CTE mismatch results in package warpage, causing open joints and strain in the solders. To understand the effect of substrate CTE on warpage, the warpage response of the package after chip-level assembly as well as consequent board-level assembly was captured. The package warpage convention is given in the Figure 3.9. When the substrate shrinks more than the die at lower temperatures, it results in a dome-shaped' or convex warpage with a positive convention. At higher temperatures, the substrate expands and acquires a bowl-shaped' or concave warpage with a negative convention.

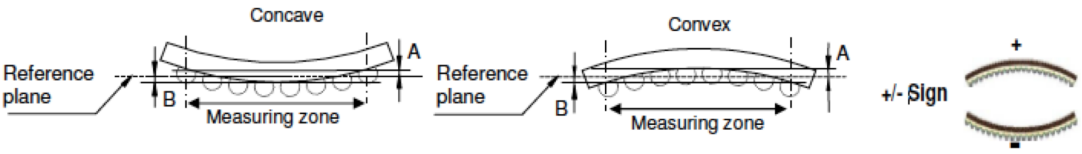


Figure 3.9: JEDEC-defined package warpage convention.

The model was built to mimic reflow conditions where the package was heated till 260 °C and then cooled to room temperature. While cooling from the stress-free temperatures,

the solder starts constraining after which the package starts shrinking. This thermal gradient and CTE mismatch results in package warpage which has an adverse effect on the subsequent board-level assembly. An example of the warpage captured at room temperature is shown in Figure 3.10. Results of the net warpage at room temperature is plotted against different substrate CTEs (Figure 3.11).

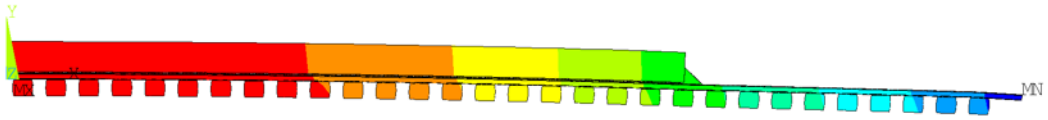
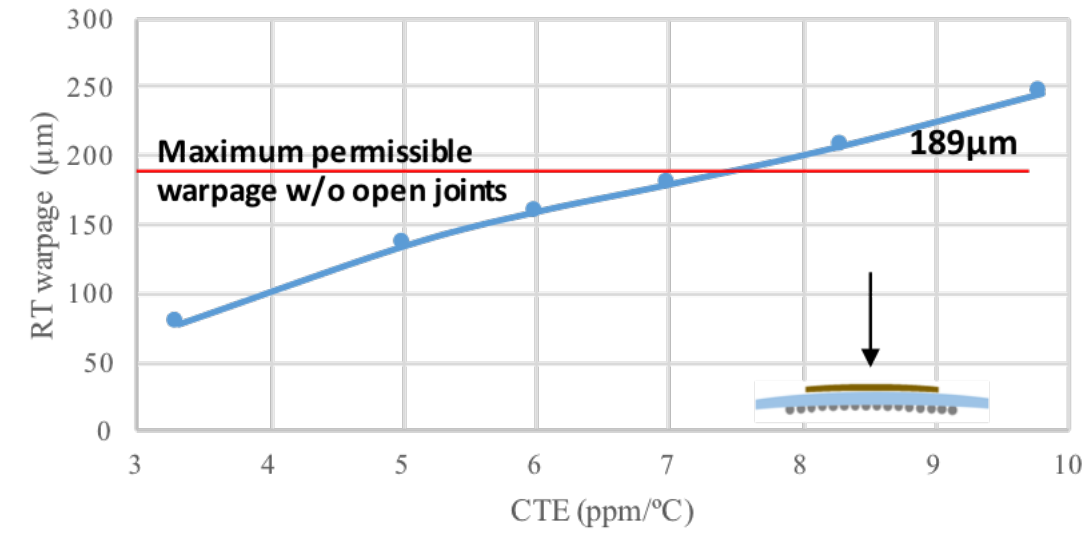


Figure 3.10: Example warpage at room temperature after chip-level assembly.

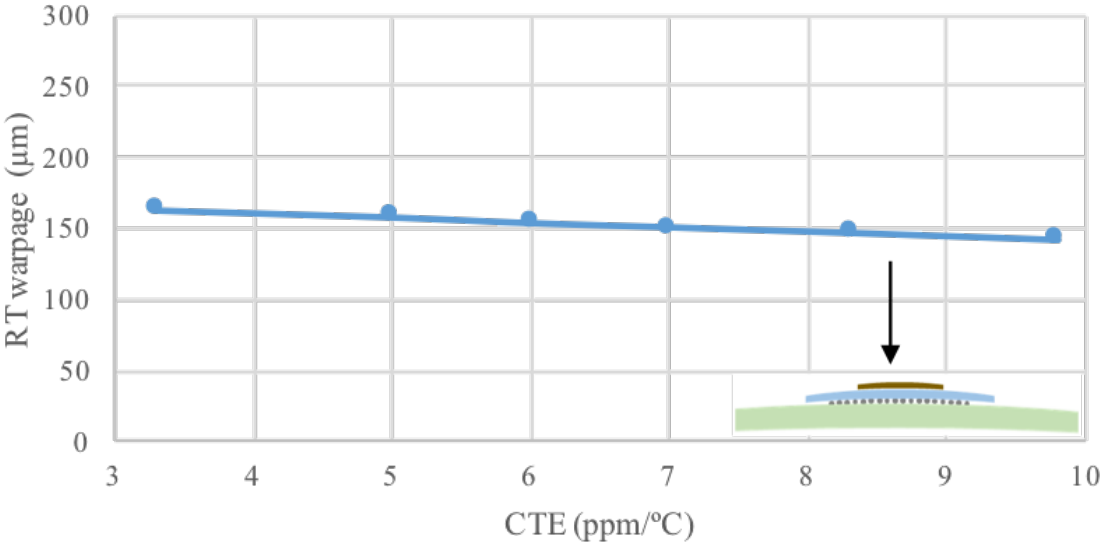
The room temperature warpage after chip-level assembly increases with increase in substrate CTE, as expected. After package-on-board assembly, the low-CTE substrates are predicted to have highest warpage. The maximum permissible warpage to prevent open solder joints for the given 2.5D package is 189 μm. This value was calculated based on JEITA-EDR standards [51]. Beyond a substrate CTE of 6 ppm/°C, the plot predicts yield loss after chip-level assembly suggesting that warpage is a major constraint in terms of thermomechanical reliability while considering higher CTE packages. Thus, an optimum CTE range balancing warpage as well as reliability needs to be defined. In this case, the range for obtaining highest reliability with least warpage is approximately between 4 ppm/°C 7.5 ppm/°C. If the package requires low-CTE substrates, in addition to using advanced interconnection materials, there is a need to minimize net warpage during TCB or alter the sequence of assembly.

3.4 Summary

As described in this chapter, higher CTE substrates are required to achieve superior system-level reliability. However, warpage contributes as a major factor in degrading yield and reli-



(a)



(b)

Figure 3.11: Room temperature warpage predictions after (a) chip-level assembly and (b) after package-to-board assembly.

ability at high-CTEs. Model predictions indicate that defining a range of CTE is necessary to balance warpage and reliability at chip- and- board-level. Since glass can be fabricated at different CTEs with the same modulus, it can be optimized and is the perfect solution to achieve system-level reliability with a 2-level hierarchy. Since board-level reliability is predicted to be more critical and sensitive to the substrate CTE, advanced materials that can be used to enhance its reliability are investigated in the following chapter. A single-chip glass package is considered for this investigation since it includes detailed experimental results to validate the methodology described for 2.5D packages.

CHAPTER 4

MODELING, DESIGN AND DEMONSTRATION OF BOARD-LEVEL RELIABILITY FOR STRAIN RELIEF AND RELIABILITY

This chapter reports details of the test vehicle design and fabrication, and results of finite element modeling, using ANSYS™ 15.0, of a single-chip glass BGA package with direct SMT-on-board. Thermomechanical reliability and room temperature warpage is evaluated over a range of substrate CTEs (3.3 ppm/°C to 9.8 ppm/°C). Strain relief mechanisms like polymer collars and doped solders employed to enhance board-level reliability are characterized through extensive failure analysis. Finally, a correlation between the model and experiments is performed to validate the design methodology for a 2.5D package as explained in Chapter 3.

4.1 Test vehicle design and fabrication of a single-chip glass BGA package

4.1.1 Daisy chain test die

Daisy-chain test vehicles (Figure 4.1), 10 mm x 10 mm in size, with 5448 I/Os arranged in four staggered peripheral rows at 80/40 μm pitch and a central area array at 150 μm pitch was used in this study. The silicon test wafers were 300 mm in size and were fabricated by Advanced Semiconductor Engineering Inc. (ASE). The wafers were plated with Cu in a dogbone wiring structure and bumped with standard Cu pillars. The copper pillar interconnections were 28 μm in diameter with 17 μm copper height, a Ni barrier layer of 3 μm and a SnAg solder cap, 17 μm in height. Test dies, 100m and 200 μm in thickness were used to assess the effect of die thickness.

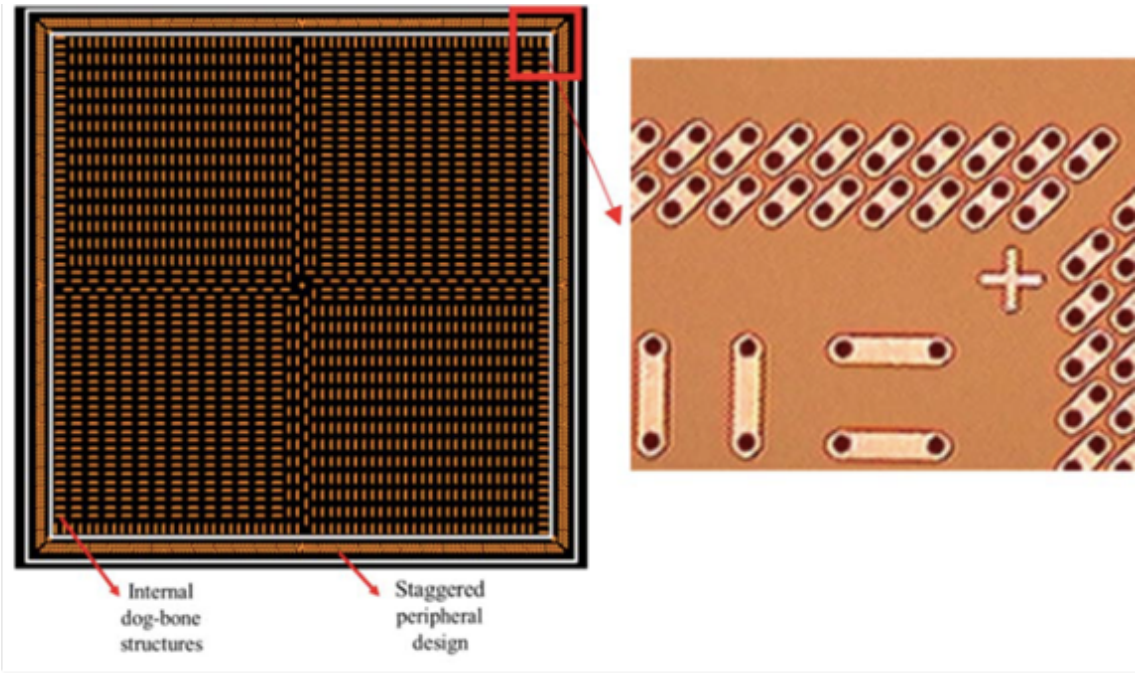


Figure 4.1: Daisy-chain test die: a) design and b) optical image of a bumped corner (Image courtesy of ASE).

4.1.2 Glass substrate fabrication

Glass substrates, 18.5 mm x 18.5 mm in size, with the die footprint on the top side, and a 45 x 45 area array of daisy chain BGA interconnections on the bottom side is shown in Figure 4.2. Due to the absence of through-package-vias (TPVs), the chip- and- board-level daisy chains are not connected to each other. The design consists of four metal layers, with a dummy mesh pattern in the inner layers for adequate Cu coverage and realistic warpage representation. The substrate was fabricated using a semi-additive process for low- and- high-CTE glass on 6 inch x 6 inch panels. The glass core, 100 μm in thickness, was laminated with 17.5 μm-thick dielectric layers. The Cu dog-bone structures were defined through double-side lithography processes using dry-film photoresist, 10-12 μm thick formed by electrodeposition. The photoresist was then stripped and the electroless copper seed layer was etched. The double-side process was repeated again for lamination of dielectric build-up layers. The daisy-chain patterns on the die and BGA sides were si-

multaneously formed by double-side Cu electrolytic plating. Further, a solder-mask defined (SMD) dry film resist (15 μm) by Hitachi Chemical was applied as a passivation layer to define landing and probing pads. Surface finish of electroless palladium autocatalytic gold (EPAG) was plated on the Cu pads by Atotech GmbH with a production-controlled process. The cross-sectional expanded view of the substrate is shown in Figure 4.3. The substrate stack-up summary is recapped in Table 4.1.

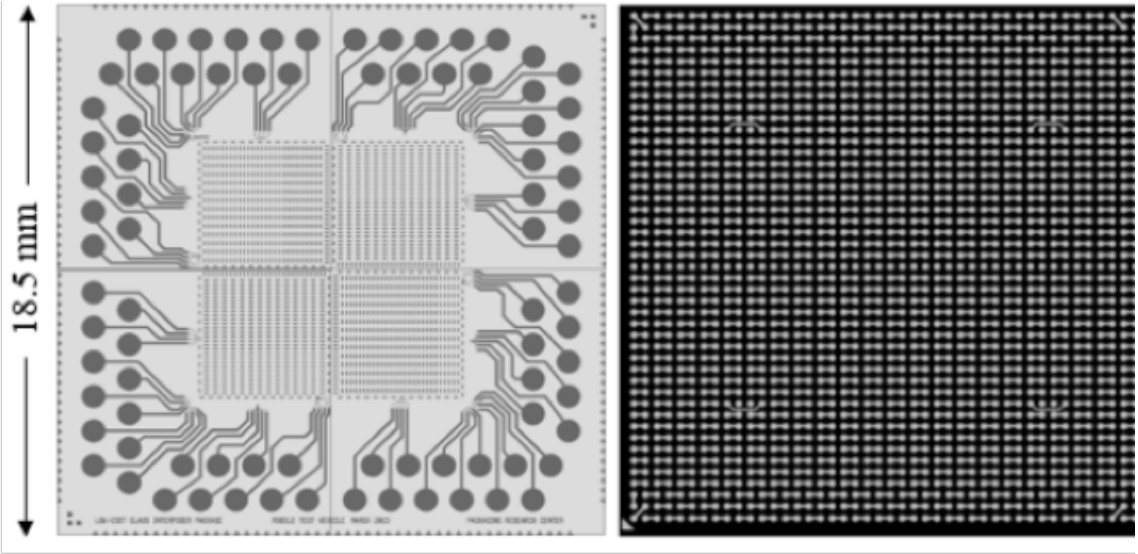


Figure 4.2: Glass substrate test vehicle design on (left) the die side, and (right) the BGA side.

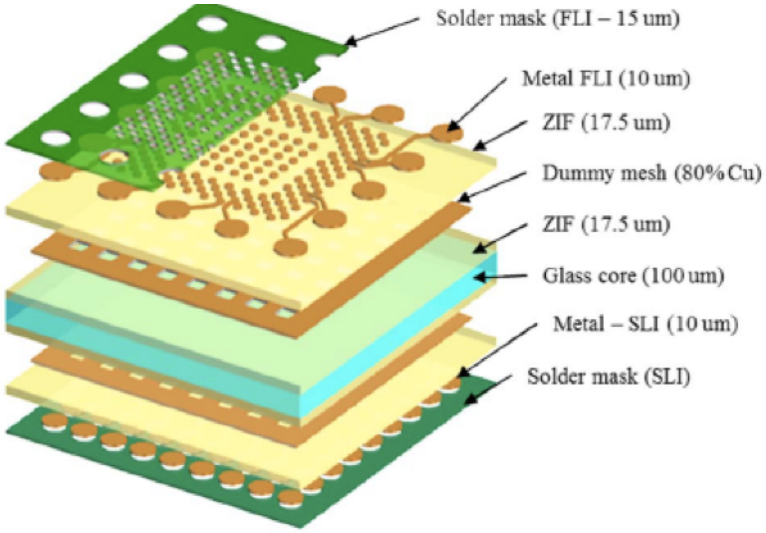


Figure 4.3: Schematic of 18.5 mm x 18.5 mm 4-metal layer glass substrate.

Table 4.1: Summary of stack-up materials and design rules

Parameters	Description
Substrate core	Low- & high-CTE glass
Build-up layers	17.5 mm/17.5 mm
Solder Resist	15 mm - SMD
Metal layers	4 ML with 80% Cu coverage on inner dummy
Surface finish	EPAG (Atotech, Germany)
BGA	250 mm @ 400 mm pitch (Paste printed)
Solder	SACm™ , SAC305, SAC105 (Indium)
Die thickness	100 mm, 200 mm
Underfill	50 mm

4.1.3 PCB board design

The backside of the glass substrates consists of 2025 solder balls at 400 μm pitch, divided in a network of 52 daisy chains. A single-layer PCB was designed to match the daisy-chain pattern, including four corner circuits and 48 inner chains as seen in Figure 4.4. The copper pads were non-solder mask defined (NSMD) with a surface finish of electroless nickel electroless palladium immersion gold (ENEPIG).

4.2 Finite-element modeling for analysis of system-level reliability and room temperature warpage

4.2.1 Geometric model

A 2D model of a single-chip glass BGA package, 18.5 mm x 18.5 mm in body size with 100 μm-thick glass was created with chip- and- board-level assemblies. The geometry includes the silicon die, 100 μm in thickness, copper pillar interconnections, polymer dielectric layers, glass substrate, copper redistribution layers with SAC105 solder on the PCB side (Figure 4.5). A 2.5D or 3D model would be ideal for accurate results. However, 2D models with plane strain approximations are sufficient for this parametric study of comparing the thermal cycling reliability and warpage for different glass CTEs. The die-substrate-PCB assembly represents a cut along the diagonal with symmetry boundary conditions at the

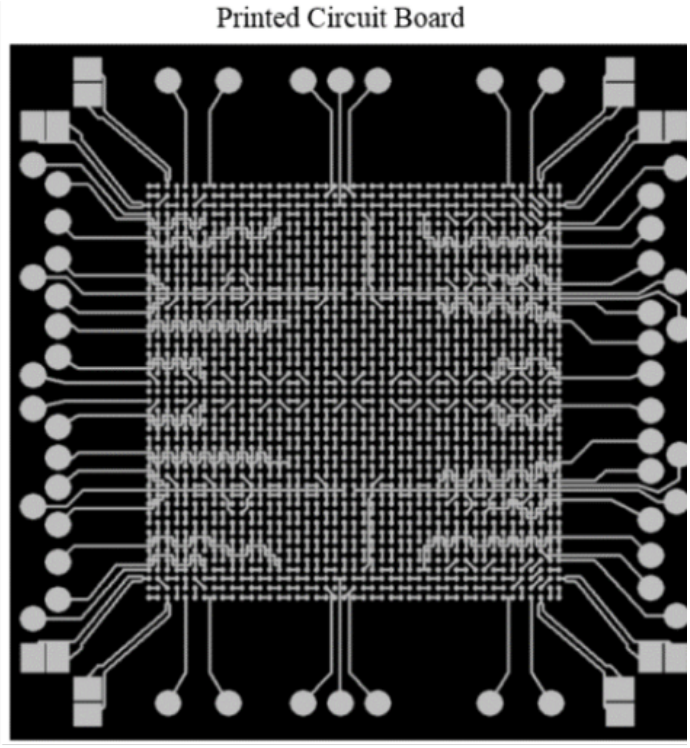


Figure 4.4: Daisy-chain interconnection design of PCB with probing pads.

center. The node at the left bottom was fixed in the y direction to prevent rigid body motion. The mesh was refined at the solder and thin layers. The mesh was coarser away from the critical regions and the number of elements as compared to the thickness was built to effectively evaluate fatigue life and warpage. An example model mesh is as shown in Figure 4.6.

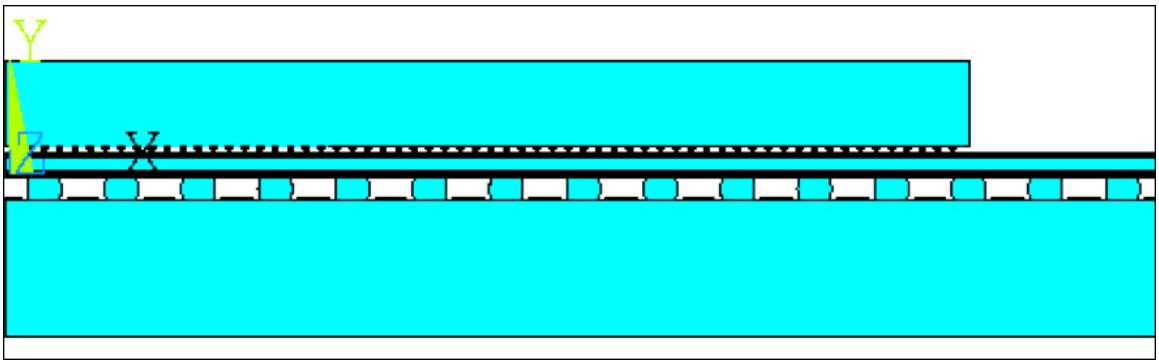


Figure 4.5: Example geometry in modeling. Two-dimensional geometry represents a cut along the diagonal of the package with symmetry at the left.

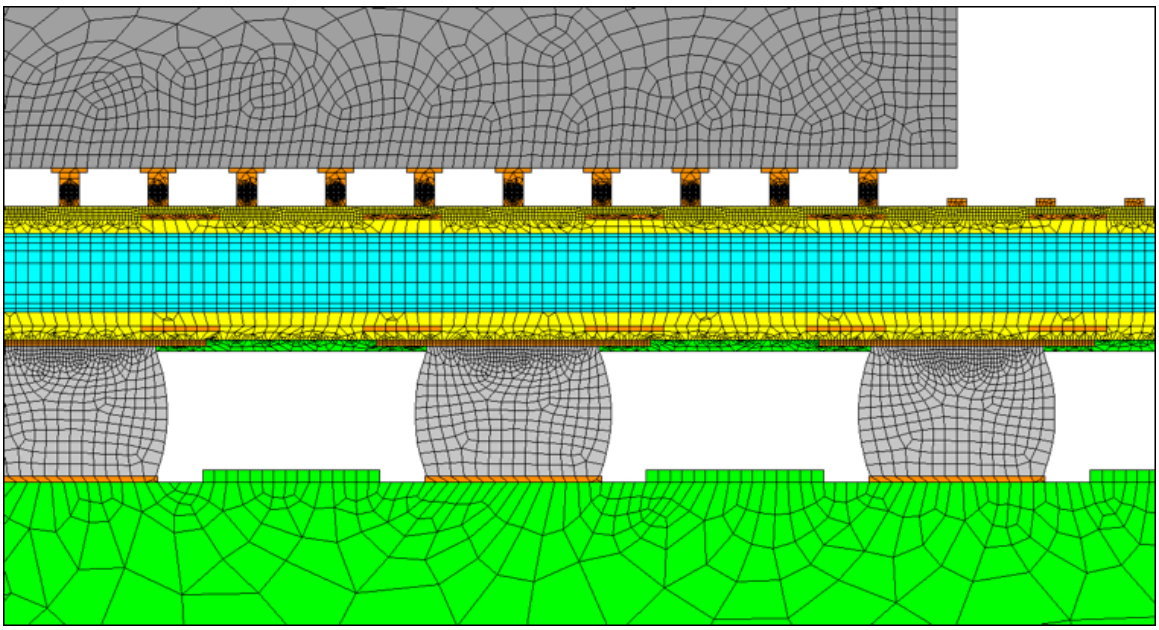


Figure 4.6: Example mesh used in modeling for thermal and mechanical analysis.

Material models, stress-free temperatures and thermal loading conditions:

Defining material models are critical to achieve accurate results as explained in the previous chapter. Since the materials used for the single-chip package are the same as that in Chapter 3, the same material models are employed. The stress-free temperature is set as the Tg of the underfill which is known to dominate the warpage behavior [45, 46]. Coupled thermal-structural finite-element models were constructed for the single-chip package as well to determine the fatigue life of the solder joints and warpage of the package. The thermal loading conditions are as defined by JEDEC standards for thermomechanical reliability from -40 °C and 125 °C with ramp-up, ramp-down and dwell times of 15 minutes each emulating real-time thermal cycling.

Evaluation of thermal cycling reliability:

The Coffin-Manson and Engelmaier-Wild model equations as described in Chapter 3 section 3.3.4 are used to estimate the chip- and- board-level fatigue life of the single-chip package. The plastic strain values for chip- and- board-level solder joints extracted from

Table 4.2: Maximum plastic strain range values at chip- and- board-level solder joint

Package CTE (ppm/°C)	**Plastic strain range**	
	Chip-level	**Board-level**
3.3	3.92 x 10^{-3}	1.081 x 10^{-2}
9.8	4.613 x 10^{-3}	7.88 x 10^{-3}

the model for low- and- high-CTE substrates are calculated and given in Table 4.2.

The plastic strain distribution of the outermost solder joints in the Cu pillar as well as BGA interconnections are shown in Figure 4.7 and Figure 4.8 and the number of cycles to failure are tabulated in Table 4.3.

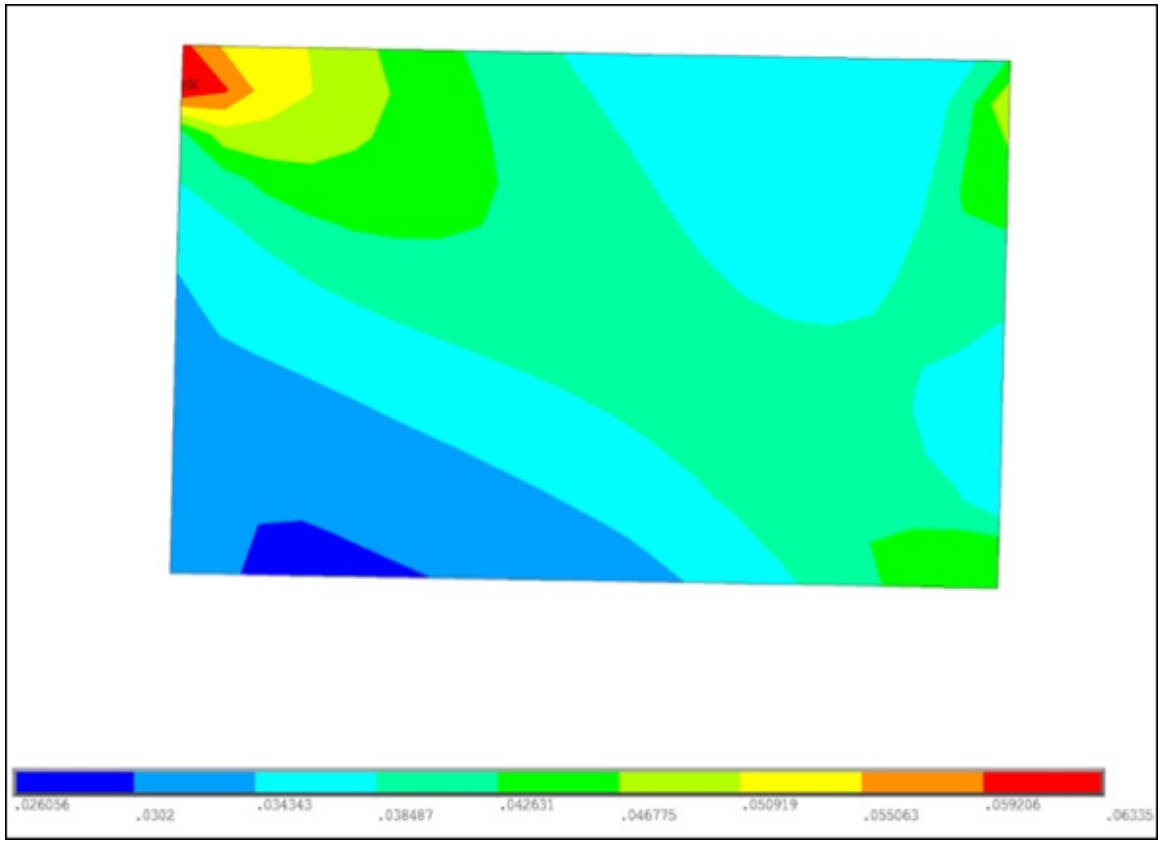


Figure 4.7: Plastic strain distribution of outermost solder joint at chip-level.

Based on the fatigue life models, all package configurations would survive over a 1000 cycles regardless of glass CTE, satisfying JEDEC reliability standards. Model predictions depicted that the chip-level fatigue life exceeded 1000 cycles by a large margin. There was no cross-over between the chip- and- board- level reliability as a function of substrate CTE.

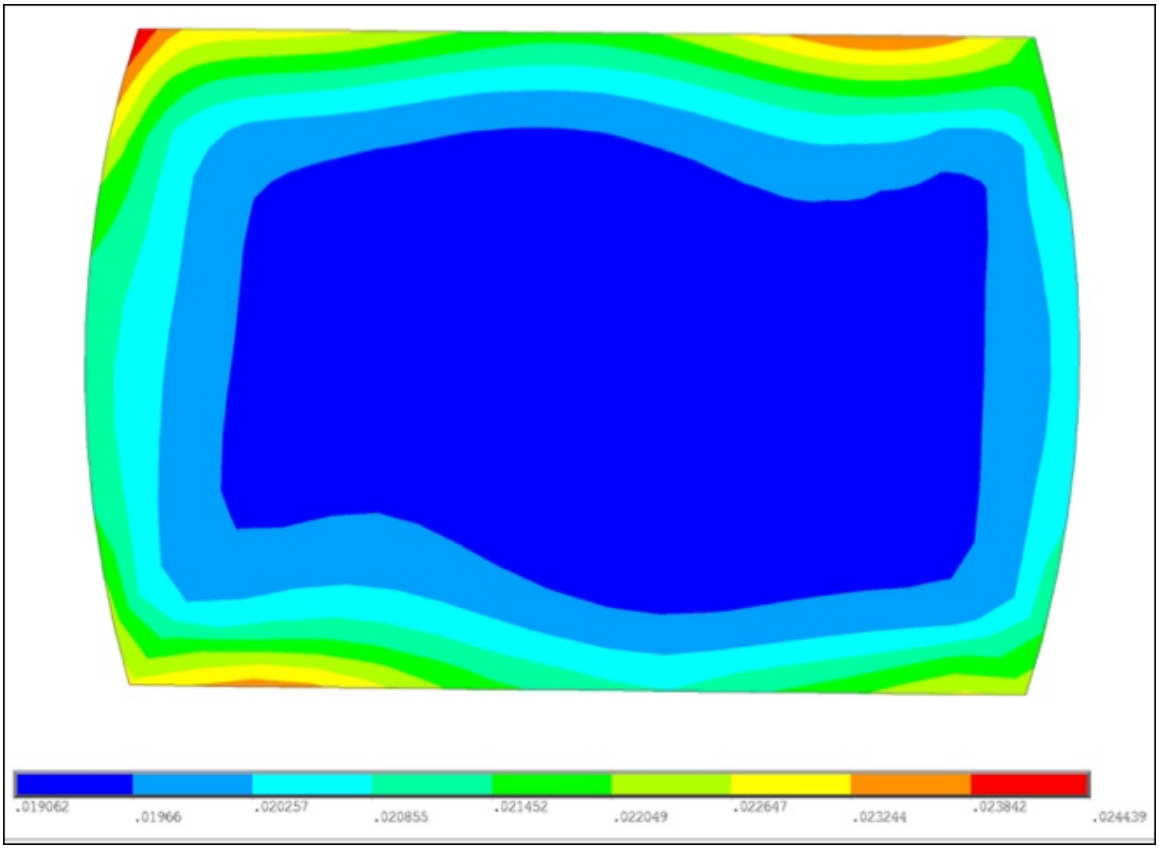


Figure 4.8: Plastic strain distribution of outermost solder joint at board-level.

As expected, the board-level fatigue life of high-CTE substrates was predicted to be better than that of low-CTE substrates. However, the low-CTE glass packages at board-level barely passed 1000 cycles. Therefore, board-level reliability is a major issue and needs to be addressed. It can be seen from the predictions that the substrate CTE has a larger impact on board-level reliability as compared to that of chip-level. Since only glass with low (3.3 ppm/K) and high (9.8 ppm/K) CTEs is readily available today, advanced interconnection materials such as doped solders and polymer collars were introduced to further enhance

Table 4.3: Predicted fatigue life for chip- and- board-level assembly using Coffin-Manson and Engelmaier-Wild models

	Chip-level Fatigue Life	**Board-level Fatigue Life**	
	Coffin-Manson	**Coffin-Manson**	**Engelmaier-Wild**
Low-CTE	7856	1205	1926
High-CTE	5815	2158	3871

board-level reliability with minimum system-level impact. Moving to 2.5D architectures, the chip- and board-level reliabilities of the system as predicted in Chapter 3 is shown to deteriorate further. This is because even though the board-level interconnection pitch is targeted to be coarser for 2.5D packages, the increase in package size and decrease in chip-level interconnection pitch will adversely impact the board-level fatigue life. Given that, low-CTE substrate packages may not pass JEDEC reliability standards at board-level. In order to balance the reliability, the optimum glass CTE can be extracted from the model similar to the plot in Figure 4.9.

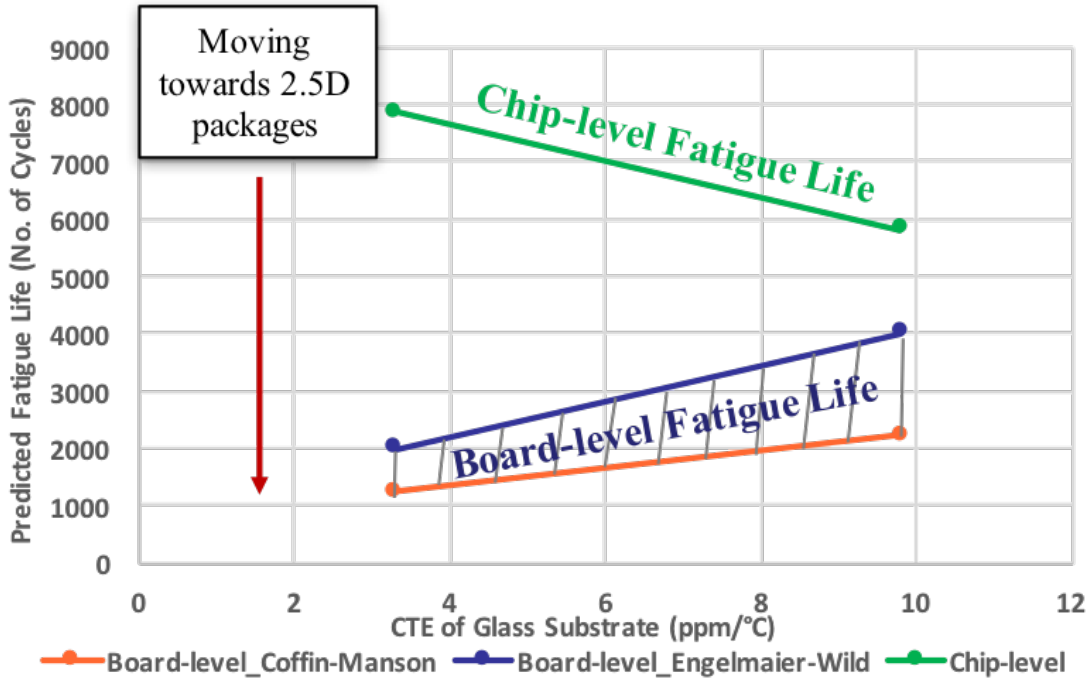


Figure 4.9: Fatigue life vs. CTE using Coffin-Manson and Engelmaier-Wild models.

Based on modeling predictions, conclusions can be made on the need to alter the assembly process and sequence by moving towards higher CTEs or the necessity of advancing to compliant interconnections.

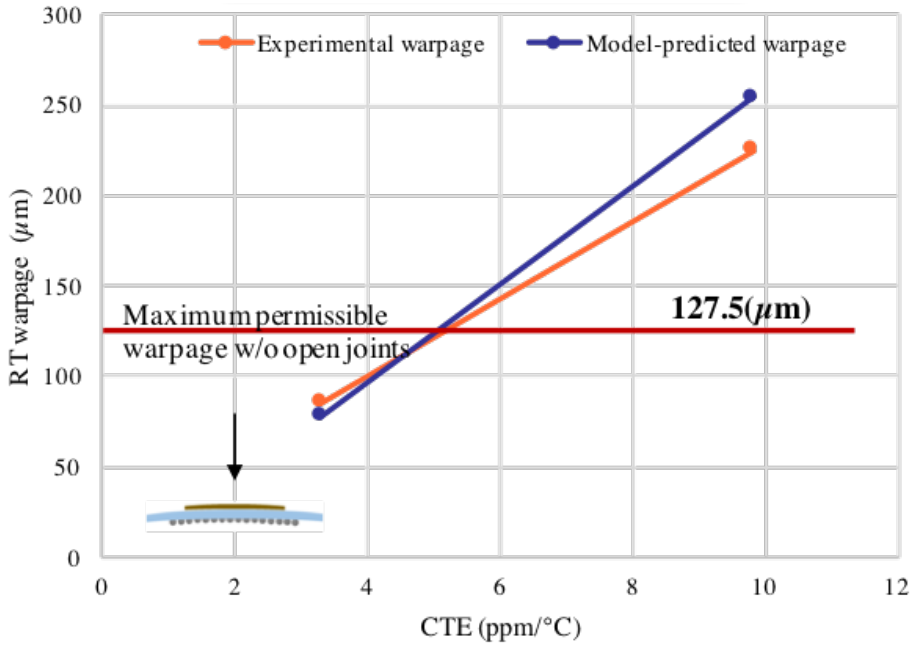
Analysis of assembly warpage yield:

Warpage depends on the CTE mismatch and thermal gradient applied during assembly. To evaluate the effect of substrate CTE on assembly warpage, the model was built to replicate the reflow conditions in which the package was heated to 260 °C and then cooled to room temperature. The plot below (Figure 4.10) shows the room temperature warpage as a function of substrate CTE.

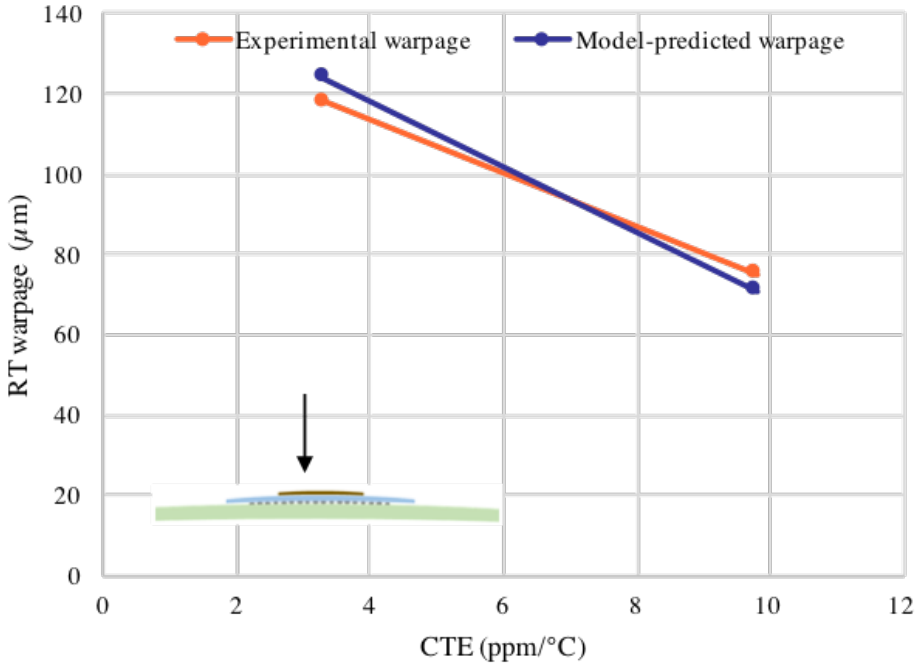
As expected, the room temperature warpage after chip-level assembly for the low-CTE package was estimated to be the least. Maximum warpage was predicted for the low-CTE package after package-on-board assembly. These predictions correlate well with the experimental warpage results obtained by Bhupender et al. [42]. Higher warpage at chip-level can result in yield loss during board-level assembly. For a given package size, pitch and BGA height, the maximum permissible warpage to prevent open solder joints prior to board-level assembly is given by JEITA-EDR specifications. As observed from the plot, this maximum permissible warpage for the package with 18.5 mm x 18.5 mm in body size, 400 μm BGA pitch and 200 μm BGA height is 127.5 μm with a corresponding CTE of 5 ppm/°C. In order to move towards higher CTEs for better reliability, the assembly process and sequence needs to be optimized.

4.3 Assembly and yield evalutation

Test dies, 100 μm and 200 μm in thickness were assembled by dip-flux thermocompression bonding process using a semi-automatic Finetech Fineplacer Matrix flip-chip bonder with a placement accuracy of 3 μm. The stage temperature was at a constant of 100 °C, the tool head peak temperature on the die side was 360 °C with a heating rate of 6 K/s and a pressure of 0.9 MPa was applied throughout the process. The bonded assemblies were then underfilled with the Namics Corporation 8410-219 material through dot-dispensing process manually. The underfill was then cured at 165 °C for two hours. After chip-level assembly,



(a)



(b)

Figure 4.10: Model vs. experimental correlation of room temperature warpage as a function of glass CTE after (a) chip-level assembly and (b) board-level assembly.

Table 4.4: Preliminary evaluation of number of samples in test for thermal cycling reliability.

Package Type	BGA Solder	# of Samples in Test
Low-CTE glass (with 200 μm-thick die)	SACm™	6
	SACm™ with collars	6
	SAC305	6
	SAC105	5
Low-CTE glass (with 100 μm-thick die)	SACm™	5
	SACm™ with collars	5
	SAC305	3
	SAC105	4
High-CTE glass (with 100 μm-thick die)	SACm™	4 (yield loss due to warpage)

BGA balling was done at panel-level with SAC305, SAC105 and SACm™ solder alloys using an optimized paste printing process. 9 samples with SACm™ solder alloy that were spin-coated with circumferential polymer collars were used for thermal cycling test in order to compare the effect of this strain relief mechanism on fatigue life. The number of samples in test for each configuration is listed in Table 4.4.

After laser dicing, the glass packages were then assembled on PCBs using a standard pick-and-place reflow process. This pick-and-place process was performed using the same Finetech Matrix FINEPLACER using no-clean tacky flux. The optimized reflow conditions using standard SMT process was optimized to minimize voiding. A summary of the assembly process is shown in Figure 4.11. Electrical measurement of the daisy-chain resistances was performed to evaluate yield. The overall balling yield was of 85%, with lower yield in high-CTE substrates due to increased warpage after chip-level assembly. The yield after SMT assembly was measured to be 90%. A total of 32 low-CTE and 4 high-CTE samples were chosen for thermal cycling. The number of high-CTE samples were limited due to yield loss. It is important to optimize the TCB conditions to prevent yield loss due to assembly warpage based on substrate CTE.

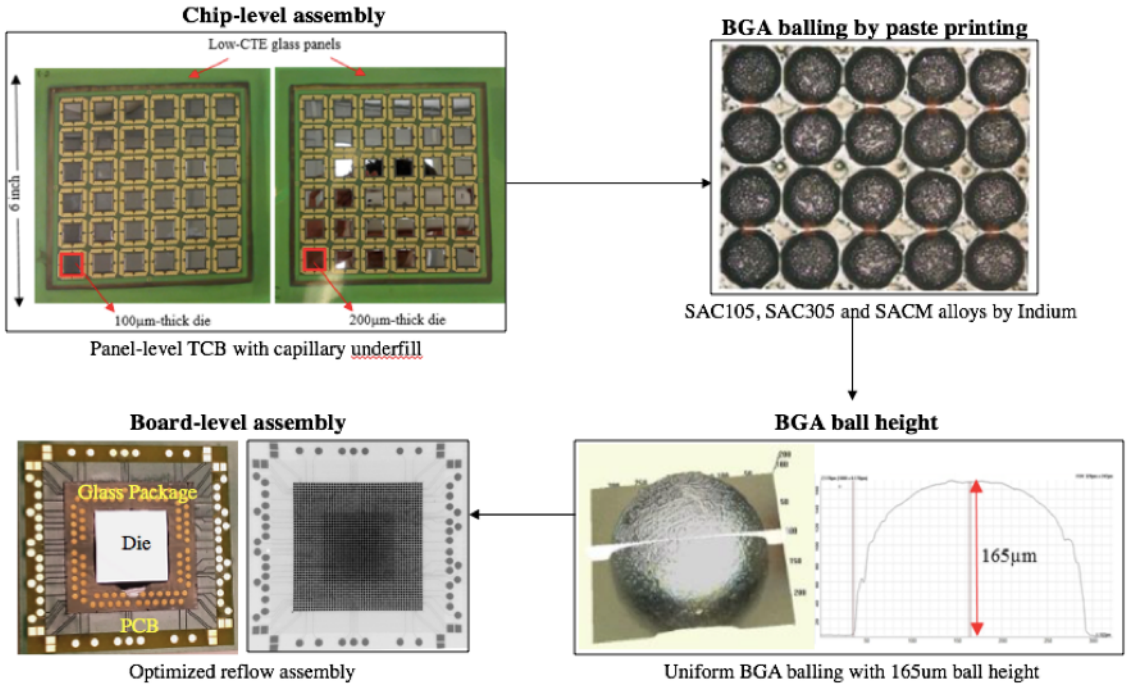


Figure 4.11: Summary of test vehicle assembly of low-CTE glass panels with 100 μm and 200 μm dies.

4.4 Thermal cycling reliability test

Thermal cycling was performed following the JEDEC JESD22-A104D standards that define five temperature steps between -40 °C and 125 °C with a dwell time of 15 minutes at each extreme, with one cycle an hour. The resistance of each test vehicle was monitored every 100 cycles. The failure criteria were defined as an increase in the resistance by 20% or an electrically open daisy-chain. All samples passed 1000 cycles with stable daisy-chain resistances, regardless of glass CTE and solder alloys. The low-CTE samples failed in the range of 1100-1800 cycles while the high-CTE cycles passed 2700 cycles and are still in test. These results correlate well with modeling predictions.

4.4.1 Failure distribution

Failed samples were characterized for void-detection using X-ray analysis. The sample failure distribution for all daisy-chain test vehicles are plotted in Figure 4.12 and Figure 4.13

for the 200 μm-thick and 100 μm-thick dies, respectively. Samples with excessive voiding failed earlier than expected. The average failure distribution for the 200 μm-thick die packages are as follows - 1350 cycles for SAC105, 1550 cycles for SAC305, 1450 cycles for SACm™ without polymer collars and 1600 cycles for SACmTM with polymer collars. The number of cycles to failure for the 100 μm-thick die packages were similar to that of 200 μm-thick dies, but with a broader failure distribution. The thermomechanical reliability results were as expected Soft solder SAC105 having the least fatigue life, hard solder SAC305 with superior TCT reliability and SACm™ exhibiting similar fatigue behavior as that of the higher Ag content solder, SAC305. Additionally, samples with polymer collars improved the fatigue life. A 25% increase in fatigue life can be achieved with advanced interconnection materials for a direct SMT assembly glass package with finer BGA pitches of 400 μm.

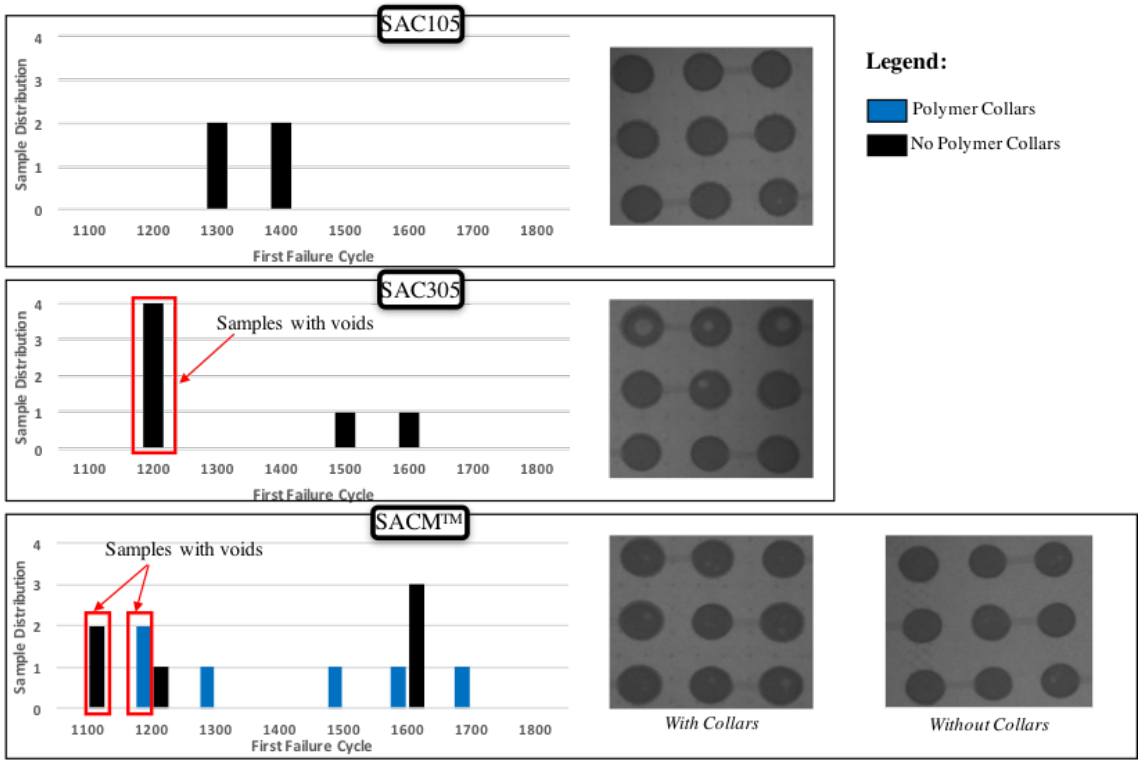


Figure 4.12: Failure distribution for low-CTE glass packages with 200 μm-thick die.

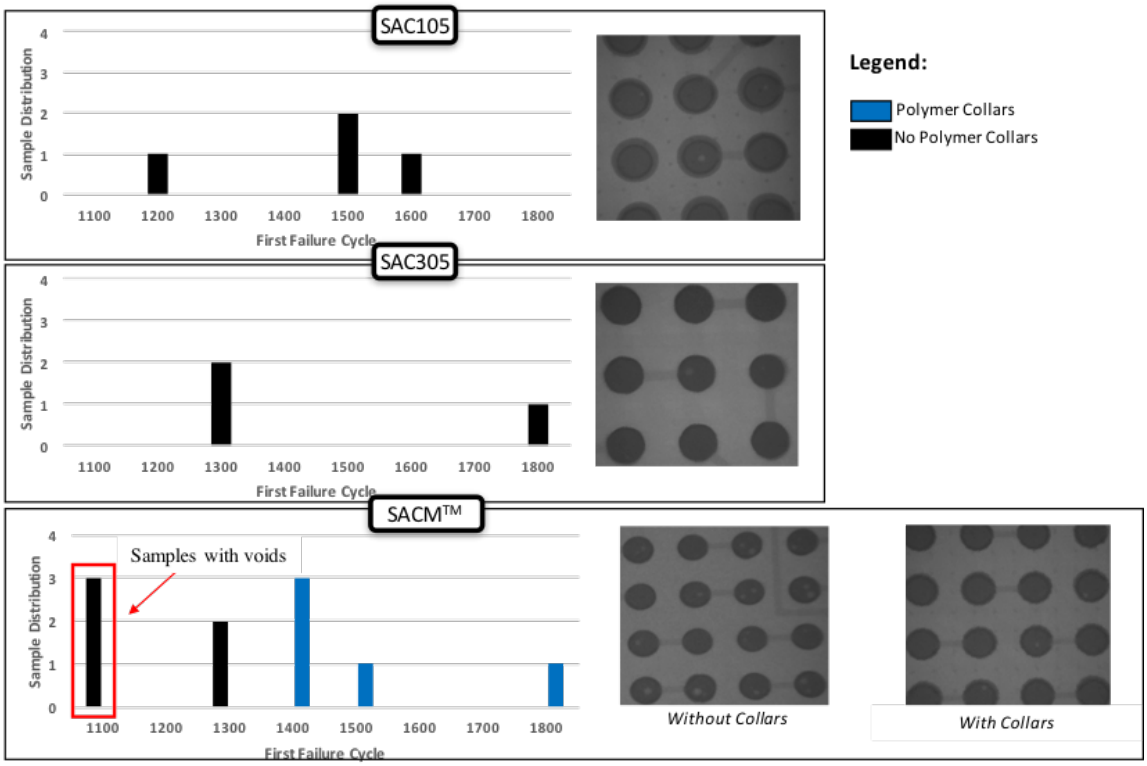


Figure 4.13: Failure distribution for low-CTE glass packages with 100 μm-thick die.

4.4.2 Optical inspection

The failed samples were underfilled to prevent solder smearing during polishing and molded in an acrylic resin for cross-sectioning to identify the predominant failure modes. Inspection of the BGAs in individual rows indicated that the defects occurred mainly in the outer rows and corner circuits, as expected; since the BGAs are less prone to failures when they are closer to the neutral point. Figure 4.14 shows the different failure modes identified in the BGAs after optical inspection. Two main types of failure modes were established. Warpage-related failures can be seen from the non-wets and stretched solders while fatigue failures can be observed with cracks initiating close to the IMC-to-solder interface on the corner BGAs. Early failures were mostly due to warpage-related defects at around 1200-1300 cycles while the fatigue defects failed at around 1500-1600 cycles. Warpage measurements after BGA balling and after board-level assembly reported in [42], and as

shown in Figure 2.20 indicate highest warpage in low-CTE samples with 100 μm-thick dies. This explains the broader failure distribution observed in these assemblies, with, on average, more warpage-related defects than in other failed packages.

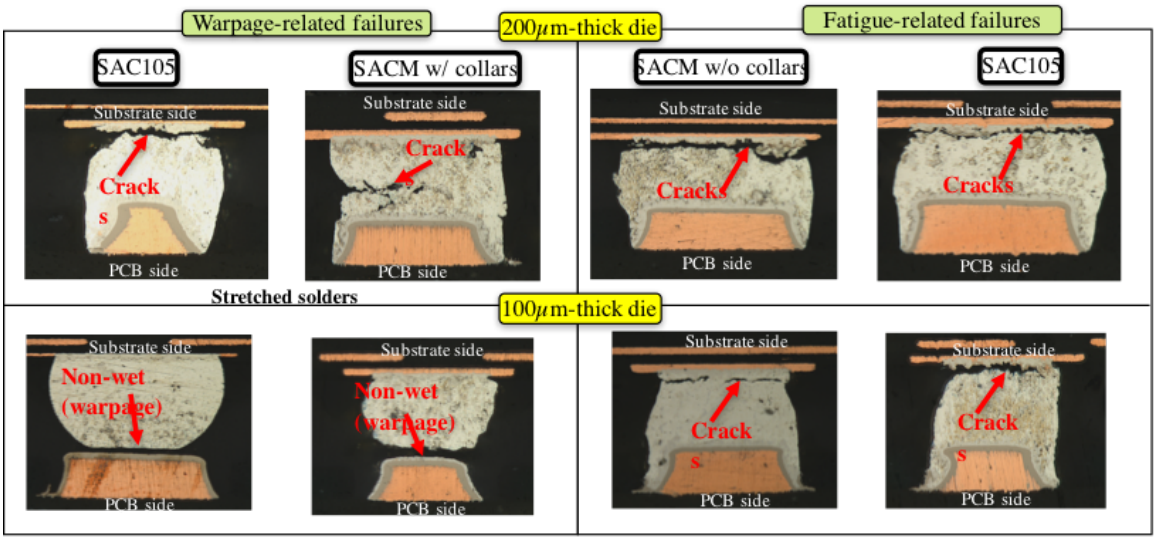


Figure 4.14: Optical characterization of cross-sections of failed assemblies with SAC105, SAC305 and SACm™ BGAs for identification of failure modes.

4.5 Summary

This chapter describes the successful demonstration of system-level reliability of a single-chip glass BGA package, 18.5 mm in body size, meeting JEDEC thermomechanical reliability standards. Using a parametric study, the effect of glass CTE on fatigue life was evaluated. In best conditions of advanced solder interconnection materials, the low-CTE samples can survive a maximum of 1600-1800 cycles for a large single-chip glass BGA package. High-CTE glass packages, currently in test at 2900 cycles, are within the model predicted range of 2100-3800 cycles. Even with the best configuration of low-CTE samples, high-CTE samples are shown to present higher fatigue life. There is limited scalability in terms of reliability of low-CTE packages as we move towards 2.5D and 3D architectures. Higher CTE glass may need to be considered for direct SMT assembly to the board. Thus, there is a need to mitigate warpage and optimize the substrate CTE through a parametric

analysis in order to achieve balanced chip- and- board-level reliability.

CHAPTER 5
THERMOCOMPRESSION BONDING PROCESS DESIGN AND OPTIMIZATION FOR WARPAGE MITIGATION

This chapter details the test vehicle design and fabrication of a single-chip low-CTE organic package. It investigates the thermocompression bonding-induced warpage and its dependence on the thermal bonding profiles. A basic 2D single-chip package was used for this focused study to extract key parameters and trends for warpage control in TCB. Warpage trends as a function of the stage temperature were predicted with a simple coupled thermal-structural finite-element model. Experimental validation was carried out by Shadow-Moire measurements of assemblies built with varying stage temperatures from 70 °C to 150 °C. Guidelines for design of TCB profiles for warpage minimization were finally derived with considerations of assembly throughput to improve board-level SMT yield and system-level reliability.

5.1 Test vehicle design and fabrication of a low-CTE single-chip package

Daisy chain test vehicles at 50 μm pitch were used in this study. The silicon test die, 7.3 mm x 7.3 mm in size, and 200 μm in thickness comprised 528 Cu pillar interconnections in a single peripheral row. The bumps were 30 μm in diameter, and composed of Cu pillars, 30 μm in height, a Ni barrier layer of approximately 2 μm and a SnAg solder cap, 12 μm in height. The test die with peripheral arrangement is shown in the picture of Figure 5.1, while Figure 5.2 gives the detail of the Cu pillar interconnection stack-up. The low-CTE substrates, 17 mm x 17 mm in size, supplied by Walts Co. LTD, consisted of a copper-clad, 200 μm-thick MCL-E-679FG type S organic core by Hitachi Chemical with a CTE of 3.3 ppm/°C. The first copper wiring layer was patterned by subtractive processing. Dielectric ABF-GX-3 build-up layers from Ajinomoto were then laminated to support a second layer

of copper wiring, formed by double-side semi-additive processes. Dry-film solder mask was then applied with an 80 μm-wide slit opening in the bonding area, exposing fan-in fan-out bump-on-trace finger structures, as shown in Figure 5.3. Electroless nickel immersion gold (ENIG) surface finish was finally plated on the exposed traces. The substrate stack-up design is recapped in Figure 5.4.

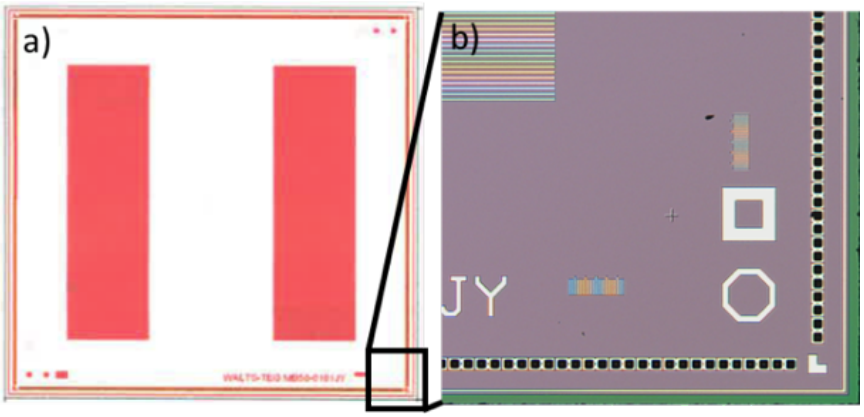


Figure 5.1: Daisy chain test die at 50 μm pitch: a) design and b) optical image of a bumped corner (Walts Co. LTD).

5.2 Finite-element modeling for warpage mitigation in thermocompression bonding process

5.2.1 Geometric model

Finite-element models were created to understand the effect of bonding conditions, in particular of the stage temperature, on warpage of the substrate. Modeling was done in ANSYS™ 15.

Ideally, 2.5D or 3D models are desirable in order to evaluate the results accurately. However, 2D models with plane strain approximation are appropriate for comparison between different assembly processes based on the geometry of Figure 5.5. The mesh was refined at the solder and thin layers. As we move away from the critical regions, the mesh was coarser and the number of elements as compared to the thickness was built to sufficiently capture warpage. An example of the model mesh is depicted in Figure 5.6. The

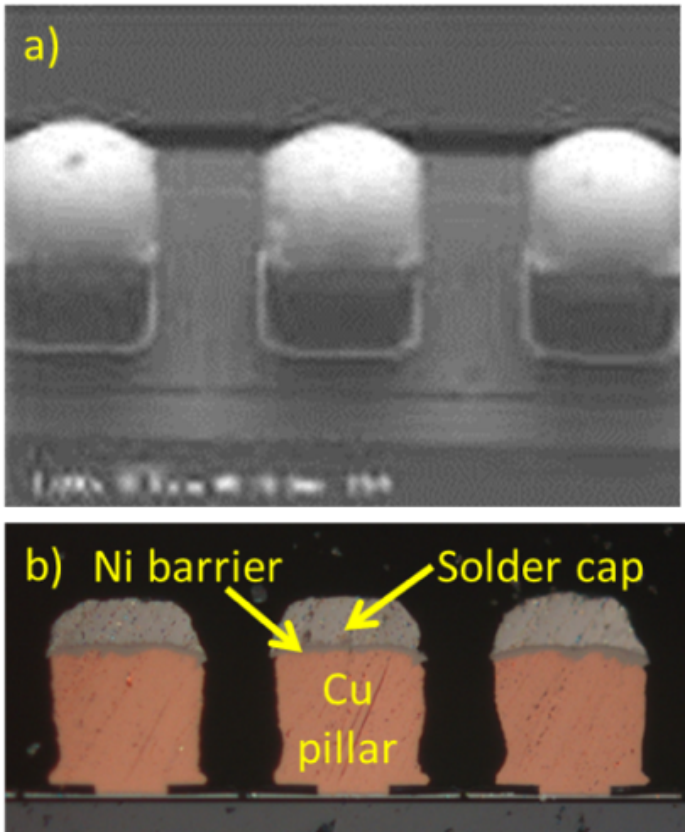


Figure 5.2: Cu pillar interconnections at 50 μm pitch: a) top view, and b) cross-section.

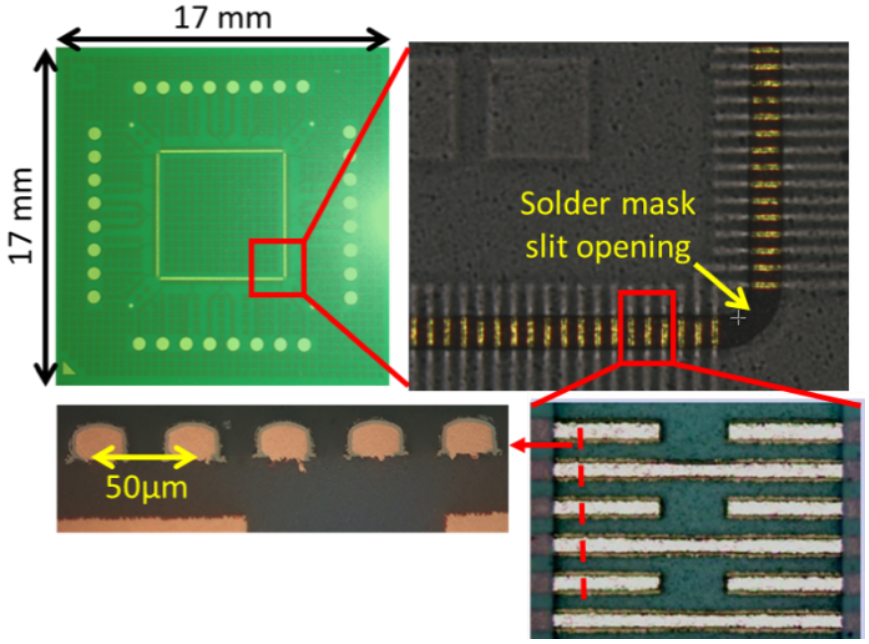


Figure 5.3: Low-CTE organic substrate with bump-on-trace wiring (Walts Co. LTD).

	Thickness (mm)	Material
Solder Resist (on Pattern)	0.015	PSR4000-AUS703
Pattern	0.015	Copper
Insulating Layer (Built up Material)	0.030	ABF-GX-3
Pattern	0.018	Copper
Base Core	0.200	Base Core Material
		E6979FGS
Pattern	0.018	Copper
Insulating Layer (Built up Material)	0.030	ABF-GX-3
Pattern	0.015	Copper
Solder Resist	0.015	PSR4000-AUS703

Figure 5.4: Substrate stack-up design (Courtesy of Walts).

low-CTE substrate, polymer dielectric layers, copper redistribution layers, copper pillar interconnections, and silicon die were included in the model. The die-substrate assembly represents a cut along the diagonal with symmetry boundary conditions at the center. The node at the left bottom was fixed in the y-direction to prevent rigid body motion. A simple elastic model was used for the solder because the effect of a more accurate model is negligible on warpage. Material properties used in the model including the stress-free temperature for each material are given in Table 5.1.

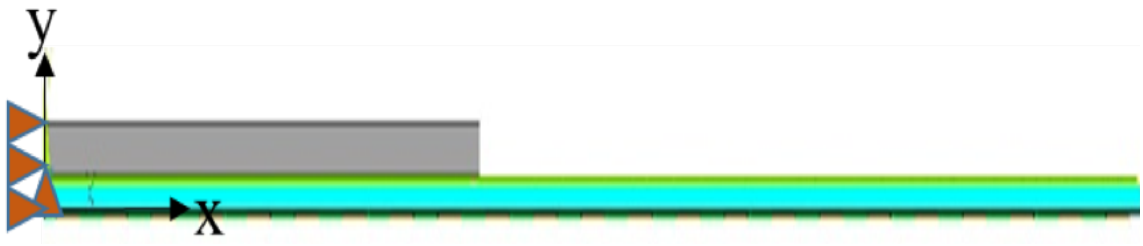


Figure 5.5: Example geometry in modeling. Two-dimensional geometry represents a cut along the diagonal of the package with symmetry at the left.

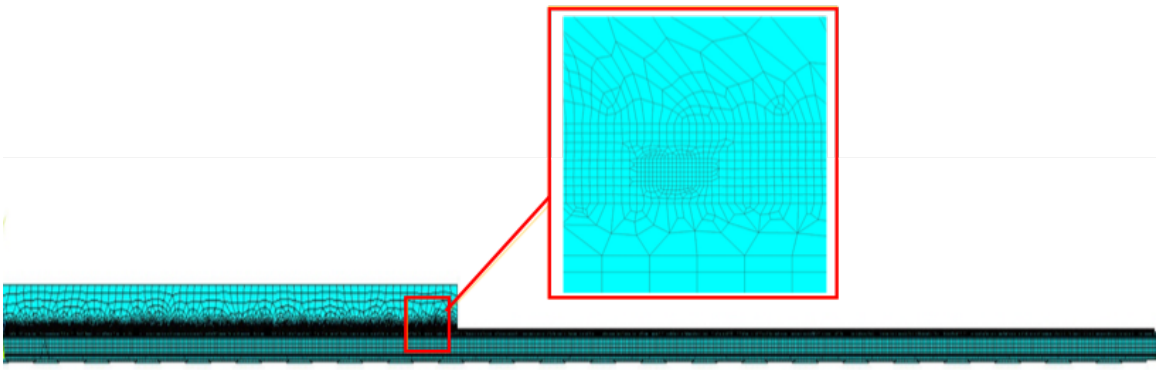


Figure 5.6: Example mesh used in modeling for thermal and mechanical analysis.

Table 5.1: Material properties used in modeling

Material	**E [GPa]**	**CTE [ppm/ °C]**	ν	**Stress-Free Temp. [°C]**
Silicon chip [52]	Ex,z = 169 Ey = 130	2.6	0.28	220
SAC 305 Solder [53]	47.6	24	0.36	220
Low-CTE core	36	3.3	0.12	160
Copper [52]	59.2-80	17	0.33	200
Polymer	5	39	0.34	150

5.2.2 Thermal loading conditions and model-predicted warpage analysis

Coupled thermal-structural finite-element models were constructed to predict the temperature gradient in the package during TCB (Figure 5.7). The solution from the thermal gradient was then used as reference for each material element, and then fed as input to the mechanical model. The solution steps for the four-metal layer package includes the application of a thermal gradient across the package, applying a cooling model from solder melt temperature to room temperature and feeding the solution to a mechanical model that calculates the displacement in y-direction.

This simulation mimics the warpage during assembly process. The substrate bottom was simulated to be heated at temperatures varying from 70 °C to 150 °C and the die top correspondingly from 360 °C to 305 °C, with sufficient thermal budget to melt the solder in every condition. The model was solved and the predicted warpage was plotted for different CTEs as a function of the stage temperature in Figure 5.8. The warpage curve is predicted

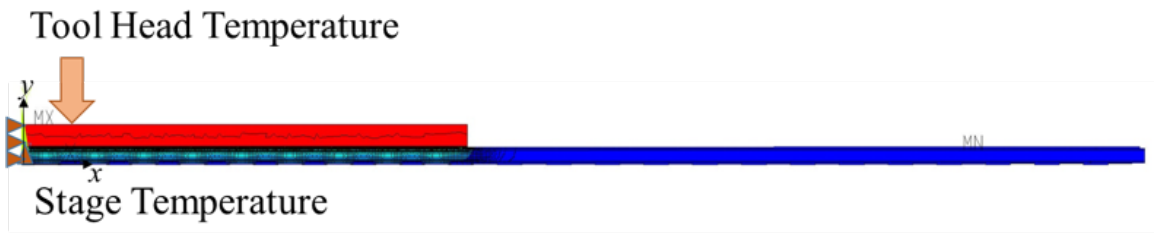


Figure 5.7: Example of thermal gradient in the package during TCB.

to be fairly linear and scalable for different CTEs.

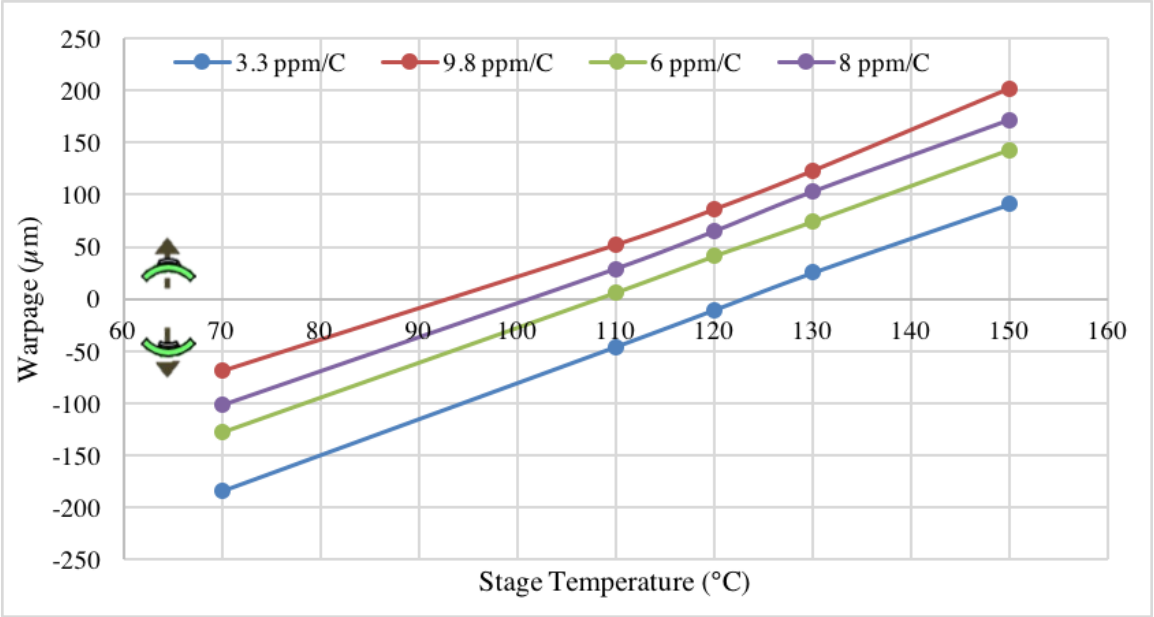


Figure 5.8: Predicted warpage as a function of stage temperature for a range of substrate CTEs.

The warpage plot predicts that there is a change in direction of warpage with increasing stage temperatures. This change occurs at temperatures beyond 120 °C for a stage temperature 70 °C. At higher stage temperatures, the magnitude of warpage continues to increase, but in the opposite direction. These predictions correlate well with observations in silicon-to-silicon bonding, showing lesser warpage at higher stage temperatures [39]. An example of warpage prediction at 70 °C stage temperature is shown in Figure 5.9.

From the plot (Figure 5.10) we can see that the stage temperature at which minimum warpage occurs decreases with increasing CTEs. Therefore, through fine tuning of the TCB

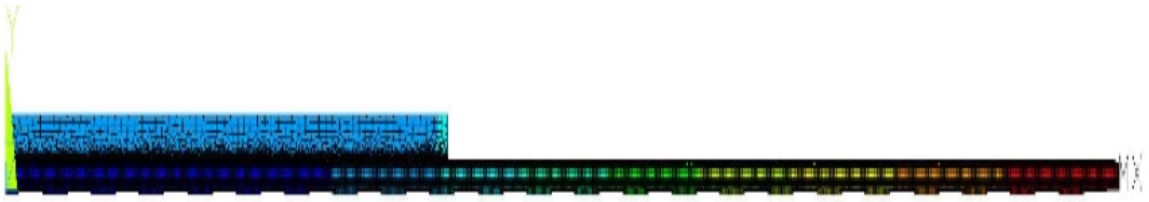


Figure 5.9: Predicted warpage of low-CTE package assembled using 70 °C stage temperature conditions.

process, warpage can be minimized for a given substrate CTE.

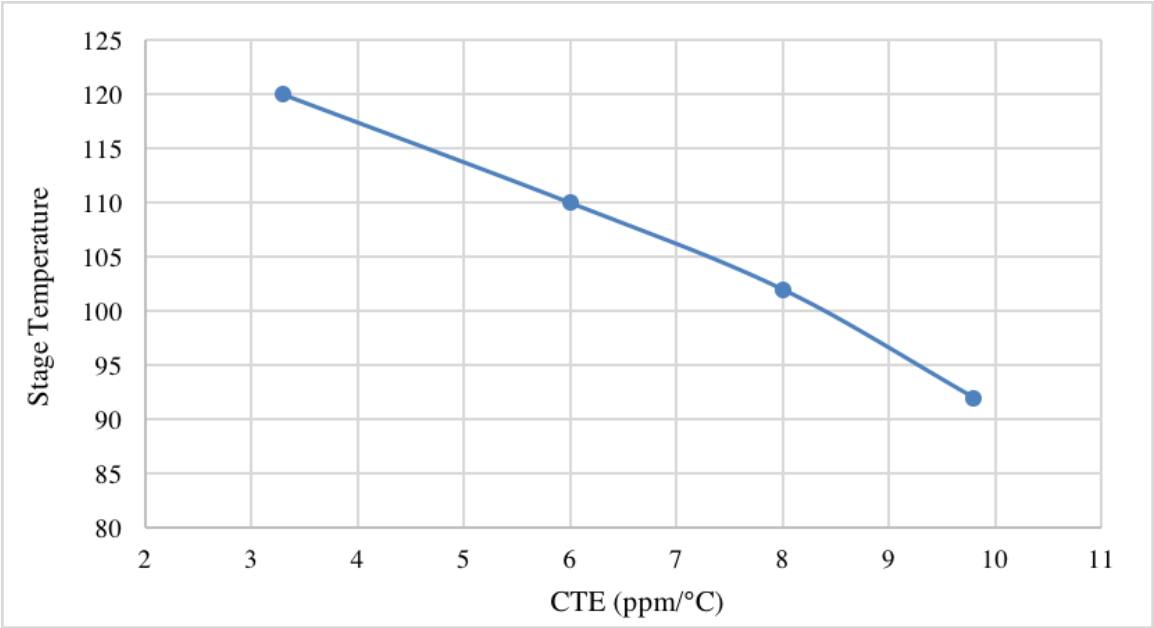


Figure 5.10: Minimum warpage stage temperature as a function of substrate CTE.

5.3 Thermocompression bonding assembly process and parameters

The test vehicles described in the previous section were assembled by dip-flux thermocompression bonding using a semi-automatic Finetech Fineplacer Matrix flip-chip bonder with a placement accuracy of 3 μm. The conventional underfilling step was skipped in this study as variability in fillet size have been shown to have a strong effect on substrate warpage and thus needs to be decoupled and studied separately [54]. The TCB force and thermal profiles were designed by varying the tool head peak temperature and dwell time at a given stage

temperature, heating and cooling ramp rates, and applied force. The bonding conditions were optimized based on systematic electrical yield analysis by DC measurements of the as-bonded daisy-chain resistances, and die-shear testing followed by qualitative observation of the fracture profiles and solder spread on the traces.

A fine control of the solder spread on the trace was found critical in achieving good yield. Due to the slit passivation opening and the lack of confinement that a PAM would provide, the solder is free to flow on the exposed trace during the process. Excellent wettability of the ENIG surface finish creates a driving force for solder lateral spread, leading to its depletion from the joints if not restricted. This mechanism is illustrated in Figure 5.11, showing the cross-sections of a yielded and non-yielded samples, respectively. As interdiffusion rates are much higher in liquid phase, lateral spread of the solder is mostly governed by the duration spent above its melting point that needs to be precisely controlled. Such fine control is challenging with the lab-scale bonder used in this study, due to the slow achievable ramp rates with a maximum of 6 °C/s compared to 400 °C/s in production. An increase by only 1 °C of the tool peak temperature was found to degrade the yield from all 45 connected daisy chains, to only three corner chain readings. Proper melting and wetting of the solder was obtained with tool head peak temperatures of 360 °C, 340 °C, 325 °C and, 305 °C, corresponding to stage temperatures of 70 °C, 90 °C, 120 °C and, 150 °C, respectively, with no dwell time at peak temperature. The force profile was optimized along the thermal one, with an initial 1N force applied until reaching a temperature in the solder of approximately 150 °C, then released to 0.5 N during reflow.

To study interactions with the bonder, in particular coupling effects with the stage induced by vacuum holding of the substrate, assembly was carried out with and without vacuum in the same conditions. To virtually strengthen the vacuum hold, the substrates were taped to the stage with Kapton® tape. At least two samples were bonded in the same conditions to give representative trends for TCB-induced warpage.

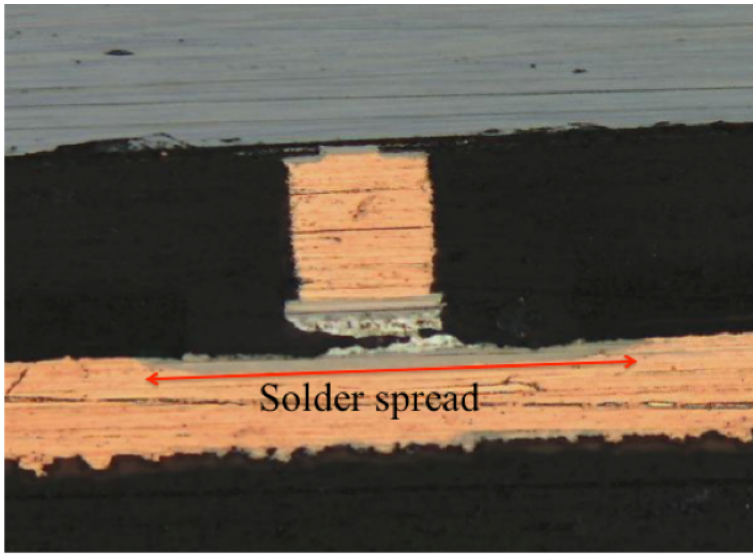


Figure 5.11: Cross-section of a non-yielded assembly with excessive solder lateral spread.

5.4 Analysis of warpage during assembly

5.4.1 Experimental warpage measurements

Substrate warpage was measured by Shadow-Moire using Akrometrix's Thermoire PS200S. Details on Shadow-Moire can be found in various publications [54, 55, 56]. Warpage is defined as the difference between the highest and lowest points on the package after tilt has been accounted for. Warpage was measured over a temperature range of 25 °C to 260 °C from the substrate side. For each bonding condition, two or three samples were measured. The error margin of the tool and grating used is 1.5 μm. For warpage convention, the frown shape (die is on top) is positive and the smile shape is negative, as indicated by the icons in Figure 5.12.

Figure 5.12 shows a typical warpage curve evaluated by Shadow Moire for the 70 °C stage temperature bonding process. On heating, the warpage increases with increasing temperature since the substrate expands at a higher rate than the silicon. Above 200 °C, the solder is in molten state and provides no mechanical coupling between the die, causing the warpage to decrease.

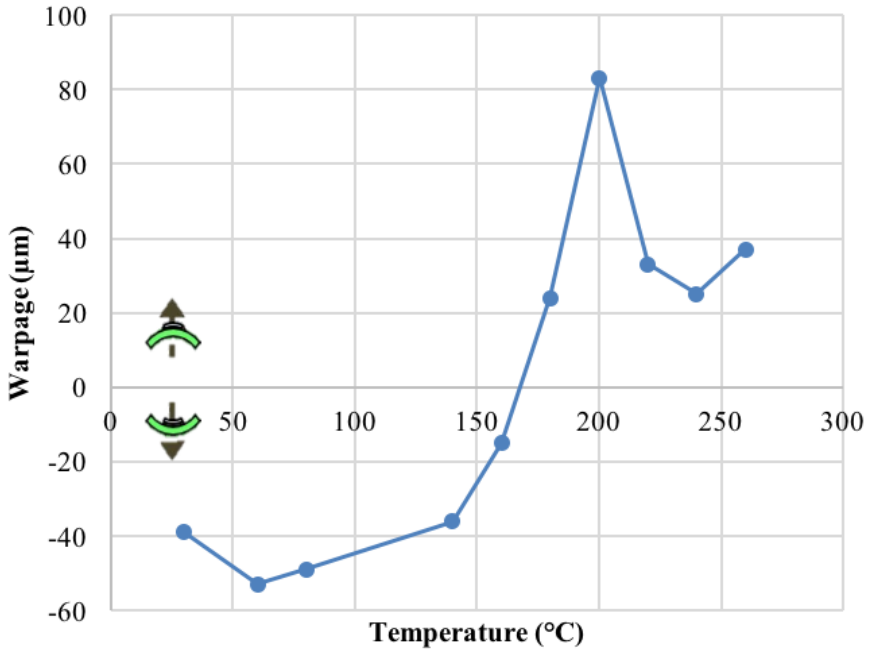


Figure 5.12: Package warpage as a function of temperature for TCB profile with 70 °C stage temperature.

5.4.2 Effect of TCB stage temperature on warpage without substrate-stage coupling

Assembly was performed without coupling the substrate to the stage and the resulting room temperature warpage minus the unassembled substrate warpage is shown in Figure 5.13 as a function of stage temperature. It is seen that room temperature warpage for the 70 °C stage temperature assembly condition is a smile-shaped warpage with higher magnitude than that at 90 °C and 120 °C. This is because the thermal gradient in the package when the solder is frozen is larger, which means the die has a larger temperature drop to room temperature and the substrate has a smaller temperature drop to room temperature. As the stage temperature is increased, the die head temperature is decreased, varying the thermal excursion in the package. In the 150 °C stage temperature case, the substrate contracted more than the die, resulting in positive or frown-shaped warpage. Between the substrate shrinking more (150 °C stage temperature case) and the die shrinking more (70 °C, 90

°C, and 120 °C stage temperature cases), there exists a balance point where the different thermal excursions offset the CTE mismatch between the package and die and result in minimum net warpage. Therefore, based on the trend observed in Figure 65, the minimum warpage can be controlled by finely tuning the TCB profile.

Overall, the trend observed for TCB assembly without coupling between the substrate and stage was the same as was predicted in modeling (Figure 5.8). In both theory and experiments, the stage temperature to minimize warpage was about 120 °C.

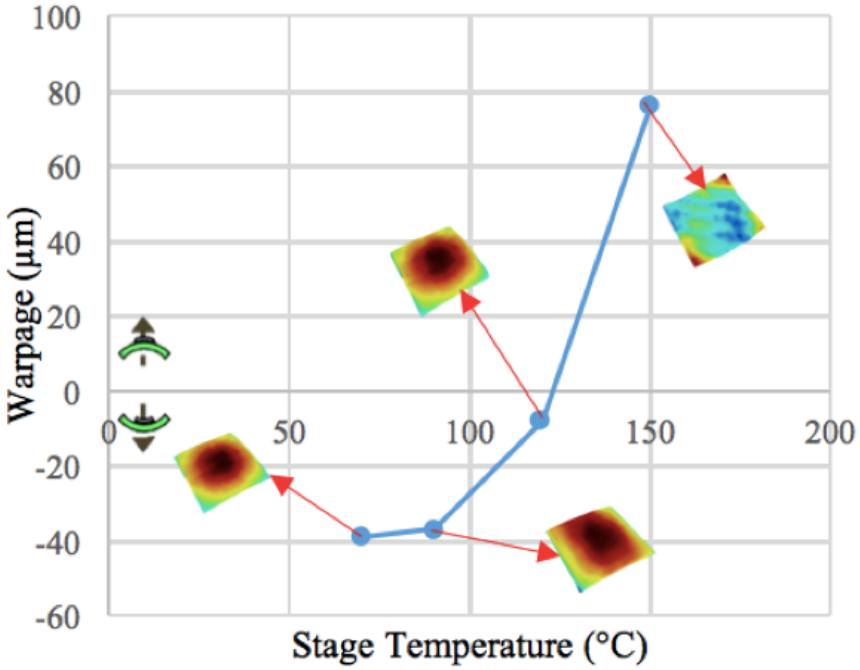


Figure 5.13: Room temperature package warpage as a function of stage temperature without coupling the substrate to the stage.

5.4.3 Effect of stage temperature on substrate-stage coupling

Assembly was performed with coupling between the substrate and stage and the resulting room temperature warpage minus the unassembled substrate warpage is shown in Figure 66 as a function of stage temperature. The coupled case showed smile-shaped warpage with similar magnitude for temperature gradients from 70 °C and 150 °C stage temperature

bonding profiles. This result is different from the uncoupled case (Figure 5.14), which shows increasing room temperature warpage with increasing stage temperature. The only difference between the two cases was the coupling applied between the stage and substrate.

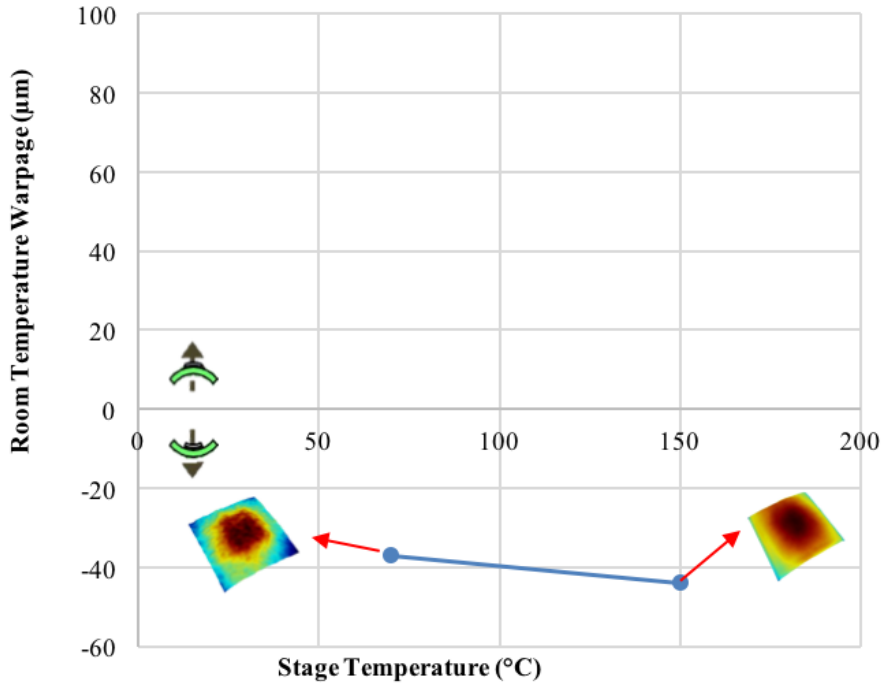


Figure 5.14: Room temperature package as a function of stage temperature warpage with coupling the substrate to the stage.

Why the coupled and uncoupled cases produce different trends is not yet fully understood, though there are several possible factors that play a role. Mechanically coupling the substrate and stage means that the substrate cannot warp during the bonding process. While pressure is applied to the die during the TCB process, the region outside the die is free to warp in the uncoupled case. Hence the substrate warpage is different between the coupled and uncoupled cases. Also, the coupling to the stage may be preventing further expansion of the substrate during assembly, which would produce strain in the opposite direction as the thermal contraction. There may be thermal differences between the coupled and uncoupled cases as well. In the coupled case, the substrate is better adhered to the stage, lowering the contact resistance, and changing the thermal gradient when the solder freezes.

The relatively small change in room temperature warpage based in the coupled case implies that the temperature profile applied during TCB has little effect on the resulting warpage. Previous literature [57] has reported yield changes based on the temperature profile, which indicates that the temperature conditions should have an effect, however, the coupling effects were not studied.

5.5 Conclusions

For TCB assembly of low-CTE substrates, higher stage temperature is desired because it increases throughput by reducing the amount of time necessary for heating and cooling during the assembly process. For example, comparing 70 °C and 150 °C stage temperatures with 200 °C/s heating rate on a 7 's process, the 70 °C stage temperature requires 6% more time from heating alone. Although this work deals with the dip-flux process and no underfill for simplicity, it is common to use pre-applied underfill materials like NCP, NCF. However, they cannot be used beyond temperatures of 90 °C, thereby enforcing constraints on the temperature profile, limiting the use of higher stage temperatures. In this chapter, we concluded that the stage temperature for minimum warpage decreases with increasing substrate CTEs, allowing the use of pre-applied underfill materials at high-CTE glass packages.

CHAPTER 6
SUMMARY AND CONCLUSIONS

This chapter summarizes the overall research work carried out to achieve, for the first time, reliable, direct SMT assembly of a large 2.5D, high-density glass BGA package to the PWB. Finite-element models were built to optimize the glass CTE in the 3.3 - 9.8 ppm/°C for balanced chip- and board-level reliability. While chip-level reliability is largely achieved regardless of CTE of glass, board-level reliability was found most critical, driving towards higher CTEs of glass exceeding 6 ppm/°C. Advanced doped solder alloys and polymer collars were implemented to improve the board-level thermal cycling performance with minimum system-level impact. Modeling of package warpage characteristics after chip-level reflow assembly indicated an expected yield loss in SMT assembly for glass CTEs above 6ppm/K. These conflicting requirements in glass CTE were resolved through design and optimization of thermocompression bonding processes for warpage mitigation. In summary, this work highlighted the importance of co-designing package CTE and chip-level assembly processes to achieve system-level reliability with such aggressive design rules as required to meet the performance and miniaturization needs of emerging high-performance systems. Directions for future work are briefly discussed to validate these preliminary results and gain a better fundamental understanding of advanced assembly processes such as thermocompression or laser-assisting bonding.

Based on the objectives defined in Chapter 1, three technical challenges were defined along with their associated research tasks. These research tasks performed to address the challenges are summarized in Table 6.1.

Table 6.1: Summary of the technical challenges addressed through associated research tasks

Technical Challenges	Research Tasks
Balanced chip- and- board-level assembly	Modeling and design of glass package (size, thickness and CTE) for chip- and-board-level reliability
Aggravated plastic strains in BGAs	Design and demonstration of board-level interconnections for strain relief and reliability
Warpage mitigation in assembly	Modeling and design of assembly process and sequence for controlled warpage

6.1 Summary of task 1 results

Finite-element modeling was performed to investigate the effect of the CTE of a high-density 2.5D glass interposer-package, 100 μm in thickness and 30 mm x 40 mm in body size, on system-level thermomechanical reliability and substrate warpage with direct SMT to board at 650 μm pitch. A 2D plane strain approximation model was used to analyze and predict the i) fatigue life at chip and board levels and ii) package warpage induced by chip-level assembly as a function of glass CTE to extract the optimum CTE range for balanced reliability and SMT yield. Coffin-Manson and Engelmaier-Wild models were used to predict the solder fatigue life. Package warpage was extracted through calculation of displacements of the substrate and board in the y-direction. Modeling predictions indicated that, while chip-level reliability is not of concern in this package structure, higher CTEs of glass were required to achieve board-level reliability. The required 1000 thermal cycles by JEDEC standards are achieved for glass CTEs above 6 ppm/°C. However, the package warpage characteristics indicated that, after chip-level reflow assembly, maximum warpage

is, as expected, observed in high-CTE glass package assemblies. The maximum admissible package warpage beyond which yield loss is expected in SMT assembly was calculated through JEITA-EDR standards for open solder joints. It gave a maximum glass CTE of 6 ppm/°C, bringing conflicting requirements in glass CTEs for reliability and warpage. These modeling predictions, albeit preliminary due to lack of timely experimental validation, brought the need to i) improve the more critical board-level reliability through advances in interconnection materials and strain-relief mechanisms with minimum system-level impact, and ii) mitigate warpage during chip-level assembly by migrating from reflow to thermocompression bonding, giving independent control over the chip and package thermal expansion in process.

6.2 Summary of task 2 results

This task focused on improving board-level reliability by using the most advanced interconnection materials Mn-doped SACm™ solder alloy by Indium Corporation and polymer collars by Namics to provide strain relief in BGA solder joints with minimum system-level impact. While the benefits of polymer collars on board-level thermomechanical reliability were previously demonstrated, this research was the first implementation and evaluation of the fatigue performance of the SACm alloy at 400 μm pitch. Conventional SAC105 and SAC305 alloys were used as reference. A single-chip daisy-chain glass BGA test vehicle, 18.5 mm x 18.5 mm in body size, with direct SMT to the board was considered in this study. As the SACm™ alloy is a new solder material, no constitutive model describing its mechanical behavior is available at this time, limiting its fatigue life evaluation to empirical data. Finite-element models were however developed to predict the fatigue life of SAC105 BGAs as a function of glass CTE. Similar trends as in Task 1 were observed, with chip-level reliability exceeding JEDEC standards by a large margin. At board-level, assemblies with low-CTE (3.3 ppm/°C) glass were predicted to survive upto 1900 thermal cycles, while assemblies with high-CTE glass could survive up to 3000 cycles, and

are still in test. Experimental thermal cycling test results correlates well with the model predictions for board-level reliability passing the JEDEC standard of 1000 cycles, regardless of glass CTE and solder alloy. As expected, for low-CTE assemblies, SAC105 BGAs failed the earliest at 1300 cycles and the Mn-doped SACm™ alloy exhibited superior fatigue life of 1450 cycles comparable to that of SAC305 with 1550 cycles. Furthermore, samples with polymer collars showed improved performance with 1600 cycles to failure. Failure analysis was subsequently carried out to conclude on failure modes. Optical inspection through cross-sectioning of failed low-CTE samples revealed two predominant failure modes - warpage and fatigue defects. Samples with warpage-related defects failed prior to the expected number of cycles. The high-CTE glass packages are still in test at 3000 cycles and are within the range of model predictions (2100-3800). Advanced interconnection materials yielded a consistent increase in fatigue life by ≈25%, marginally extending applicability of glass BGA packages. Further improvements in fatigue performance at board level can be achieved with compliant interconnections beyond conventional solder balls, with manufacturability, performance and cost trade-offs.

6.3 Summary of task 3 results

In this task, a novel methodology to co-design chip-level assembly processes as a function of the CTE of the substrate was proposed and demonstrated to minimize the resulting package warpage. As opposed to isothermal heating in conventional reflow, the thermal gradient established in the package assembly in thermocompression bonding gives control over the package warpage so that board-level assembly yield could theoretically be maintained regardless of the CTE of the substrate. Mitigation of warpage produced in chip-level assembly is a key enabler to realize the targeted direct SMT assembly of a large, 2.5D glass CTE package.

Finite-element models were first developed to predict package warpage based on the temperature profile and, subsequently, thermal gradient, applied during thermocompression

bonding for a CTE range of 3.3 - 9.8 ppm/°C. The model predicted a change in direction of warpage with increasing stage temperatures, regardless of the glass CTE. The stage temperature at which minimum (zero) warpage is observed decreases with increase in CTE between ≈120 °C for low-CTE glass and 90 °C for high-CTE glass.

To validate these modeling results, silicon chips were bonded on low-CTE glass and organic substrates by dip-flux thermocompression bonding with stage temperatures varying between 70 °C and 150 °C. A lab-scale tool with low vacuum holding the substrate onto the stage was first used and the room temperature warpage matched the theoretical predictions, showing increasing warpage with increasing stage temperature, with a smile-shape at low stage temperature and frown-shape at high stage temperature. In other words, the net or absolute warpage first decreases, then increases. The minimum absolute warpage was obtained with a stage temperature of 120 °C, in good accordance with modeling predictions.

Assembly was also performed using Kulicke and Soffa's production-scale APAMA C2S thermocompression bonder where the substrate is tightly coupled to the stage by application of strong vacuum. Over the same temperature range, the bonding thermal profiles were found to have little effect on the room temperature warpage. While it is still unclear why different trends were observed, these discrepancies were tentatively attributed to interactions with the assembly tool, including: i) the tool head being smaller than the die size with strong vacuum preventing chip deformation, ii) stage vacuum being stronger and preventing substrate deformation etc. These tool interactions are not accounted for in the simple model that was built. More research is, therefore, required to gain a better fundamental understanding of thermocompression bonding and fully exploit its benefits in providing fine control over die and substrate warpage.

Underfill was also not considered in this work to decouple the effect of assembly and underfilling on substrate warpage. While a similar trend is expected in package warpage as a function of CTE in presence of underfill, a shift in stage temperature giving minimum net warpage is expected, corresponding to the change in reference temperature to that of

the glass transition temperature of the underfill as opposed to the solidification temperature of the solder. When using pre-applied underfill materials such as non-conductive pastes (NCPs) and films (NCFs), additional restrictions apply in designing the bonding thermal profiles, with stage temperatures limited to below 90 °C by the thermal stability on stage of the underfill material. These restrictions are, however, compatible with the recommended stage temperature for high-CTE glass required for board-level reliability.

6.4 Conclusions

This research work demonstrates the feasibility direct SMT assembly of a 2.5D glass BGA package to the board through parametric finite element modeling and focused experimental validation. The need to move towards higher CTE glass substrates of $>$6 ppm/°C for achieving balanced chip- and- board-level reliability is established. Board-level reliability was enhanced by 25% with the use of innovative materials like doped solders and polymer collars. Failure analysis was performed on single-chip mobile glass packages to confirm the improvement in board-level reliability through the use of these advanced interconnection materials. Through the extensive failure analysis, warpage was found to be a critical factor in affecting yield and reliability of the system. Ways to mitigate assembly warpage through careful control of thermocompression bonding process was demonstrated. The experimental data correlates well with model predictions and provides a holistic approach for enhancing the system-level reliability of a 2.5D glass package. As future work, there is a need to experimentally validate the model predictions for system-level reliability and warpage during assembly. Fabrication of low- and- high-CTE 2.5D glass packages are in process to provide conclusive results. Additionally, the TCB process needs to be designed and developed for investigating the role of i) capillary and pre-applied underfills and ii) tool interaction in shifting the warpage curves at different glass CTEs. Alternative bonding processes like laser-assisted bonding can also be considered to achieve higher throughput and better control over the reaction. However, the assembly induced package warpage due

to the thermal gradient and CTE mismatch across the package needs to be fully understood and evaluated. This technology can be used to solve automotive electronics packaging needs, where the reliability requirements are double. In that case, compliant interconnections may need to be developed due to the material limitations of standard BGA solders.

REFERENCES

[1] R. R. Tummala, "System scaling: The next frontier for a new era of electronic systems," *Chip Scale Review*, 2015.

[2] H. Vuong, "Mobile memory technology roadmap," in *JEDEC's Mobile Forum 2013*.

[3] N. Kim, C. Shin, D. Wu, J. H. Kim, and P. Wu, "Performance analysis and optimization for silicon interposer with Through Silicon Via (TSV)," in *SOI Conference (SOI), 2012 IEEE International*, IEEE, pp. 1–2, ISBN: 1467326917.

[4] R. Tummala, *Fundamentals of microsystems packaging*. McGraw Hill Professional, 2001, ISBN: 0071418075.

[5] P. Stanley-Marbell, V. C. Cabezas, and R. P. Luijten, "Pinned to the walls - impact of packaging and application properties on the memory and power walls," in *Low Power Electronics and Design (ISLPED) 2011 International Symposium on*, IEEE, pp. 51–56, ISBN: 1612846602.

[6] G. Hariharan, R. Chaware, L. Yip, I. Singh, K. Ng, S. Pai, M. Kim, H. Liu, and S. Ramalingam, "Assembly process qualification and reliability evaluations for heterogeneous 2.5 D FPGA with HiCTE ceramic," in *Electronic Components and Technology Conference (ECTC), 2013 IEEE 63rd*, IEEE, pp. 904–908, ISBN: 1479902322.

[7] D. U. Lee, K. S. Lee, Y. Lee, K. W. Kim, J. H. Kang, J. Lee, and J. H. Chun, "Design considerations of hbm stacked dram and the memory architecture extension," in *Custom Integrated Circuits Conference (CICC), 2015 IEEE*, IEEE, pp. 1–8, ISBN: 1479986828.

[8] Y. Lin, C. Hsieh, C. Yu, C. Tung, and C. Doug, "Study of the thermo-mechanical behavior of glass interposer for flip chip packaging applications," in *Electronic Components and Technology Conference (ECTC), 2011 IEEE 61st*, IEEE, pp. 634–638, ISBN: 1612844987.

[9] T. Hisada, T. Aoki, J. Asai, and Y. Yamada, "FEM analysis on mechanical stress of 2.5 D package interposers," *Transactions of The Japan Institute of Electronics Packaging*, vol. 5, no. 1, pp. 107–114, 2012.

[10] *Embedded Multi-Die Interconnect Bridge (EMIB)*, `http://www.intel.com/content/www/us/en/foundry/emib-an-interview-with-babak-sabi.html`.

[11] S. Bezuk, "Applications drive packaging challenges in growth markets," in *Con Fab Conference*.

[12] "Amkor's next generation of packaging solutions, SLIM & SWIFT," in *SEMICON*.

[13] C. Zwenger, R. Huemoeller, J. Kim, D. Kim, W. Do, and S. Seo, "Silicon wafer integrated fan-out technology," *Additional Papers and Presentations*, vol. 2015, no. DPC, pp. 000 217–000 247, 2015.

[14] G. Subbarayan, *A systematic approach for selection of best lead-free printed circuit board (PCB) surface finish*. State University of New York at Binghamton, 2007, ISBN: 0549316698.

[15] Y. P. Ho, J. T. Luo, K. Hsu, and A. Chen, "Reliability test and imc investigation of lead and lead free solder joints on different surface finish processes," in *Microsystems, Packaging, Assembly and Circuits Technology Conference, 2009. IMPACT 2009. 4th International*, IEEE, pp. 637–640, ISBN: 1424443415.

[16] H.-T. Lee, M.-H. Chen, H.-M. Jao, and T.-L. Liao, "Influence of interfacial intermetallic compound on fracture behavior of solder joints," *Materials Science and Engineering: A*, vol. 358, no. 1, pp. 134–141, 2003.

[17] W. Liu and N.-C. Lee, "The effects of additives to snagcu alloys on microstructure and drop impact reliability of solder joints," *JOM*, vol. 59, no. 7, pp. 26–31, 2007.

[18] L.-W. Lin, J.-M. Song, Y.-S. Lai, Y.-T. Chiu, N.-C. Lee, and J.-Y. Uan, "Alloying modification of snagcu solders by manganese and titanium," *Microelectronics Reliability*, vol. 49, no. 3, pp. 235–241, 2009.

[19] W. Liu, N.-C. Lee, A. Porras, M. Ding, A. Gallagher, A. Huang, S. Chen, and J. C. Lee, "Achieving high reliability low cost lead-free SAC solder joints via Mn or Ce doping," in *Electronic Components and Technology Conference, 2009. ECTC 2009. 59th*, IEEE, pp. 994–1007, ISBN: 1424444756.

[20] L. Wang and C. Wong, "Novel thermally reworkable underfill encapsulants for flip-chip applications," in *Electronic Components & Technology Conference, 1998. 48th IEEE*, IEEE, pp. 92–100, ISBN: 0780345266.

[21] H. Peng, R. W. Johnson, G. T. Flowers, A.-G. Ricketts, E. K. Yeager, M. M. Konarski, A. Torres-Filho, and L. Crane, "Underfilling fine pitch bgas," *Electronics Packaging Manufacturing, IEEE Transactions on*, vol. 24, no. 4, pp. 293–299, 2001.

[22] H. Ishida and K. Matsushita, "Characteristics of ceramic BGA using polymer core solder balls," in *Electronic Components and Technology Conference (ECTC), 2014 IEEE 64th*, IEEE, pp. 404–410, ISBN: 1479924075.

[23] M. Brunnbauer, T. Meyer, G Ofner, K Mueller, and R Hagen, "Embedded wafer level ball grid array (eWLB)," in *Electronic Manufacturing Technology Symposium (IEMT), 2008 33rd IEEE/CPMT International*, IEEE, pp. 1–6, ISBN: 1424433924.

[24] B. Cheng, D. De Bruyker, C. Chua, K. Sahasrabuddhe, I. Shubin, J. E. Cunningham, Y. Luo, K. F. Bohringer, A. V. Krishnamoorthy, and E. M. Chow, "Microspring characterization and flip-chip assembly reliability," *IEEE Transactions on Components, Packaging and Manufacturing Technology*, vol. 3, no. 2, pp. 187–196, 2013.

[25] L. Ma, S. K. Sitaraman, Q. Zhu, K. Klein, and D. Fork, "Design and development of stress-engineered compliant interconnect for microelectronic packaging," in *Nanopackaging*. Springer, 2008, pp. 465–490.

[26] M Topper, V Glaw, P Coskina, J Auersperg, K Samulewicz, M Lange, C Karduck, S Fehlberg, O Ehrmann, and H Reichl, "Wafer level package using double balls," in *Advanced Packaging Materials: Processes, Properties andInterfaces, 2000. Proceedings. International Symposium on*, IEEE, pp. 198–200, ISBN: 0930815599.

[27] V. S. Rao, V Kripesh, S. W. Yoon, D. Witarsa, and A. Tay, "Bed of nails: Fine pitch wafer-level packaging interconnects for high performance nano devices," in *Electronic Packaging Technology Conference, 2005. EPTC 2005. Proceedings of 7th*, vol. 2, IEEE, 6 pp. ISBN: 0780395786.

[28] X. Qin, P. M. Raj, V. Smet, and R. Tummala, "Direct SMT interconnections of large Low-CTE interposers to printed wiring board using copper microwire arrays," *IEEE Transactions on Components, Packaging and Manufacturing Technology*, vol. 5, no. 11, pp. 1709–1719, 2015.

[29] G. Menezes, V. Smet, M. Kobayashi, V. Sundaram, P. M. Raj, and R. Tummala, "Large low-CTE glass package-to-PCB interconnections with solder strain-relief using polymer collars," in *Electronic Components and Technology Conference (ECTC), 2014 IEEE 64th*, IEEE, pp. 1959–1964, ISBN: 1479924075.

[30] W. Lin, S. Wen, A. Yoshida, and J. Shin, "Evaluation of raw substrate variation from different suppliers and processes and their impact on package warpage," in *Electronic Components and Technology Conference (ECTC), 2012 IEEE 62nd*, IEEE, pp. 1406–1411, ISBN: 146731966X.

[31] P. Lall, K. Patel, and V. Narayan, "Model for prediction of package-on-package warpage and the effect of process and material parameters," in *Electronic Components and Technology Conference (ECTC), 2013 IEEE 63rd*, IEEE, pp. 608–622, ISBN: 1479902330.

[32] A. Takahashi, K. Kobayashi, S. Arike, N. Okano, H. Nakayama, A. Wakabayashi, and T. Suzuki, "High density substrate for semiconductor packages using newly de-

veloped low CTE build-up materials," in *Advanced Packaging Materials: Processes, Properties andInterfaces, 2000. Proceedings. International Symposium on*, pp. 216–220.

[33] L. Wei and L. Min Woo, "Pop/csp warpage evaluation and viscoelastic modeling," in *Electronic Components and Technology Conference, 2008. ECTC 2008. 58th*, pp. 1576–1581, ISBN: 0569-5503.

[34] K. Saban, "Xilinx stacked silicon interconnect technology delivers breakthrough FPGA capacity, bandwidth, and power efficiency," *Xilinx, White Paper*, vol. 1, wP380, 2011.

[35] B. Sawyer, H. Lu, Y. Suzuki, Y. Takagi, M. Kobayashi, V. Smet, T. Sakai, V. Sundaram, and R. Tummala, "Modeling, design, fabrication and characterization of first large 2.5 D glass interposer as a superior alternative to silicon and organic interposers at 50 micron bump pitch," in *Electronic Components and Technology Conference (ECTC), 2014 IEEE 64th*, IEEE, pp. 742–747.

[36] B. Sawyer, B. C. Chou, S. Gandhi, J. Mateosky, V. Sundaram, and R. Tummala, "Modeling, design, and demonstration of 2.5 D glass interposers for 16-channel 28 Gbps signaling applications," in *Electronic Components and Technology Conference (ECTC), 2015 IEEE 65th*, IEEE, pp. 2188–2192, ISBN: 1479986097.

[37] V. Smet, T.-C. Huang, S. Kawamoto, B. Singh, V. Sundaram, M. R. Pulugurtha, and R. Tummala, "Interconnection materials, processes and tools for fine-pitch panel assembly of ultra-thin glass substrates," in *Electronic Components and Technology Conference (ECTC), 2015 IEEE 65th*, IEEE, pp. 475–483, ISBN: 1479986097.

[38] O. Suzuki and S. Kawamoto, "Development of low stress no-flow underfill for flip-chip application," in *Electronic Components and Technology Conference*, vol. 55, IEEE; 1999, p. 185, ISBN: 0569-5503.

[39] L. Wang, C. G. Woychik, G. Gao, S. McGrath, H. Shen, E. Tosaya, and S. Arkalgud, "Assembly and scaling challenges for 2.5D IC," in *International Symposium on Microelectronics*, International Microelectronics Assembly and Packaging Society, vol. 2014, 2014, pp. 000 606–000 611.

[40] D. B. Rao and M Prakash, "Effect of substrate warpage on the second level assembly of advanced plastic ball grid array (PBGA) packages," in *Electronics Manufacturing Technology Symposium, 1997., Twenty-First IEEE/CPMT International*, IEEE, pp. 439–446, ISBN: 0780339290.

[41] K. Murayama, M. Aizawa, K. Hara, M. Sunohara, K. Miyairi, K. Mori, J. Charbonnier, M. Assous, J.-P. Bally, and G. Simon, "Warpage control of silicon interposer for

2.5 D package application," in *Electronic Components and Technology Conference (ECTC), 2013 IEEE 63rd*, IEEE, pp. 879–884, ISBN: 1479902330.

[42] B. Singh, T.-C. Huang, S. Kawamoto, V. Sundaram, R. Pulugurtha, V. Smet, and R. Tummala, "Demonstration of enhanced system-level reliability of ultra-thin BGA packages with circumferential polymer collars and doped solder alloys," in *Electronic Components and Technology Conference (ECTC), 2016 IEEE 66th*, IEEE, pp. 1377–1385, ISBN: 1509012044.

[43] *Technology options and their influence on routing for interposer-based memory processor integration*, https://www.3dincites.com/2015/01/technology-options-influence-routing-interposer-based-memory-processor-integration/.

[44] W. Zhang, D. Wu, B. Su, S. A. Hareb, Y. Lee, and B. P. Masterson, "The effect of underfill epoxy on warpage in flip-chip assemblies," *IEEE Transactions on Components, Packaging, and Manufacturing Technology: Part A*, vol. 21, no. 2, pp. 323–329, 1998.

[45] M.-Y. Tsai, Y.-C. Lin, C.-Y. Huang, and J.-D. Wu, "Thermal deformations and stresses of flip-chip BGA packages with low-and high-T/sub g/underfills," *IEEE transactions on electronics packaging manufacturing*, vol. 28, no. 4, pp. 328–337, 2005.

[46] J. Pyland, R. V. Pucha, and S. Sitararnan, "Thermomechanical reliability of underfilled BGA packages," *IEEE transactions on electronics packaging manufacturing*, vol. 25, no. 2, pp. 100–106, 2002.

[47] "Guidelines for accelerated reliability testing of surface mount solder attachments," *IPC-SM-785*, 1992.

[48] W. Engelmaier, "Surface mount solder joint long-term reliability: Design, testing, prediction," *Soldering & Surface Mount Technology*, vol. 1, no. 1, pp. 14–22, 1989.

[49] W. Engelmaier, "Solder creep-fatigue model parameters for sac & snag lead-free solder joint reliability estimation," in *Proceedings from IPC Midwest Conference and Exhibition*, 2009, pp. 1–44.

[50] B. Richards, C. Hunt, and C. Levoguer, *An analysis of the current status of lead-free soldering*. Great Britain, Department of Trade and Industry, 1999.

[51] G. Subbarayan, *A systematic approach for selection of best lead-free printed circuit board (PCB) surface finish*. State University of New York at Binghamton, 2007.

[52] M. A. Hopcroft, W. D. Nix, and T. W. Kenny, "What is the young's modulus of silicon?" *Journal of microelectromechanical systems*, vol. 19, no. 2, pp. 229–238, 2010.

[53] N. Bai, X. Chen, and H. Gao, "Simulation of uniaxial tensile properties for lead-free solders with modified anand model," *Materials & Design*, vol. 30, no. 1, pp. 122–128, 2009.

[54] S. R. McCann, V. Sundaram, R. R. Tummala, and S. K. Sitaraman, "Flip-chip on glass (FCOG) package for low warpage," in *Electronic Components and Technology Conference (ECTC), 2014 IEEE 64th*, IEEE, pp. 2189–2193, ISBN: 1479924075.

[55] P. S. Theocaris, "Moire topography of curved surfaces," *Experimental Mechanics*, vol. 7, no. 7, pp. 289–296, 1967.

[56] H. Ding, R. E. Powell, C. R. Hanna, and I. C. Ume, "Warpage measurement comparison using shadow moire and projection moire methods," *IEEE Transactions on Components and Packaging Technologies*, vol. 25, no. 4, pp. 714–721, 2002.

[57] K Becker and J Wilde, "Applying Anand model to represent the viscoplastic deformation behavior of solder alloys," 2001.