ABSTRACT

Title of Dissertation: ADVANCED STATISTICAL ANALYSIS FOR TAIL-END PROBABILITY PREDICTION AND PERFORMANCE RESPONSE CALCULATION OF SEMICONDUCTOR PACKAGING PRODUCTS WITH A LARGE NUMBER OF INPUT VARIABLES

Hsiu-Ping Wei, Doctor of Philosophy, 2018

Dissertation directed by: Professor Bongtae Han, Department of Mechanical Engineering

Stochastic reliability modeling capabilities are developed and implemented for semiconductor packaging problems with a very large number of input variables (> 10 input variables). The capabilities are aimed at three critical areas in the semiconductor packaging product development: (1) prediction of tail-end probability (i.e., assembly yield loss) by advanced uncertainty propagation (UP) analyses, (2) determination of the statistical distributions of unknown design and/or manufacturing parameters by advanced statistical model calibrations, and (3) determination of the performance response of high-dimensional problems by developing an advanced metamodeling scheme.

In the first part, a comprehensive stochastic model is proposed and implemented to predict package-on-package (PoP) stacking yield loss based on non-contact open. The model takes into account all pad locations at the stacking interface while considering the statistical variations of warpages as well as solder ball and joint heights. The goal is achieved by employing (1) advanced approximate integration-based approach, called eigenvector dimension reduction (EDR) method,

for the UP analysis; (2) the stress-strength interference (SSI), and (3) the union of events. The proposed approach is capable of handling the number of input variables much larger than that has been conceived as the practical limit of the UP analysis. The model can be used effectively to control the input uncertainties, and thus to achieve a yield goal for a given set of PoP designs.

In the second part, the unknown statistical distributions of two effective elastic properties of Sn-3.0Ag-0.5Cu solder joint of leadless chip resistors (LCRs), induced by an assembly condition, are determined by the advanced statistical model calibration. The UP analysis also utilizes the EDR method, which allows to take into account the statistical variations of six additional known input variables, including die thickness, solder joint height, termination length, and thickness and elastic moduli of a printed circuit board. The cyclic bending test results of LCR assemblies are used in conjunction with the maximum likelihood metric to obtain the statistical distributions of the effective properties. The cycles-to-failure distribution of the identical LCR assemblies subjected to a different loading level is predicted accurately by the calibrated model, which corroborates the validity of the proposed approach.

In the third part, an advanced metamodeling scheme, called partitioned bivariate Cut-high dimensional model representation (PB Cut-HDMR), is developed to consider the statistical correlation among input variables and to further reduce the computational burden encountered for high-dimensional problems without compromising accuracy. The statistical correlation is handled by eigen-decomposition of a covariance matrix. The latter is achieved by the HDMR-factorial design (HDMR-FD) hybrid method. The validity of the proposed scheme is verified by comparing the performance of the proposed scheme with the full bivariate Cut-HDMR. The proposed scheme is implemented successfully to construct an accurate metamodel for a problem with 12 input variables among which 2 pairs are correlated.

# ADVANCED STATISTICAL ANALYSIS FOR TAIL-END PROBABILITY PREDICTION AND PERFORMANCE RESPONSE CALCULATION OF SEMICONDUCTOR PACKAGING PRODUCTS WITH A LARGE NUMBER OF INPUT VARIABLES

By

Hsiu-Ping Wei

Dissertation submitted to the Faculty of the Graduate School of the University of Maryland, College Park, in partial fulfillment of the requirements for the degree of Doctor of Philosophy
2018

**Advisory Committee**

Professor Bongtae Han, Chair and Advisor

Professor F. Patrick McCluskey

Professor Jin-Oh Hahn

Dr. Michael Osterman

Professor Gang-Len Chang, Dean’s Representative

© Copyright by

Hsiu-Ping Wei

2018

# Dedication

To

*My grandmother Mrs. Min Chen Wei, my father Mr. C.Y. Wei, my mother Mrs. Betty Hsu and my brothers Li-Fan Wei, Zhi-Xuan Wei and Liang-Tian Wei*

*for all their support and love; and*

*My beloved wife and daughter, An-Chi Chen and Zoey, Ruo-Shi, Wei,*

*for her persistence, love and faith in me*

Acknowledgement

First and foremost, I would like to express my deepest and sincere gratitude to my advisor Prof. Bongtae Han for his guidance and support. I enjoy so much to work with and learn from Prof. Han with wonderful teach-student trust relationship. His wise advices on logical thinking, critical experiment, asking "so what" after "why", will accompany me in the rest of my life. Words could never fully express my gratitude. All I ever wanted to do was duplicate it 10 times over just to show my appreciation. Thank you, Prof. Han!

I would like to thank Prof. Abhijit Dasgupta, Prof. F. Patrick McCluskey, Professor Jin-Oh Hahn, Dr. Michael Osterman, and Professor Gang-Len Chang for serving in my dissertation committee. I gratefully acknowledge their critical comments in the completion of this dissertation.

I am lucky to work with amazing LOMSS members and very grateful for all their support: Dr. Oh, Yong, Dae-Suk, Kenny, Bulong, Byung, Hyun-Seop, Jack, Ryan and Artur. I appreciate all your support and help. I am also especially grateful to Prof. Byeng Dong Youn in Seoul National University and Prof. Chao Hu in Iowa State University for teaching me all the knowledge about design under uncertainty and providing valuable inputs into my dissertation. Furthermore, I thank the Taiwanese students who joined University of Maryland at the same time with me, Che-Wei Liu, Chien-Lun Lan, Wan Ting Liao, Michelle Lin, I-Lin Liu, and Shane Hsu for their support and encouragement.

Finally, my family! All of this would never come true without the endless supports from my parents, parents-in-law, grandmother, and brothers. I also would like to thank my beloved daughter, Zoey, for bringing great joy to me. My deepest gratitude goes to my beloved wife, An-Chi Chen, for her sacrifice, support and love. I love you all!

# Table of Contents

**Table of Contents** ..... iv

List of Tables ..... viii

List of Figures ..... x

Chapter 1. Introduction ..... 1

1.1 Motivation and Objectives ..... 1

1.2 Organization of the Dissertation ..... 3

Chapter 2. Stacking Yield Loss Prediction of Package-on-Package Assembly with a Large Number of Input Variables ..... 6

2.1. Assembly Yield Prediction Of Plastically Encapsulated Packages With A Large Number Of Manufacturing Variables By Advanced Approximate Integration Method ..... 6

2.1.1. Introduction ..... 6

2.1.2. Eigenvector Dimension Reduction Method ..... 9

2.1.3. Assembly Yield Loss Prediction of Thin Flat Ball Grid Array ..... 16

2.1.3.1. Package Description ..... 16

2.1.3.2. Numerical Analysis: Warpage Prediction ..... 18

2.1.3.3. Uncertainty Propagation Analysis by EDR ..... 21

2.1.3.3.1. Input Random Variables ..... 21

2.1.3.3.2. Eigenvector Sampling and Sample Points ..... 24

2.1.3.3.3. Statistical Moments ..... 27

2.1.3.3.4. PDF Estimation ..... 29

2.1.4. Validity of the Proposed Approach ..... 31

2.1.5. Conclusion ..... 36

2.2. Stacking Yield Prediction of Package-on-Package Assembly Using Advanced Uncertainty Propagation Analysis: Part I Stochastic Model Development ..... 38

2.2.1. Introduction ..... 38

2.2.2. Conditions for Non-Contact Open ..... 40

2.2.3. Stacking Yield Loss Prediction Model ..... 43

2.2.3.1. Stacking Yield Loss Prediction Model for Single Pad ..... 44

2.2.3.2. Stacking Yield Loss Prediction for Multiple Pads ..... 49

2.2.3.3. Stacking Yield Loss Prediction Model ..... 52

2.2.4. Conclusion ..... 54

2.3. Stacking Yield Prediction of Package-on-Package Assembly Using Uncertainty Propagation Analysis: Part II Implementation of Stochastic Model ..... 55

2.3.1. Introduction ..... 55

2.3.2. Probability Density Functions of Warpages ..... 55

2.3.2.1. Warpage Prediction ..... 56

2.3.2.2. Uncertainty Propagation Analysis for Warpage PDFs ..... 59

2.3.3. Probability Density Functions of Solder Balls and Joint Heights ..... 66

2.3.3.1. Prediction of Solder Ball Heights and Joint Heights at Corner Pad ..... 66

2.3.3.2. Uncertainty Propagation Analysis for Solder Ball Heights and Joint Heights at Corner Pad ..... 68

2.3.4. Stacking Yield Loss Prediction ... 71

2.3.4.1. Stacking Yield Loss Prediction using a Single MCS Run ... 71

2.3.4.2. True Stacking Yield Loss Estimation using Multiple MCS Runs ... 75

2.3.5. Conclusion ... 77

Chapter 3. Advanced Statistical Model Calibration to Determine Manufacturing-induced Variations of Effective Elastic Properties of SAC Solder Joints in Leadless Chip Resistor Assemblies ... 78

3.1. Introduction ... 78

3.2. Background: Statistical Model Calibration ... 80

3.3. Implementation Using Vibration-Induced Solder Fatigue Failure ... 82

3.3.1. Setup of Four-Point Bend Test ... 82

3.3.2. Physics-based Solder Joint Fatigue Life Modeling ... 84

3.3.3. Statistical Model Calibration ... 86

3.3.3.1. Known and Unknown Input Variables ... 86

3.3.3.2. Calibration Results and Discussion ... 87

3.3.4. Discussions ... 90

3.3.4.1. Modeling Runs ... 90

3.3.4.2. Reliability Prediction under Usage Condition ... 91

3.3.5. Conclusions ... 92

Chapter 4. An Advanced HDMR-based Metamodeling Technique for Semiconductor Packages with A Large Number of Input Variables ... 93

4.1. Introduction ........................................................................................................ 93

4.2. Background: Cut-HDMR Based Methods........................................................................ 96

4.3. Proposed Partitioned Bivariate Cut-HDMR (PB Cut-HDMR) ...................................... 101

4.3.1. Transformation of Correlated Input Variables to Uncorrelated Input Variables 101

4.3.2. Ranking and Partitioning of Bivariate Terms using HDMR-Factorial Design Hybrid Method.................................................................................................. 103

4.3.3. Procedure of PB Cut-HDMR ................................................................ 107

4.3.4. Validity Check of PB Cut-HDMR ....................................................... 112

4.3.4.1. Performance Metric ........................................................................ 112

4.3.4.2. Numerical Examples ....................................................................... 113

4.4. Implementation of PB Cut-HDMR: Warpage Prediction of Thin Flat Ball Grid Array . 115

4.4.1. Problem Description ............................................................................ 115

4.4.2. Metamodel Construction by PB Cut-HDMR....................................... 118

4.5. Conclusion .................................................................................................... 121

Chapter 5. Contributions and Future Works........................................................... 123

5.1. Dissertation Contributions ............................................................................ 123

5.2. Future Works ................................................................................................ 124

Reference 128

**LIST OF TABLES**


Table 1 Material properties of TFBGA ........ 19

Table 2 Dimensions of TFBGA ........ 19

Table 3 Input variables........ 22

Table 4 Package warpage simulation results at sample points ........ 27

Table 5 First four statistical moments........ 29

Table 6 Coefficients of Pearson system........ 30

Table 7 Predicted statistical moments and yield loss by MCS and EDR ........ 33

Table 8 Effect of sample size on MCS predictions........ 34

Table 9 Material properties of top TFBGA and bottom fcBGA........ 57

Table 10 Dimensions of top TFBGA and bottom fcBGA ........ 57

Table 11 Input variables of bottom fcBGA ........ 59

Table 12 Eigenvalues and eigenvectors of the covariance matrix........ 61

Table 13 Locations and predicted warpages of sample points along the 3rd eigenvector and the 11th eigenvector........ 61

Table 14 Locations, weights, and warpage values of integration points along the 3rd eigenvector and the 11th eigenvector ........ 62

Table 15 The 1st to the 4th statistical moments and coefficients of the stabilized Pearson distribution of the warpage at the outmost corner pad........ 63

Table 16 Input variables of top solder ball ........ 69

Table 17 Input variables of bottom solder ball ........ 69

Table 18 Predicted statistical moments and the corresponding coefficients of the stabilized Pearson distribution for top and bottom solder balls as well as solder joint........ 70

Table 19 Rank of probability of non-contact open for the *j*th pad........ 76

Table 20 Material Properties ........ 85

Table 21 Input Variables ........ 86

Table 22 Correlation coefficient of correlated input variables ........ 86

Table 23 $z_{i_main}$ and corresponding main effects ........ 108

Table 24 Illustration of the results of the pair having interaction effect and the pair having no interaction ........ 109

Table 25 First order and second order sensitivity indices ........ 109

Table 26 Comparison of the prediction results of univariate Cut-HDMR, bivariate Cut-HDMR and PB Cut-HDMR ........ 112

Table 27 Test functions ........ 114

Table 28 Comparison of performance metrics for full bivariate Cut-HDMR and PB Cur-HDMR ........ 115

Table 29 Material properties ........ 116

Table 30 Dimensions ........ 117

Table 31 Input variables and design space ........ 118

Table 32 Sensitivity indices of the TFBGA warpage prediction ........ 120

**LIST OF FIGURES**

Figure 1. Illustration of a yield loss (tail-end probability) ........................................................ 7

Figure 2. Schematic illustration of numerical integration ........................................................ 13

Figure 3. Schematic illustration of eigenvectors of two random variables ................................ 14

Figure 4. Schematic illustration of locations of sample points: (a) three and (b) five sample-point schemes ........................................................................................................... 15

Figure 5. TFBGA package: (a) cross-sectional view and (b) bottom view showing the measuring zone ........................................................................................................... 17

Figure 6. Package warpage definition and sign convention: (a) convex and (b) concave ........... 17

Figure 7. Package warpage definition and sign convention: (a) convex and (b) concave ........... 18

Figure 8. Temperature dependent Young's modulus of DAF ................................................... 19

Figure 9. Master curves of bulk modulus and shear modulus of EMC [1] ................................ 19

Figure 10. Completed thermal excursion ........................................................................ 20

Figure 11. (a) Deformed configuration of TFBGA package at 260 °C (20x magnification) and (b) package warpage calculated from on reference plane ............................................... 21

Figure 12. MCS sweeping analysis results for correlation coefficient of EMC thickness and PCB thickness ........................................................................................................... 23

Figure 13. Marginal joint PDF of package length and width (uncorrelated case) and locations of sampling points ........................................................................................... 26

Figure 14. Marginal joint PDF and locations of sampling points for correlated input variables: (a) EMC and PCB thickness and (b) EMC CTE below $T_g$ and above $T_g$ ........................ 26

Figure 15. Integration points and weights for 1-D numerical integration: (a) $x_{11}$ and (b) $x_3$ ..... 28

Figure 16. PDFs of and assembly yield loss predicted by two schemes of EDR: (a) entire PDF and (b) enlarged view of the tail-end region ................................................................. 31

Figure 17. Yield loss distribution predicted by MCS with the sample size of 1,000,000 ........... 34

Figure 18. Number of required modeling runs for tail-end probability prediction...................... 36

Figure 19. Schematic illustration of a typical Package-on-Package assembly ............................ 38

Figure 20. Schematic illustration of non-contact open ............................................................... 39

Figure 21. Three gap formation scenarios of Case I where the maximum warpage difference occurs at the corner of the package: (a) Scenario-1, (b) Scenario-2, and (c) Scenario-3 ....... 41

Figure 22. Gap formation scenario of Case II (Scenario-4) where the maximum warpage difference occurs at the center of the package ........................................................... 41

Figure 23. Statistical interference between the PDFs of load $f_L(L)$, and strength $f_S(S)$ ........ 45

Figure 24. Schematic illustration of the union of three events .................................................. 53

Figure 25. Details of the FEA model of bottom fcBGA package................................................ 56

Figure 26. Deformed configurations (5x magnification) of (a) the top and (b) the bottom packages with the nominal design parameters at the peak reflow temperature.......................... 58

Figure 27. PDFs of warpage at the outmost corner pad............................................................. 64

Figure 28. Statistical correlations of the warpages at the $j^{th}$ pad and the outmost corner pad: (a) pad locations, (b) top package, and (c) bottom package ............................................ 65

Figure 29. Cross-sectional view of a PoP using the BoB approach just prior to reflow ............. 66

Figure 30. Solder balls after (a) the first reflow process and (b) the second reflow during stacking ........................................................................................................................ 67

Figure 31. Shape prediction results of (a) the top and (b) bottom solder balls after the 1st reflow; and (c) the solder joint after the 2nd reflow ................................................................ 68

Figure 32. (a) PDF of the top solder ball height, (b) PDF of the bottom solder ball height, and (c) PDF of the solder joint height ......................................................................... 70

Figure 33. Procedure of stacking yield prediction by MCS........................................................ 72

Figure 34. Interference of PDFs of load and strength for the 109th pad ($j$ = 109) for a single MCS run: (a) scenario-1 and (b) scenario-4 ........ 73

Figure 35. (a) PDF of gap at the 109th pad and (b) the enlarged view of the tail-end marked by the red box in (a) ........ 74

Figure 36. Regions of ball pads having non-contact opens for all of the MCS runs ........ 76

Figure 37. Durability vs. reliability ........ 79

Figure 38. Procedure of statistical model calibration ........ 81

Figure 39. Illustration of likelihood function with the maximum and minimum likelihood values ........ 82

Figure 40. Detail of a test coupon and a single leadless chip resistor assembly ........ 83

Figure 41. Test coupon mounted on 4 point bend fixture ........ 83

Figure 42. Accelerated life testing results of LCR assemblies subjected to strain levels of 2000 and 1200 με. (a) Weibull plots and (b) Cross section of failed assembly ........ 84

Figure 43. Damage analysis (a) finite element model (b) von Mises stress distribution in the solder joint ........ 85

Figure 44. Negative log-likelihood (NLL) values as a function of iteration number ........ 87

Figure 45. Initial guesses and calibrated values of (a) Young's modulus and (b) Poisson's ratio of SAC305 solder ........ 88

Figure 46. PDFs of cycles-to-failure before and after calibration are compared with the histogram of experimental data obtained at 2000 με strain level ........ 89

Figure 47. PDF of cycles-to-failure predicted by the calibrated model is compared with the histogram of experimental data obtained at 1200 με strain level ........ 89

Figure 48. PDF of cycles-to-failure under a portable device usage condition, predicted by the calibrated model ........ 91

Figure 49. Illustration of the independence requirement of input variables on the sampling strategy ........ 95

Figure 50. Illustration of Cut-HDMR using a 2-D function (a) underlying 2-D function and zeroth-order effect, (b) main effect of $x_1$, (c) main effect of $x_2$, and (d) interaction effect. ... 99

Figure 51. Illustration of HDMR-FD hybrid method ........ 107

Figure 52. Comparison of true response and approximated response surface with 2-D MLS interpolation for significant bivariate terms........ 110

Figure 53. Comparison of true response and approximated response surface from superposition of the corresponding main effects for minor bivariate terms........ 111

Figure 54. Configuration of TFBGA assembly and details of the FEA model ........ 116

Figure 55. Transformation of input variable directions and design space change after eigen-decompostion of covariance matrix. (a) $x_3 x_4$ pair and (b) $x_9 x_{10}$ pair. ........ 119

Figure 56. Warpages at the sample points along each input variable........ 119

Figure 57. Histogram of errors at the sample points for validity check ........ 121

# CHAPTER 1. INTRODUCTION

## 1.1 Motivation and Objectives

Due to the ever-increasing demand in product development cycle time reduction of semiconductor packages, the performance evaluation and reliability assessment must be cost and time effective to maintain a competitive edge. The computer-aided engineering (CAE) tools, such as the finite element analysis (FEA), have been used extensively for comparing competitive designs. Tremendous efforts have been made to provide an accurate deterministic computational model. In reality, the reliability performance responses of a semiconductor package (e.g. package warpage, solder joint fatigue life, cycles to failure of drop test, etc.) show statistical variations due to inherent manufacturing variability or design variations. Understanding the statistical distribution of the performance responses is critical to yield prediction, warranty period determination, and design for reliability or design for yield.

To achieve this goal, statistical distributions of input variables, including material properties, dimensions, loading conditions, etc, have to be considered. These inherent variations can be categorized into "known input variables" and "unknown input variables". The "known input variables" are the variables whose statistical distributions are known or can be measured. The variables are called "unknown" when their statistical distributions are extremely difficult or impractical to be obtained experimentally.

When all the input variables are known, the uncertainty propagation (UP) analysis is typically used to determine the statistical distributions of the performance responses, which enables the intrinsically deterministic computational model to characterize the output distribution in the presence of input uncertainties. As the most widely used configuration in advanced mobile applications now days, the stacking yield loss of package-on-package (PoP) assembly is one of the major concerns and falls into this category [2-4]. The yield loss prediction, however, is a problems

related to tail-end probability with a large number of input variables and nonlinear material properties. For these problems, the conventional uncertainty propagation analysis methods such as the Monte Carlo simulation (MCS) and the response surface method (RSM), etc. become impractical due to excessive computational cost. The first objective of this dissertation is to implement and develop an advanced stochastic yield loss prediction model to for more accurate PoP stacking yield loss prediction.

In addition to the problems involving only known input variables, many semiconductor packaging applications contain unknown input variables. For example, the widely used Sn-rich solders, such as SnAgCu (SAC) alloys, have been known to have large variations in grain sizes and orientations, intermetallic compound (IMC) sizes, and distributions of anisotropic Sn crystals [5]. These microstructural variations occur even under the same assembly condition, which results in inherent mechanical property variations of solder joints; not only package-to-package variations but also joint-to-joint variations in the same package [6]. The statistical distributions of mechanical properties of Sn-rich solder materials, however, are difficult to obtain from direct measurements. In this dissertation, the second objective is to inversely determine the unknown statistical distributions of two effective elastic properties of Sn-3.0Ag-0.5Cu (SAC305) solder joint of leadless chip resistors (LCRs) assemblies by implementing the advanced statistical model calibration.

Unlike the previous two problems focusing on the full statistical distributions of performance responses, what combinations of input variables can cause the performance response out of desired specifications is important for design-for-reliability. Conventionally, this study was done by Design of Experiment (DoE) with numerical modeling. When the numerical model becomes computationally intensive and the number of design configurations increases, the metamodel (or metamodeling) techniques are usually used to reduce the computational burden.

The metamodel is also known as response surface method (RSM), surrogate model, or reduced-order model, which utilizes simple analytical models to approximate the input/output relationship of computationally intensive numerical models. However, the current practice of metamodeling techniques has been limited to problems with only a few input factors or a few levels because of the number of modeling runs to construct a metamodel raise rapidly as the number of input variables increases. In addition, the interaction effects of input variables on the performance response are always present when multiple input factors are considered, which makes the problem even more challenging. Thus, the third objective is to propose an advanced metamodeling technique in response to the need.

## 1.2 Organization of the Dissertation

This first chapter describes the motivations behind this work and three main objectives sought to be accomplished by it. Each of the three objectives is addressed in the subsequent chapters.

Chapter 2 presents a comprehensive stochastic model is proposed to predict PoP stacking yield loss based on non-contact open taking into account all pad locations at the stacking interface while considering the statistical variations of five critical geometrical quantities, including warpages as well as the heights of solder balls of the top and bottom package and solder joint at the corner. The goal is achieved by employing three statistical approaches: including (1) advanced uncertainty propagation (UP) by approximate integration-based approaches, (2) stress-strength interference (SSI), and (3) the union of events. The advanced approximate integration scheme called eigenvector dimension reduction (EDR) method is first implemented to predict the assembly yield loss cause by the warpage of a plastically encapsulated package (i.e., the top package of PoP). The probability density function of the warpage at reflow temperature was obtained using only 25 modeling runs for 12 input variables. The results prove that the EDR provides the numerical

efficiency required for the tail-end probability prediction of manufacturing problems with a large number of input variables, while maintaining high accuracy. This section has been published in the *Microelectronics Reliability* [7]. Then, the theoretical development about the integration of three statistical approaches to form the proposed stochastic model is presented. The model takes into account the statistical dependency of hundreds PDFs of the five critical quantities on every pad at the stacking interface. Implementation of the proposed model to a typical PoP is presented in a companion paper. These two sections have be has been prepared to be submitted as a journal paper.

Chapter 3 is devoted to the determination of the unknown statistical distributions of two effective elastic properties of SAC305 solder joint of LCR assemblies by the advanced statistical model calibration while considering the statistical variations of serval other known input variables including die thickness, solder joint height, termination length, and thickness and elastic moduli of a printed circuit board (PCB). The background of the statistical model calibration is described first. The cyclic bending test results of the LCR assemblies are followed, and the results are subsequently used to obtain the statistical distributions of the effective elastic properties of SAC305 solder. Validity of the calibrated model is corroborated by comparing the predicted probability density function (PDF) of cycles-to-failure of the identical LCR assemblies subjected to a different loading level with the cycles to failure distribution of actual testing data. This chapter was submitted to *IEEE Transactions on Components, Packaging and Manufacturing Technology* on June 2018.

Chapter 4 focuses on development of an advanced metamodeling technique, called partitioned bivariate Cut-high dimensional model representation (PB Cut-HDMR), to tackle two major challenges associated with the metamodel construction for a semiconductor package problems, namely (1) the computational burden caused by a large number of input variables with

non-negligible interaction effects and (2) the statistical correlation among input variables. The current Cut-HDMR based methods are reviewed. After the introduction of the state of art, the development of PB Cut-HDMR is presented. Then, the proposed approach is implemented to construct the metamodel for the warpage prediction of a thin flat ball grid array (TFBGA) assembly to check the accuracy and efficiency of the proposed method in real application. This chapter has been prepared to be submitted as a journal paper.

Chapter 5 contains a summary of the contributions and a discussion of future works that can be extended from this work.

# CHAPTER 2. STACKING YIELD LOSS PREDICTION OF PACKAGE-ON-PACKAGE ASSEMBLY WITH A LARGE NUMBER OF INPUT VARIABLES

## 2.1. Assembly Yield Prediction Of Plastically Encapsulated Packages With A Large Number Of Manufacturing Variables By Advanced Approximate Integration Method

### 2.1.1. Introduction

Epoxy molding compound (EMC) has been used extensively as a protection layer in various semiconductor packaging components. The mismatch of coefficient of thermal expansion (CTE) causes the warpage of components after molding, which is one of the most critical issues to board assembly yield. The warpage issue has become more critical as Package-on-Package (PoP) and fan-out wafer level package (FO-WLP) are widely adopted for portable devices.

The computer-aided engineering (CAE) tools, such as the finite element method (FEM), have been used extensively to predict the warpage. Typically, the CAE tools provide deterministic outputs, which establish quantitative relationships between the system response (i.e., warpage) and the input parameters such as geometries, material properties, process and/or environmental conditions, etc. The deterministic approaches have been proven effective for comparing competitive designs. In reality, the package warpage behavior shows statistical variations (or probability distributions) due to inherent manufacturing variabilities. The probabilistic aspect should be incorporated in prediction if the assembly yield is to be predicted.

The yield loss is in general a small probability event (i.e., tail-end probability) [2, 3, 8], especially for the large production volume. In many cases, even 0.1 % yield loss would cause a significant profit loss. Based on the Six Sigma concept, the target is often to control the yield loss within 3 to 6 sigma, i.e., 6.67% to 3.4 ppm [9].

Figure 1 shows a schematic illustration of the tail-end probability, where the statistical property of system performance (e.g., warpage) is represented by a probability density function (PDF). When a component has the performance exceeding or falling behind a certain specification,

it cannot be processed further and is regarded as a failure. The probability of all possible failure, i.e., yield loss, is the area under the PDF where the performance does not satisfy the specification.

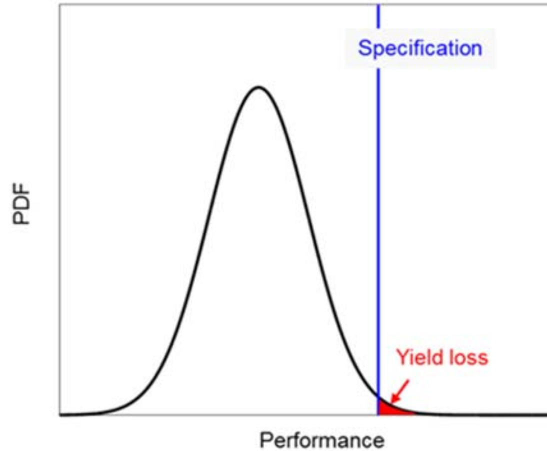



Figure 1. Illustration of a yield loss (tail-end probability)

A technical approach critically required for the yield loss prediction is the uncertainty propagation analysis, which enables the intrinsically deterministic computational model to characterize the output distribution in the presence of input uncertainties. The most popular uncertainty propagation methods are “random sampling method” and “response surface method (RSM)”. When they are applied to complex manufacturing problems with a large number of input variables, however, they become impractical due to their own limitations.

Due to its random nature, the failure probability estimated from the random sampling method, e.g., Monte Carlo simulation (MCS), exhibits statistical variations [10]. The variations can be substantial when the tail-end probability is to be predicted. In order to ensure that the tail-end probability prediction falls within the specified accuracy tolerance, an extremely large number of model computations is required. This computational burden makes the random sampling impractical for the cases that require complex nonlinear computational models (e.g., viscoelastic analysis required for warpage prediction of plastically encapsulated components) [11].

The RSM has also been widely used in conjunction with the MCS [12, 13] to reduce the computational burden. The RSM relies on Design of Experiments (DOE) to build computationally inexpensive mathematical response surface models, which can be used for the direct MCS. Two commonly used types of DOE are the Full Factorial Design (FFD) [14-16] and the Central Composite Design (CCD) [17-19]. Although the CCD can reduce the sample size of the FFD substantially, both types cannot avoid the challenge known as the curse of dimensionality (i.e., the computational costs increase exponentially as the number of random input variables increases). Due to this inherent limitation, the RSM has been applied to the designs with only a few input variables.

Another method for the uncertainty propagation analysis is "approximate integration scheme". The scheme calculates the statistical moments of the output response by performing a multi-dimensional integration. Seo and Kwak proposed a numerical algorithm to perform the integration [20]. The algorithm also suffered from the curse of dimensionality as the FFD was used to select integration points. Rahman and Xu proposed the univariate dimension-reduction (UDR) method to cope with the curse of dimensionality [21]. With the method, a multi-dimensional integration is transformed into a series of one-dimensional integrations, and thus the computational cost increases only additively with the increased number of input variables. This additive increase makes the method attractive to the problems with a large number of input variables.

In a typical UDR implementation, however, a large number of numerical integration points are still required to ensure the accuracy of each one-dimensional integration result. For a large number of input variables, the method also can be computationally expensive. Youn et al. developed a method called "eigenvector dimension-reduction (EDR)" method [22] to relax the requirement of the UDR method. In the EDR method, the eigenvector sampling scheme was

proposed to select a few sample points along the eigenvectors of the covariance matrix of the input variables, and the stepwise moving least square (SMLS) was implemented to interpolate and extrapolate the numerical integration points. As a result, the accuracy of statistical moment estimation by EDR remained virtually unaffected although the number of simulations was reduced substantially.

In this paper, the EDR method is implemented to predict the assembly yield of a plastically encapsulated package. A total of 12 manufacturing input variables are considered during the yield prediction, which is based on the JEDEC reflow flatness requirements. Section 2.1.2 provides a brief introduction of the EDR method. In Section 2.1.3, the details of an EDR implementation are described. The accuracy of the yield prediction is verified by the direct MCS in Section 2.1.4. Section 2.1.5 concludes the paper.

### 2.1.2. Eigenvector Dimension Reduction Method

The eigenvector dimension-reduction (EDR) method estimates the complete probability density function (PDF) of a system response by (1) calculating the statistical moments and (2) constructing the corresponding PDF using the probability estimation methods.

The statistical moments are the characteristics of a distribution. The 1st moment, $\mu$, is the mean, which represents the central tendency of the distribution, and the 2nd moment is the standard deviation, $\sigma$, which represents the spread of the distribution. The 3rd and 4th moments are skewness, $\beta_1$, and kurtosis, $\beta_2$, which indicate the symmetry and the peakedness of the distribution, respectively. The $m^{th}$-order statistical moment of a system response is defined as

$$E\left\{\left[f(y)\right]^m\right\} \equiv \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \left\{y\left(x_1,\ldots,x_N\right)\right\}^m f\left(x_1,\ldots,x_N\right) dx_1 \cdots dx_N \tag{1}$$

where $E(\cdot)$ is the expectation operator; $f(y)$ denotes the PDF of the performance response, $y$; $y(x_1,\ldots,x_N)$ is the performance response function, i.e., the deterministic relationship of a certain

performance response value, $y_i$, with a given set of input values $\left(x_{1i},...,x_{Ni}\right)$; and $f\left(x_1,...,x_N\right)$ is the joint probability density function of input variables.

To tackle the mathematical challenge associated with the multidimensional integration in Eq. $E\left\{\left[f\left(y\right)\right]^m\right\} \equiv \int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}\left\{y\left(x_1,...,x_N\right)\right\}^m f\left(x_1,...,x_N\right)dx_1\cdots dx_N$ (1), Rahman and Xu proposed the additive decomposition [21] to transform the multidimensional response function $y\left(x_1,...,x_N\right)$ into a series of one-dimensional functions. The approximated system response function, then, can be expressed as [21]:

$$y\left(x_1,...,x_N\right) \approx y_a\left(x_1,...,x_N\right) = \sum_{j=1}^{N} y\left(\mu_1,...,\mu_{j-1},x_j,\mu_{j+1},...,\mu_N\right) - \left(N-1\right)\cdot y\left(\mu_1,...,\mu_N\right) \qquad (2)$$

where $y_a\left(x_1,...,x_N\right)$ is the approximated system response function obtained by the additive decomposition, $\mu_j$ is the mean value of an input variable, $x_j$, $y\left(\mu_1,...,\mu_{j-1},x_j,\mu_{j+1},...,\mu_N\right)$ is the system response of the input variable, $x_j$, while the other input variables are kept as their respective mean values, and $y\left(\mu_1,...,\mu_N\right)$ is the system response with all input variables are fixed as their mean values.

Substituting Eq. (2) into Eq. (1) yields:

$$\begin{aligned} &E\left\{\left[y\left(x_1,...,x_N\right)\right]^m\right\} \cong E\left\{\left[y_a\left(x_1,...,x_N\right)\right]^m\right\} \\ &= E\left\{\left[\sum_{j=1}^{N} y\left(\mu_1,...,\mu_{j-1},x_j,\mu_{j+1},...,\mu_N\right) - \left(N-1\right)\cdot y\left(\mu_1,...,\mu_N\right)\right]^m\right\} \end{aligned} \qquad (3)$$

Using the binomial formula, the right-hand side of Eq. (3) can be rewritten as [21]:

$$\begin{aligned}
&E\left\{\left[\sum_{j=1}^{N} y\left(\mu_1,\dots,\mu_{j-1},x_j,\mu_{j+1},\dots,\mu_N\right)-(N-1)\cdot y\left(\mu_1,\dots,\mu_N\right)\right]^m\right\}\\
&=E\left\{\sum_{i=0}^{m}\frac{m!}{i!(m-i)!}\left[\sum_{j=1}^{N} y\left(\mu_1,\dots,\mu_{j-1},x_j,\mu_{j+1},\dots,\mu_N\right)\right]^i\left[-(N-1)\cdot y\left(\mu_1,\dots,\mu_N\right)\right]^{m-i}\right\}\\
&=\sum_{i=0}^{m}\frac{m!}{i!(m-i)!}E\left\{\left[\sum_{j=1}^{N} y\left(\mu_1,\dots,\mu_{j-1},x_j,\mu_{j+1},\dots,\mu_N\right)\right]^i\right\}\left[-(N-1)\cdot y\left(\mu_1,\dots,\mu_N\right)\right]^{m-i}
\end{aligned} \tag{4}$$

The recursive formula is further employed to simplify the expectation operation in Eq. (4), which yields [21]:

$$\begin{aligned}
&\sum_{i=0}^{m}\frac{m!}{i!(m-i)!}E\left\{\left[\sum_{j=1}^{N} y\left(\mu_1,\dots,\mu_{j-1},x_j,\mu_{j+1},\dots,\mu_N\right)\right]^i\right\}\left[-(N-1)\cdot y\left(\mu_1,\dots,\mu_N\right)\right]^{m-i}\\
&=\sum_{i=0}^{m}\frac{m!}{i!(m-i)!}S_N^i\left[-(N-1)\cdot y\left(\mu_1,\dots,\mu_N\right)\right]^{m-i}
\end{aligned} \tag{5}$$

where

$$\begin{cases}
S_1^i=\int_{-\infty}^{+\infty}\left\{y\left(x_1,\mu_2,\dots,x_N\right)\right\}^i f\left(x_1\mid x_2=\mu_2,\dots,x_N=\mu_N\right)dx_1\\
S_j^i=\sum_{k=0}^{i}\frac{i!}{k!(i-k)!}S_{j-1}^k\int_{-\infty}^{+\infty}\left\{y\left(\mu_1,\dots,\mu_{j-1},y_j,\mu_{j+1}\dots,\mu_N\right)\right\}^{i-k}\\
\qquad\times f\left(x_j\mid x_1=\mu_1,\dots,x_{j-1}=\mu_{j-1},x_{j+1}=\mu_{j+1},\dots,x_N=\mu_N\right)dx_j\\
S_N^i=\sum_{k=0}^{i}\frac{i!}{k!(i-k)!}S_{N-1}^k\int_{-\infty}^{+\infty}\left\{y\left(\mu_1,\dots,x_N\right)\right\}^{i-k} f\left(x_N\mid x_1=\mu_1,\dots,x_{N-1}=\mu_{N-1}\right)dx_N
\end{cases} \tag{6}$$

From Eqs. (1) to (6), a total of $m \times N$ one-dimensional integrations are needed to obtain the $m^{\text{th}}$-order statistical moment. These integrations in general can be done by numerical integration algorithms, which perform the calculations at the integration points, $x_{j,i}$, with weights, $w_{j,i}$, as [22]:

$$\begin{aligned}
&\int_{-\infty}^{+\infty} \left\{ y\left(\mu_1, \ldots, \mu_{j-1}, x, \mu_{j+1} \ldots, \mu_N\right) \right\}^r \\
&\qquad \times f\left(x_j \mid x_1 = \mu_1, \ldots, x_{j-1} = \mu_{j-1}, x_{j+1} = \mu_{j+1}, \ldots, x_N = \mu_N\right) dx_j \\
&\cong \sum_{i=1}^{n} w_{j,i} \left\{ y\left(\mu_1, \ldots, \mu_{j-1}, x_{j,i}, \mu_{j+1} \ldots, \mu_N\right) \right\}^r
\end{aligned} \tag{7}$$

where $x_{j,i}$ is the $i^{\text{th}}$ integration point of an input variable, $x_j$, $y\left(\mu_1, \ldots, \mu_{j-1}, x_{j,i}, \mu_{j+1} \ldots, \mu_N\right)$ is the system response at $x_{j,i}$, while the other input variables are kept as their respective mean values, and $w_{j,i}$ is the corresponding weight which approximates the area under the PDF of $x_j$ from $x_{j,i} - \left(x_{j,i} - x_{j,i-1}\right)/2$ to $x_{j,i} + \left(x_{j,i+1} - x_{j,i}\right)/2$, respectively.

Figure 2 illustrates the numerical integration with 5 integration points. The PDF of $x_j$ and the system response along $x_j$ are shown as the blue and red dashed curves, respectively. The products of these two curves are integrated along the variable, $x_j$. More specifically, the products of the responses and their corresponding weights (the bars) at the integration points (the black cross) are added to complete the numerical integration.

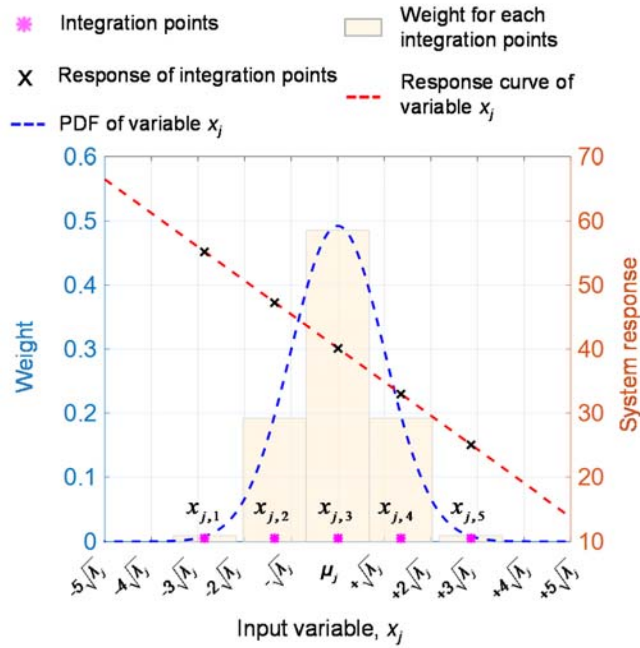



Figure 2. Schematic illustration of numerical integration

High accuracy of the 1-D numerical integration in Eq. (7) can be achieved by selecting numerous integration points, which can be computationally challenging for the applications with a large number of input variables, especially those that require nonlinear modeling. Two major improvements were made in the EDR method to reduce the number of simulations while maintaining the accuracy.

First, the eigenvector sampling scheme was proposed to handle the statistical correlation and variation of the input variables. By solving the eigenvalue problem of the covariance matrix, the eigenvalues and the corresponding eigenvectors are obtained. The eigenvector associated with the largest eigenvalue is the direction of the largest variation, wherein the square root of the eigenvalue is the standard deviation along this direction. The eigenvector associated with the second largest eigenvalue is the orthogonal direction with the next highest variation. The sample points are selected along the eigenvectors, and the simulations are conducted only at the sample

points. Eigenvectors of two random variables are illustrated in Figure 3, where the 1st and 2nd eigenvector directions are shown with a joint PDF.

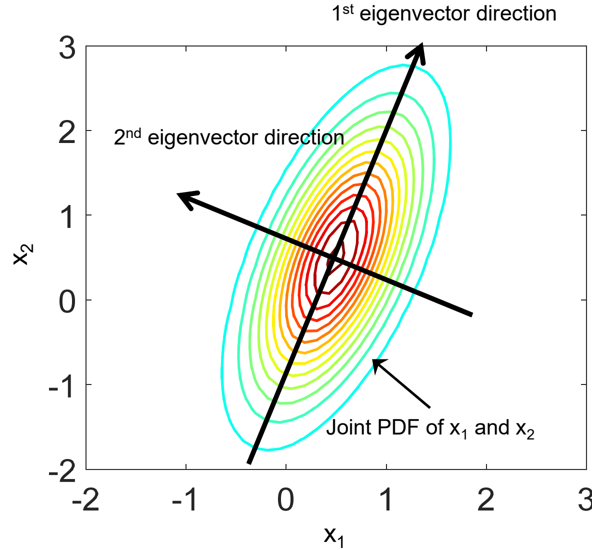



Figure 3. Schematic illustration of eigenvectors of two random variables

The three sample-point scheme is typically used in practice. The locations of the three sample points are mean and the mean ± 3 standard deviations, which are expressed as [22]:

$$\begin{aligned} {}^{0}\mathrm{V} &= \left[\mu_1, \ldots, \mu_N\right] \\ {}^{1}\mathrm{V}_i &= \left[\mu_1, \ldots, \mu_i - 3\sqrt{\lambda_i}, \ldots, \mu_N\right] \\ {}^{2}\mathrm{V}_i &= \left[\mu_1, \ldots, \mu_i + 3\sqrt{\lambda_i}, \ldots, \mu_N\right] \end{aligned} \tag{8}$$

where $\mu_i$ and $\lambda_i$ are the mean and eigenvalue along the $i^{\text{th}}$ eigenvector. The locations of sample points in Eq. (8) were suggested based on the parametric study [22]. Figure 4(a) illustrates the three sample-point scheme of an input variable $X_j$ following the normal distribution. It can be observed that the PDF values outside of the range of mean ± 3 standard deviations are very small (i.e., 0.27% of all the possible values of $X_j$), which suppresses their contribution during the numerical integration.

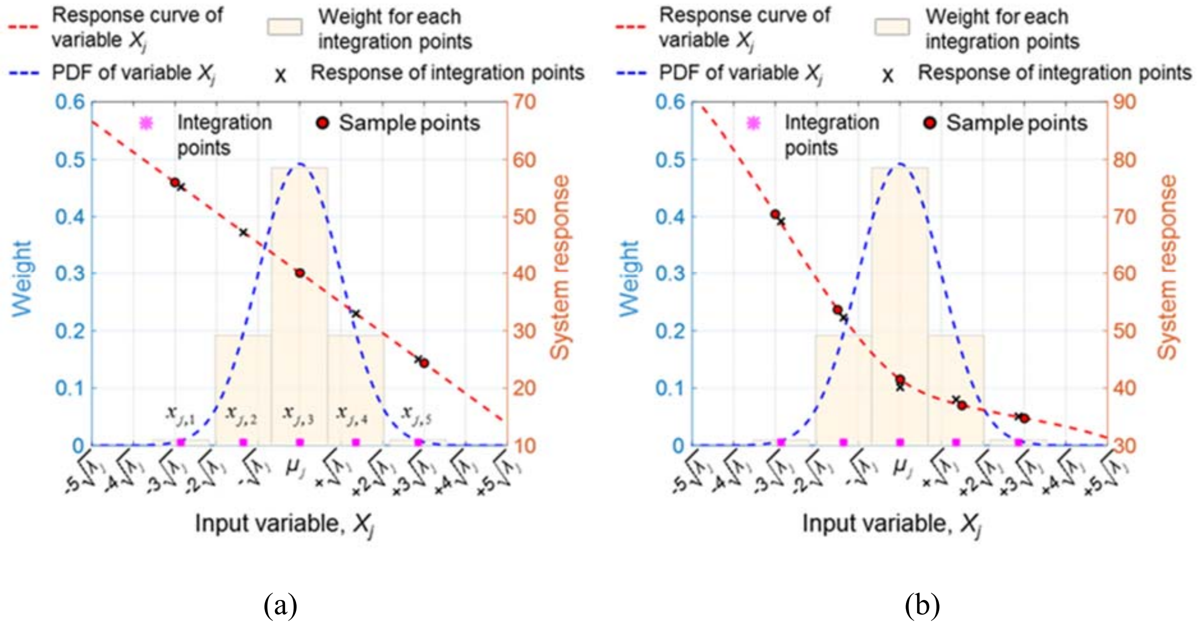



Figure 4. Schematic illustration of locations of sample points: (a) three and (b) five sample-point schemes

The above condition is no longer valid if the system response is highly nonlinear within the range of mean ± 3 standard deviations. More sample points are needed to capture the nonlinear response. Figure 4(b) illustrates the five sample-point scheme. The additional sample points can be expressed as [22]:

$$
\begin{aligned}
{}^{3}\mathrm{V}_i &= \left[\mu_1, \ldots, \mu_i - 1.5\sqrt{\lambda_i}, \ldots, \mu_N\right] \\
{}^{4}\mathrm{V}_i &= \left[\mu_1, \ldots, \mu_i + 1.5\sqrt{\lambda_i}, \ldots, \mu_N\right]
\end{aligned}
\tag{9}
$$

The nonlinear behavior can be captured accurately using the two additional points. By excluding the repeated runs of the mean value, a total of (2$N$+1) or (4$N$+1) runs are required for the three and five sample-point schemes, respectively.

Once the corresponding system responses are obtained at the sample points, the moving least square (MLS) or stepwise moving least square (SMLS) method [22] is utilized to interpolate and extrapolate the responses at the integration points. Using the approximated system

responses, $\hat{y}$, at the integration points, the numerical integration of the one-dimensional integrations finally takes the following form [22]:

$$\begin{aligned}&\int_{-\infty}^{+\infty}\left\{y\left(\mu_1,\ldots,\mu_{j-1},x_j,\mu_{j+1}\ldots,\mu_N\right)\right\}^r \\ &\qquad\times f\left(x_j \mid x_1=\mu_1,\ldots,x_{j-1}=\mu_{j-1},x_{j+1}=\mu_{j+1},\ldots,x_N=\mu_N\right)dx_j \\ &\cong\sum_{i=1}^{n} w_{j,i}\left\{\hat{y}\left(\mu_1,\ldots,\mu_{j-1},x_{j,i},\mu_{j+1}\ldots,\mu_N\right)\right\}^r\end{aligned} \tag{10}$$

### 2.1.3. Assembly Yield Loss Prediction of Thin Flat Ball Grid Array

The assembly yield of a plastically encapsulated package is determined. A viscoelastic analysis to predict the warpage is described after defining the warpage. The uncertainty propagation analysis and PDF estimation are followed.

#### 2.1.3.1. Package Description

A stacked die thin flat ball grid array (TFBGA) package is often used as the top package of a Package-on-Package (PoP). Figure 5(a) shows the schematic of a typical TFBGA package. The encapsulation of the TFBGA package is done by the transfer molding process. For successful PoP stacking with the high assembly yield, the package warpage at the solder reflow temperature must be controlled [23-25].

Typically, the TFBGA packages are produced by memory manufacturers and shipped to the outsourced semiconductor assembly and test services (OSAT) companies for the PoP assembly. Therefore, the TFBGA packages are required to meet the warpage specification (e.g., ±0.1 mm for the package body size of 15 mm × 15 mm and the ball pitch of 0.5 mm [26]) before shipment. The packages with warpage exceeding the specification cannot be processed further, which becomes a yield loss. It is suggested by JEDEC and JEITA [27] that the package warpage at solder reflow temperature be measured over the area where solder joints are located (will be referred to as

“measuring zone”). Figure 5(b) shows the measuring zone of the TFBGA package used in this study.

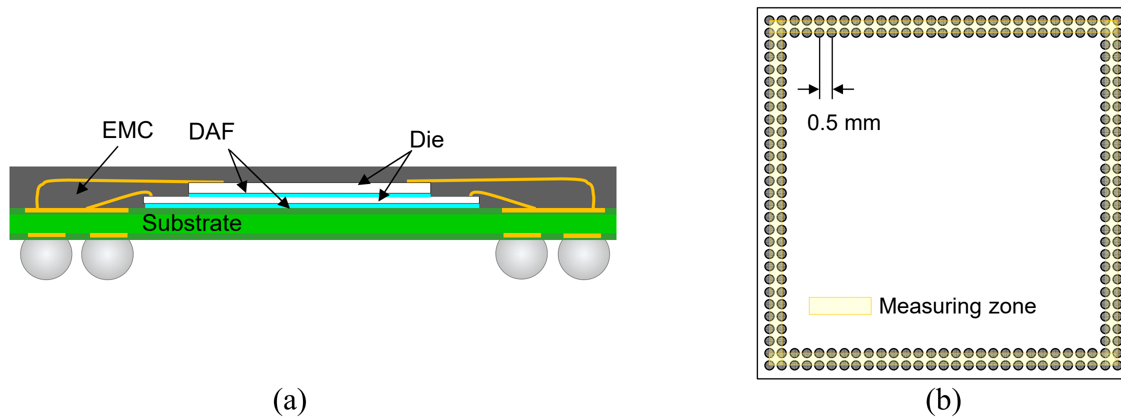



Figure 5. TFBGA package: (a) cross-sectional view and (b) bottom view showing the measuring zone

Figure 6 shows the definition of package warpage and the sign convention. Figure 6(a) shows a convex package (corners down during assembly – a positive warpage), while Figure 6(b) shows a concave package (corners up during assembly – a negative warpage). The red dashed line shown in Figure 6 indicates the reference plane; the coefficients of the equation of the reference plane are calculated by the least square method with the out-of-plane deformation across the x-y spatial dimensions of the specimen in the measuring zone. The distance between the highest point in the measuring zone and the reference plane is denoted as **A**, whereas the distance between the lowest point in the measuring zone and the reference plane is denoted as **B**. The magnitude of the package warpage is defined as the sum of **A** and **B**.

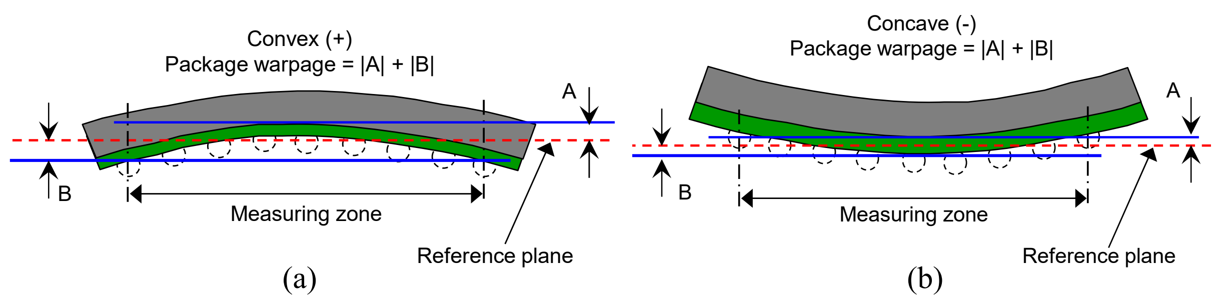



Figure 6. Package warpage definition and sign convention: (a) convex and (b) concave

#### 2.1.3.2. Numerical Analysis: Warpage Prediction

A quarter symmetry was used to build a finite element model with the element type SOLID185 in the commercial FEA package (ANSYS®), which supports the viscoelastic and elastic material properties. It takes about 1.5 hours for a single model run using an advanced workstation. Figure 7 shows the details of the model. The boundary condition and the die stack configuration are shown in (a). The enlarged view of the cross-section in (b) shows the details of the chip and the die attach film (DAF) configuration. The TFBGA assembly contains two dies. The die stack configuration is shown in (a) using white dashed lines.

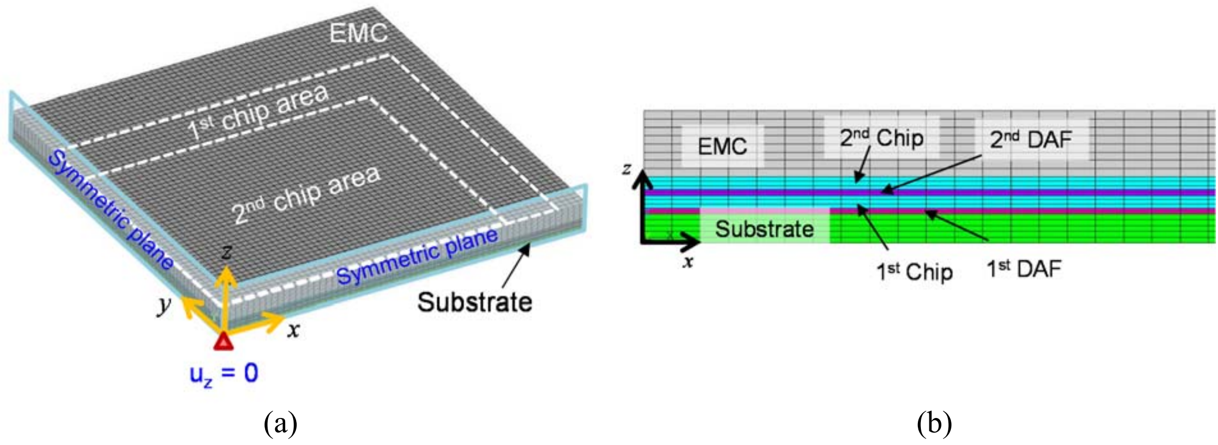



Figure 7. Package warpage definition and sign convention: (a) convex and (b) concave

The material properties and the nominal dimensions used in the model are summarized in Table 1 and Table 2, respectively. The temperature dependent Young's modulus of DAF measured by thermomechanical analysis (TMA) is shown in Figure 8. The EMC was modeled as a linear viscoelastic material. The master curves used in the model are shown in Figure 9 [1]. The Williams-Landel-Ferry (WLF) function was used to fit the shift factors at different temperatures, which can be expressed as:

$$\log\left(a_T\right) = \frac{C_1\left(T - T_{ref}\right)}{C_2 + \left(T - T_{ref}\right)} \tag{11}$$

where $a_T$ is the shift factor, $T_{ref}$ is the reference temperature (115 °C), and $C_1$, $C_2$ are the material constants. The values of $C_1$ and $C_2$ were -20.16 and -111.38, respectively.

Table 1 Material properties of TFBGA

| Material | Young's modulus (GPa) | Poisson's Ratio | CTE (ppm/°C) | | $T_g$ (°C) |
|---|---|---|---|---|---|
| | | | $\alpha_1$ (< $T_g$) | $\alpha_2$ (> $T_g$) | |
| Silicon die | 130 | 0.23 | 2.8 | | -- |
| DAF | Temp. dependent | 0.3 | 65.3 | 162.9 | 138 |
| Substrate | 46.794 | 0.3 | 16.2 (in-plane)<br>61.5 (out-of-plane) | | -- |
| EMC | Viscoelastic | 0.21 | 9.12 | 35.13 | 137.5 |

Table 2 Dimensions of TFBGA

| | length × width × thickness |
|---|---|
| 1st Die (mm) | 13 × 11 × 0.575 |
| 1st DAF (mm) | 13 × 11 × 0.025 |
| 2nd Die (mm) | 11 × 9 × 0.575 |
| 2nd DAF (mm) | 11 × 9 × 0.025 |
| Substrate (mm) | 15 × 15 × 0.13 |
| EMC (mm) | 15 × 15 × 0.59 |

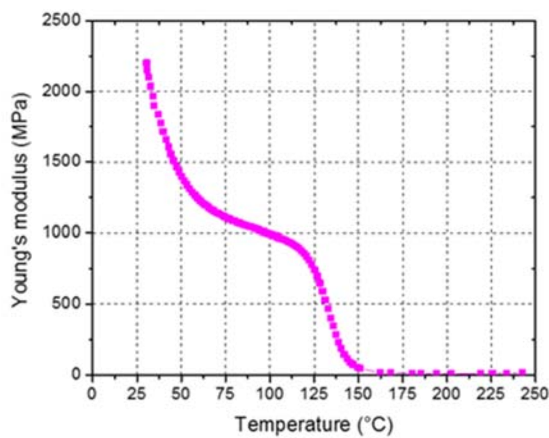



Figure 8. Temperature dependent

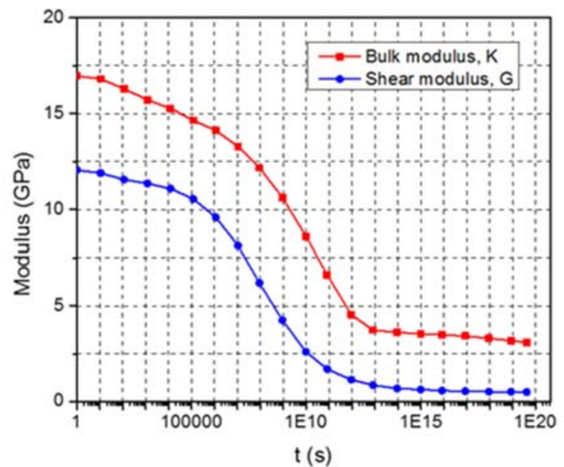



Figure 9. Master curves of bulk modulus and shear modulus of EMC [1]

The EMC molding process is done at 175 °C, which can be assumed as the stress free temperature. The package is then subjected to the solder reflowing process during the assembly. Figure 10 shows the complete thermal excursion of the package used for warpage prediction. The conventional lead-free solder reflow profile is considered [28], where the peak temperature is 260 °C.

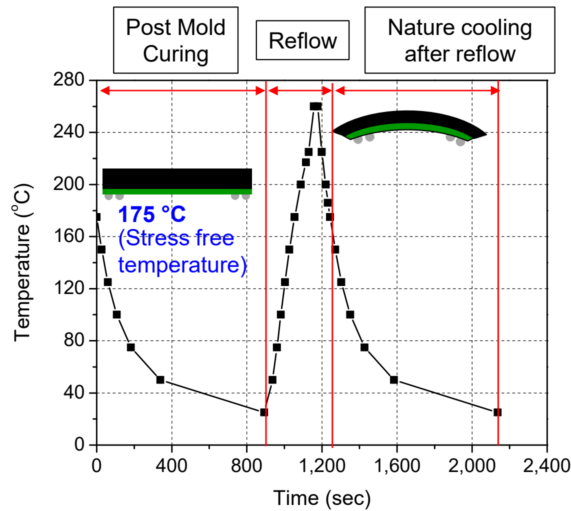



Figure 10. Completed thermal excursion

The deformed configuration of the TFBGA package with the nominal design at the peak reflow temperature is shown in

Figure 11(a). The light pink area indicates the measuring zone and the white circles represent the solder ball locations.

Figure 11(b) shows the reference plane determined based on the z-direction displacements of the nodes in the measuring zone. The package warpage was calculated based on the definition described in Section 3.1, i.e., the sum of (1) the distance of the highest point in the measuring zone to the reference plane and (2) the distance of the lowest point in the measuring zone to the reference plane. The package warpage of the nominal design was 40.16 μm.

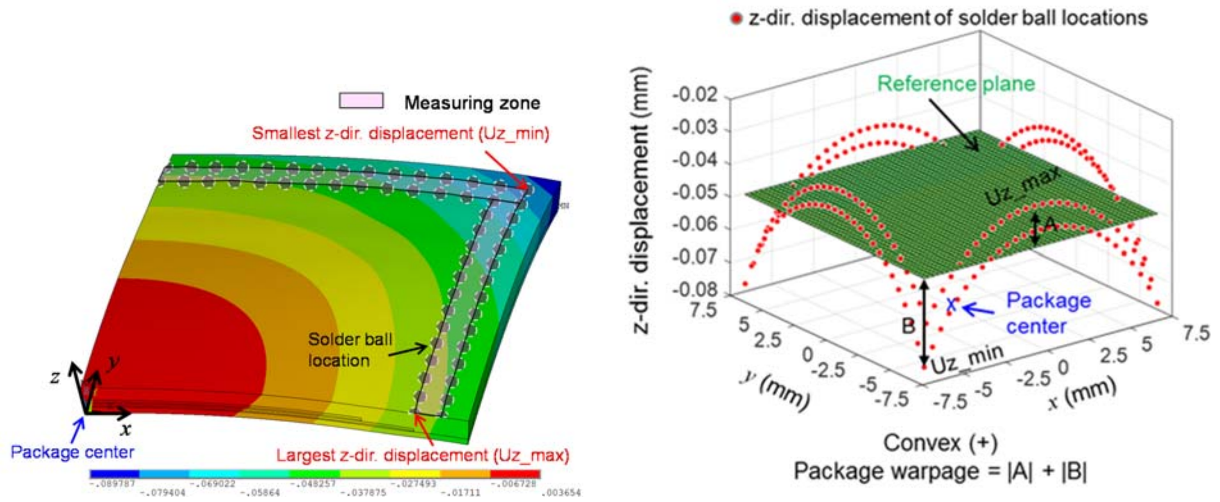



Figure 11. (a) Deformed configuration of TFBGA package at 260 °C (20x magnification) and (b) package warpage calculated from on reference plane

#### 2.1.3.3. Uncertainty Propagation Analysis by EDR

##### 2.1.3.3.1. Input Random Variables

It is well known that the manufacturing variables tend to follow a normal distribution according to the central limit theorem [29]. The 12 random input variables with the means and standard deviations used in the study are listed in Table 3. The material properties ( $x_9$ to $x_{12}$ ) were measured, and the dimensions of $x_1$ to $x_8$ were obtained from the manufacturing specifications [30, 31].

Among the 12 input variables, two pairs of properties have the statistical correlation: (1) the EMC thickness $\left(t_{EMC}\right)$, $x_3$, and the PCB thickness $\left(t_{PCB}\right)$, $x_4$, and (2) the EMC CTE below and above $T_g$, $x_9$ and $x_{10}$. To define the joint PDF of the correlated and normally distributed input variables, an additional parameter called the correlation coefficient is required. The joint PDF with these 5 parameters (i.e., the means and standard deviations of two variables, and the correlation coefficient) is called the bivariate normal distribution.

Table 3  Input variables

| Variables | Physical meaning | Mean | Std. Dev. | Distribution | Correlation coefficient |
|---|---|---|---|---|---|
| $x_1$ | PKG length (mm) | 15 | 0.033 | Normal | -- |
| $x_2$ | PKG width (mm) | 15 | 0.033 | Normal | |
| $x_3$ | EMC thickness (mm) | 0.59 | 0.029 | Bivariate Normal | -0.35 |
| $x_4$ | PCB thickness (mm) | 0.13 | 0.01 | | |
| $x_5$ | 1st Chip thickness (mm) | 0.0575 | 0.001 | Normal | -- |
| $x_6$ | 2nd Chip thickness (mm) | 0.0575 | 0.001 | Normal | -- |
| $x_7$ | 1st DAF thickness (mm) | 0.025 | 0.00375 | Normal | -- |
| $x_8$ | 2nd DAF thickness (mm) | 0.025 | 0.00375 | Normal | -- |
| $x_9$ | EMC CTE above $T_g$ (ppm/°C) | 35.13 | 4.24 | Bivariate Normal | 1 |
| $x_{10}$ | EMC CTE below $T_g$ (ppm/°C) | 9.12 | 1.1 | | |
| $x_{11}$ | PCB CTE (ppm/°C) | 16.2 | 0.81 | Normal | -- |
| $x_{12}$ | PCB modulus (MPa) | 46794 | 159 | Normal | -- |

The package thickness, $t_{PKG}$, is equal to $t_{EMC}$ plus $t_{PCB}$, i.e., $t_{PKG} = t_{EMC} + t_{PCB}$. The statistical distributions of $t_{PKG}$, $t_{EMC}$, and $t_{PCB}$ should also satisfy this relationship. Based on the manufacturing specification [30], the package thickness, $t_{PKG}$, is given as 0.72 ± 0.08 mm. The package thickness is expected to follow a normal distribution, and it can be expressed with the mean and standard deviations of $\mu_{PKGt}$ = 0.72 mm and $\sigma_{PKGt}$ = 0.027 mm. It is to be noted that $\sigma_{PKGt}$ is set to be one third of the tolerance, which makes 99.73% of $t_{PKG}$ lie within the tolerance.

The correlation coefficient of $t_{EMC}$ and $t_{PCB}$ was determined using the distribution of the package thickness. A sweeping analysis was conducted over the theoretical range of correlation coefficient, [-1, 1] with a step size of 0.05, which produced a total of 41 correlation coefficients.

For each correlation coefficient, $\rho_{t,i}$ ($i$ = 1 to 41), random sampling of the bivariate normal distribution of $t_{EMC}$ and $t_{PCB}$ was conducted to generate 100,000 pairs of $t_{EMC}$ and $t_{PCB}$ data, which subsequently produced 100,000 $t_{PKG}$ data. The $t_{PKG}$ data was then fitted into a normal distribution to calculate the mean, $\mu_{\rho_{t,i}}$, and standard deviation, $\sigma_{\rho_{t,i}}$. The least square error was calculated to represent the degree of $\mu_{\rho_{t,i}}$ and $\sigma_{\rho_{t,i}}$ deviated from $\mu_{PKGt}$ and $\sigma_{PKGt}$. The results are shown in Figure 12. The correlation coefficient of $X_3$ and $X_4$ was determined as -0.35.

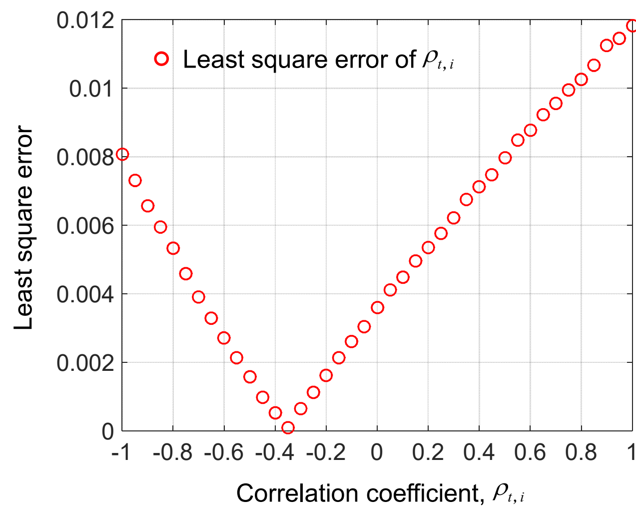



Figure 12. MCS sweeping analysis results for correlation coefficient of EMC thickness and PCB thickness

Unlike the above case, the EMC CTEs above and below $T_g$ always increase or decrease in the same direction since any pair of these two variables are measured from the same sample. Accordingly, it is reasonable to assume that the EMC CTEs above and below $T_g$ have perfect positive correlation (i.e., the correlation coefficient of "unity").

#### 2.1.3.3.2. Eigenvector Sampling and Sample Points

The covariance of input variables $x_i$ and $x_j$, which quantify the linear dependency between these two variables, is defined as:

$$\mathrm{Cov}\left(x_i, x_j\right) = \Sigma_{ij} = E\left[\left(x_i - \mu_i\right)\left(x_j - \mu_j\right)\right] \tag{12}$$

where $E(\cdot)$ is the expectation operator; $\mu_i$ and $\mu_j$ are the mean values of the input variable $x_i$ and $x_j$, respectively.

After calculating the covariance between each pair of the $N$ number of input variables, the covariance matrix can be obtained as [22]:

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \Sigma_{12} & \Sigma_{13} & \cdots & \Sigma_{1N} \\ \Sigma_{21} & \sigma_2^2 & \Sigma_{23} & \cdots & \Sigma_{2N} \\ \Sigma_{31} & \Sigma_{32} & \sigma_3^2 & \cdots & \Sigma_{3N} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \Sigma_{N1} & \Sigma_{N2} & \Sigma_{N3} & \cdots & \sigma_N^2 \end{bmatrix}$$

$$= \begin{bmatrix} 0.033^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.033^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.029^2 & -0.0001015 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -0.0001015 & 0.01^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.001^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.001^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.00375^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.00375^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 4.24^2 & 4.664 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 4.664 & 1.1^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.81^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 159^2 \end{bmatrix} \tag{13}$$

where $\Sigma_{ii} = \sigma_i^2$ is the variance of the input variable $x_i$ and $\Sigma_{ij} = \Sigma_{ji}$. By solving the eigenvalue problem of the covariance matrix (i.e., $\Sigma X_E = \lambda X_E$), the eigenvalues $\lambda$ and the corresponding eigenvectors $X_E$ are obtained. The eigenvalues, $\lambda$, and the corresponding eigenvectors, $X_E$, of the covariance matrix of Eq. (13) are:

$$\lambda = \begin{bmatrix} 0.033^2 \\ 0.033^2 \\ 0.000855 \\ 0.000086 \\ 0.001^2 \\ 0.001^2 \\ 0.00375^2 \\ 0.00375^2 \\ 4.38^2 \\ 0 \\ 0.81^2 \\ 159^2 \end{bmatrix}, \quad X_E = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.9911 & 0.1333 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -0.1333 & 0.9911 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.9680 & -0.2511 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.2511 & 0.9680 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{14}$$

The multi-dimensional integration of the joint PDF with 12 dimensions (i.e., 12 input variables) can be decomposed into 12 one dimensional integration along the 12 eigenvectors directions shown in Eq. (14). The marginal joint PDF of two input variables is used to illustrate the locations of the sampling points since it is difficult to show graphically the joint PDF with more than 3 input variables. The marginal joint PDF of $x_i$ and $x_j$ provides the probability of each $(x_i, x_j)$ pair, which is calculated by integrating the joint probability distribution of the 12 input variables over the 10 input variables other than $x_i$ and $x_j$. For example, the marginal joint PDF of $x_3$ and $x_4$ can be expressed as:

$$f(x_3, x_4) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \ldots, x_{12})\, dx_1 dx_2 dx_5 \cdots dx_{12} \tag{15}$$

Figure 13 shows the marginal joint PDF of an uncorrelated case, where the package length ($x_1$) and package width ($x_2$) are the uncorrelated input variables. After the additive decomposition is completed, this bivariate marginal joint PDF is further transformed into two one-dimensional marginal joint PDFs along the eigenvector directions (green and black lines). It is to be noted that for this uncorrelated case, the eigenvector directions are just the directions of the input variables. Figure 13 also shows the sample points of $2N$+1 sampling scheme (Eq. (8)) as well as the additional sample points required for $4N$+1 sampling scheme (Eq. (9)).

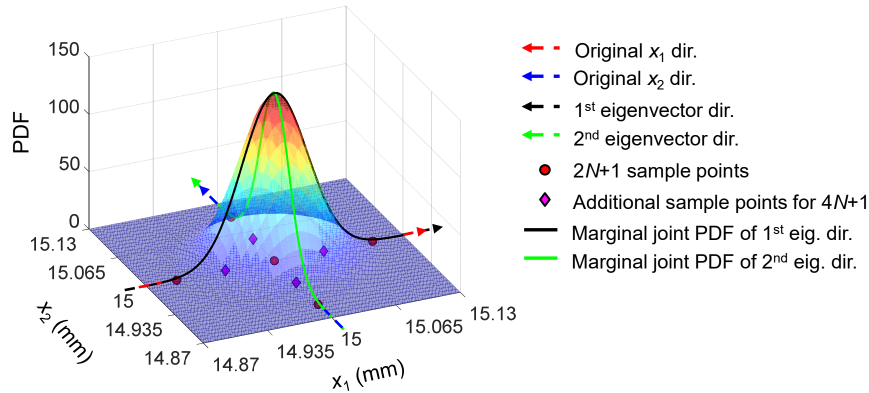



Figure 13. Marginal joint PDF of package length and width (uncorrelated case) and locations of sampling points

Figure 14 shows the marginal joint PDFs of the two correlated cases. The marginal joint PDF of EMC thickness ($x_3$) and PCB thickness ($x_4$) is shown in (a), and the marginal joint PDF of EMC CTE above and below $T_g$ ($x_9$ and $x_{10}$) in (b). Due to the statistical correlation, the eigenvector directions are different from the original variable directions. It should be noted that, in (b), the variation along the second eigenvector direction was zero due to the perfect correlation between $x_9$ and $x_{10}$.

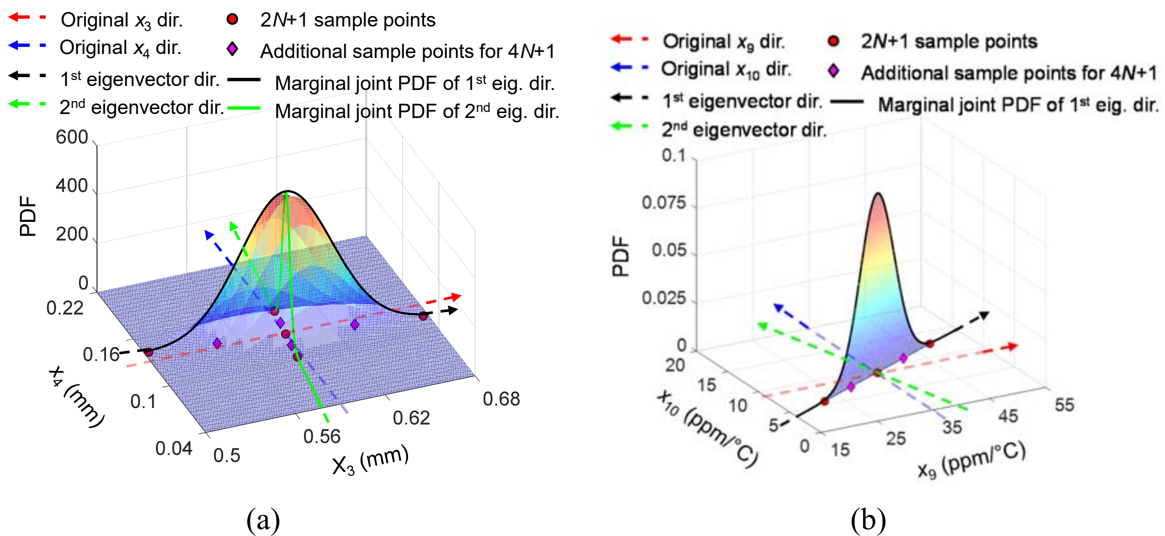



Figure 14. Marginal joint PDF and locations of sampling points for correlated input variables: (a) EMC and PCB thickness and (b) EMC CTE below $T_g$ and above $T_g$

The warpage at the each sample point was calculated by the same procedure described in Section 2.1.3.2. The results are summarized in Table 4. Based on the 2$N$+1 sampling scheme, a total of 25 simulations were conducted for the 12 input variables.

Table 4 Package warpage simulation results at sample points

| Variable | Package warpage (μm) | | | | |
|---|---|---|---|---|---|
| | $-3\sqrt{\lambda_i}$ | $-1.5\sqrt{\lambda_i}$ | $\mu_i$ | $+1.5\sqrt{\lambda_i}$ | $+3\sqrt{\lambda_i}$ |
| $x_1$ | 40.00 | 40.08 | 40.16 | 40.25 | 40.33 |
| $x_2$ | 40.06 | 40.11 | 40.16 | 40.22 | 40.27 |
| $x_3$ | 52.94 | 47.27 | 40.16 | 31.16 | 19.07 |
| $x_4$ | 27.87 | 33.67 | 40.16 | 47.44 | 55.62 |
| $x_5$ | 41.24 | 40.70 | 40.16 | 39.63 | 39.11 |
| $x_6$ | 40.83 | 40.49 | 40.16 | 39.83 | 39.51 |
| $x_7$ | 44.31 | 42.31 | 40.16 | 37.94 | 35.69 |
| $x_8$ | 41.49 | 40.88 | 40.16 | 39.39 | 38.57 |
| $x_9 / x_{10}$ | -12.78 | 14.40 | 40.16 | 65.26 | 90.13 |
| $x_{11}$ | 55.99 | 48.11 | 40.16 | 32.28 | 24.33 |
| $x_{12}$ | 40.47 | 40.31 | 40.16 | 40.01 | 39.86 |

#### 2.1.3.3.3. Statistical Moments

As described in Section 2, the statistical moments can be obtained by calculating the multiple one-dimensional integrations in Eq. (6). In this study, each one-dimensional integration was calculated by the numerical integration algorithm called the moment quadrature rule [21]. It was demonstrated that it can calculate the one-dimensional integration for an arbitrary distribution of the input variable $x_i$ with good accuracy and efficiency, compared with other conventional integration methods such as Gauss-Legendre and Gauss-Hermite quadratures [21].

Figure 15 shows the weights and the approximated package warpages of the integration points for two representative variables: (a) $x_{11}$ (CTE of PCB) and (b) $x_3$ (EMC thickness). The red dots represent the package warpages at the sample points. A total of 21 integration points were suggested in Ref. [22] for several examples with nonlinear system response. The 21 integration points were also used in this study and expected to be sufficient since the system response of this study is less nonlinear. As described in Section 2, the weight of each integration point is represented by the area of the corresponding bar. For example, the central bar in (a) approximates the area under the PDF of $x_{11}$ from $x_{11,11} - \frac{x_{11,11} - x_{11,10}}{2}$ to $x_{11,11} + \frac{x_{11,12} - x_{11,11}}{2}$.

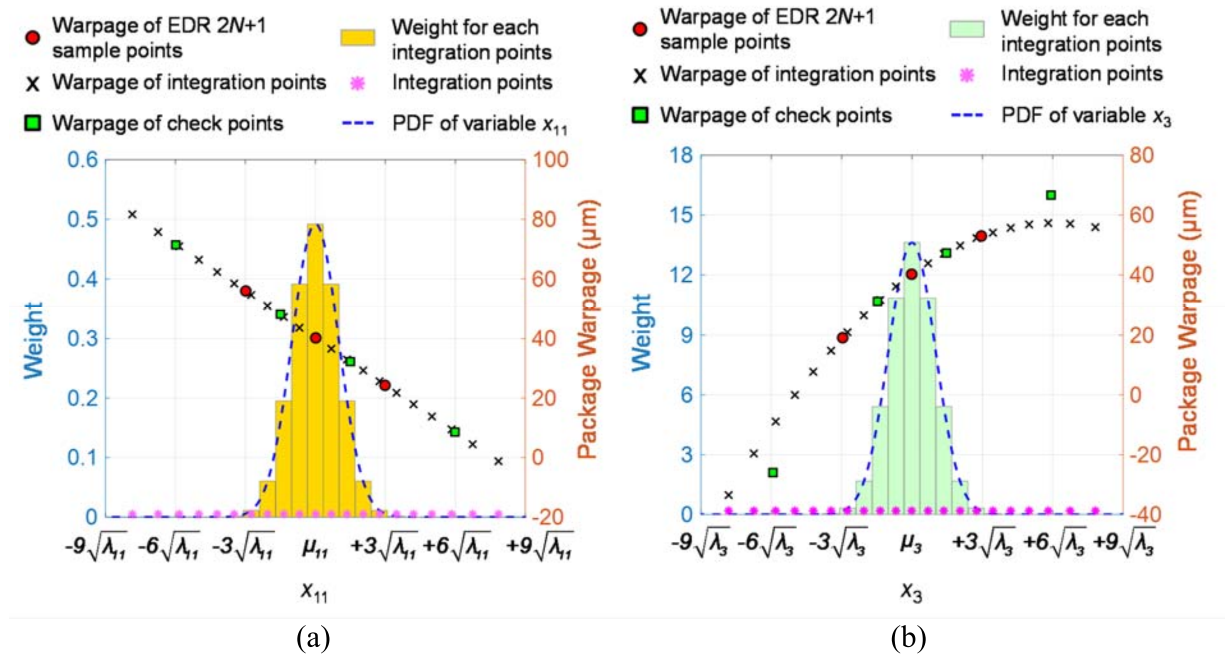



Figure 15. Integration points and weights for 1-D numerical integration: (a) $x_{11}$ and (b) $x_3$

In Figure 15(a), the package warpages at the three sample points ( $\mu_{11}$ and $\mu_{11} \pm 3\sqrt{\lambda_{11}}$ ) linearly decrease along the direction of $x_{11}$. Therefore, the package warpages of the integration points (black cross) can be accurately interpolated and extrapolated by MLS, which was confirmed

by four additional simulations conducted at $\mu_{11} \pm 1.5\sqrt{\lambda_{11}}$ and $\mu_{11} \pm 6\sqrt{\lambda_{11}}$. Most of the input variables in this study show linear response curves similar to $x_{11}$.

The most nonlinear response curve is observed with EMC thickness, $x_3$ (Figure 15b)). Due to the warpage within $\mu_3 \pm 3\sqrt{\lambda_3}$ is merely intermediately nonlinear, two additional simulations conducted at $\mu_3 \pm 1.5\sqrt{\lambda_3}$ indicate that the interpolation was accurate, whereas the extrapolation results deviated from the simulations at $\mu_3 \pm 6\sqrt{\lambda_3}$. However, as expected, it is clearly shown in Figure 15 that the contribution of weights are negligible for $X < \mu_i - 3\sqrt{\lambda_i}$ and $X < \mu_i + 3\sqrt{\lambda_i}$, and thus, even the extrapolation by MLS contains error, the effect on the integration results is minimal.

Once all the one-dimensional integrations in Eq. (10) are completed, they are combined to calculate the statistical moments. The first four statistical moments are listed in Table 5.

Table 5 First four statistical moments

| | | Mean | Std. Dev. | Skewness | Kurtosis |
|---|---|---|---|---|---|
| EDR sampling scheme | $2N$+1 | 39.68 | 19.48 | -0.0488 | 3.0053 |
| | $4N$+1 | 39.76 | 19.25 | -0.0478 | 3.0332 |

#### 2.1.3.3.4. PDF Estimation

After obtaining the statistical moments, the probability estimation method such as the method of moments (MOM) and the Pearson system can be used to construct the PDF of random response, which is the final outcome of the EDR method for an uncertainty propagation analysis.

In this study, the PDF of package warpage was constructed using the Pearson system [32]. The Pearson system is a family of continuous probability distributions, which offers flexibility in

constructing the PDF based on the first four statistical moments (mean, standard deviation, skewness and kurtosis). The Pearson distributions of the system response, $y$, are defined by the following differential equation [32]:

$$\frac{1}{p(y)}\frac{dp(y)}{dy} = -\frac{a+y}{c_0 + c_1 y + c_2 y^2} \tag{16}$$

where $a$, $c_0$, $c_1$ and $c_2$ are four parameters to describe the PDF. Based on the theoretical derivation, the four parameters can be determined by the first four moments, which can be expressed as [32]:

$$\begin{cases} c_0 = \left(4\beta_2 - 3\beta_1^2\right)\left(10\beta_2 - 12\beta_1^2 - 18\right)^{-1}\sigma^2 \\ a = c_1 = \beta_1\left(\beta_2 + 3\right)\left(10\beta_2 - 12\beta_1^2 - 18\right)^{-1}\sigma \\ c_2 = \left(2\beta_2 - 3\beta_1^2 - 6\right)\left(10\beta_2 - 12\beta_1^2 - 18\right)^{-1} \end{cases} \tag{17}$$

As denoted in Section 2.1.2, $\sigma$ is the standard deviation, $\beta_1$ is the skewness, and $\beta_2$ is the kurtosis. It is worth noting that the 1$^{st}$ moment (i.e., mean, $\mu$) is not shown in Eq. (17) since the Pearson system first constructs the PDF about the mean of zero and then shifts the PDF to the true mean.

The coefficients determined from the statistical moments obtained in Section 3.3.3 are summarized in Table 6. Figure 16(a) depicts the predicted PDFs. Figure 16(b) shows the enlarged view of the tail-end region. It is clear that the results of 2$N$+1 and 4$N$+1 sampling schemes are virtually identical, which is attributed to the fact that most of the response curves are linear. By applying the specification of JEDEC [26], i.e., ±0.1 mm for the TFBGA package in this study, the prediction yield losses of the two schemes are 765 ppm and 751 ppm, respectively.

Table 6 Coefficients of Pearson system

| | | $c_0$ | $c_1$ | $c_2$ |
|---|---|---|---|---|
| EDR sampling scheme | 2$N$+1 | 378.96 | -0.4745 | 0.0003 |
| | 4$N$+1 | 365.03 | -0.4515 | 0.0048 |

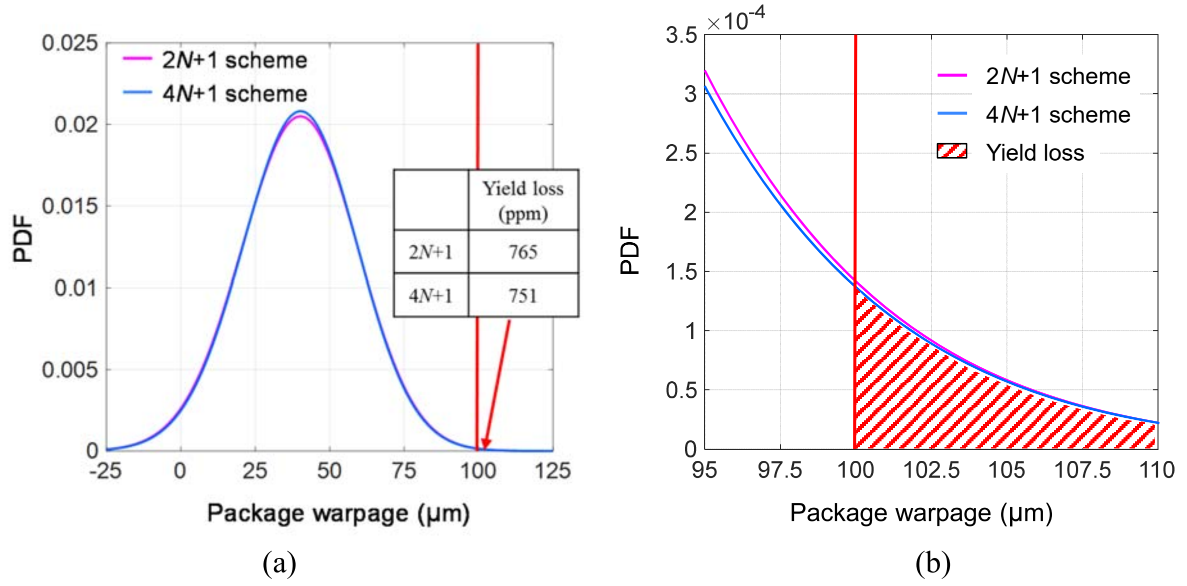



Figure 16. PDFs of and assembly yield loss predicted by two schemes of EDR: (a) entire PDF and (b) enlarged view of the tail-end region

### 2.1.4. Validity of the Proposed Approach

It has been shown that the yield loss (i.e., tail-end probability) of a package with 12 input variables can be predicted accurately by as few as 25 simulations. The direct MCS is used to evaluate the accuracy of the EDR results quantitatively.

The direct application of the MCS to the current problem was impractical due to excessive computational time. Instead, an empirical model obtained from the coplanarity data at room temperature was used for the verification. The empirical coplanarity model produced by using the stepwise regression with 43,358 coplanarity data points is given as:

$$
\begin{aligned}
y_{Copl} = \; & 2.6366x_1 + 0.4564x_2 - 74.4052x_3 + 36.0588x_4 - 0.06189(x_5 + x_6) \\
& + 0.0274(x_7 + x_8) + 1.7329x_9 + 1.73875x_{10} + 0.6273x_{11} - 0.00116x_{12} \\
& - 0.49186\left((x_5 + x_6)\Big/\left(0.4746x_1x_2x_3 - 0.036\left(1.913(x_5 + x_6) + 0.8(x_7 + x_8)\right)\right)\right) \\
& + 2.0361\left((x_5 + x_6)/x_1x_2x_4\right) + 14.662
\end{aligned}
\tag{18}
$$

where $y_{Copl}$ is the coplanarity and $x_1$ to $x_{12}$ are input variables listed in Table 3. The two pairs of input variables – (1) the EMC thickness, $x_3$, and the PCB thickness, $x_4$, and (2) the EMC CTE below and above $T_g$, $x_9$ and $x_{10}$ – are still considered to be statistically correlated with the same correlation coefficients as -0.35 and 1, respectively.

As mentioned earlier, the statistical moments and the yield loss estimated from the MCS exhibit statistical variations. It is well-known that the true values can be obtained by employing the unbiased estimators in multiple repetitions. According to the central limit theorem, 30 repetitions will produce a good approximation of the true value [33, 34]. More details about the unbiased estimators of the 1$^{st}$ to 4$^{th}$ statistical moments and the yield loss can be found in Ref. [35] and [11], respectively.

For the yield loss prediction, the coplanarity criterion was set as 80 µm according to the room temperature coplanarity criterion suggested by JEDEC [36, 37]. It is to be noted that this is different from the package warpage specification at reflow temperature discussed in Section 2.1.3.

In this study, the MCS with the sample size of 1,000,000 was conducted for 30 repetitions to estimate the true statistical moments and the true yield loss. The comparison between the MCS and EDR is summarized in Table 7. As expected, the results from 2*N*+1 and 4*N*+1 sampling schemes were nearly identical. Differences of statistical moments between the MCS result and the EDR results are very small (only fourth decimal place). The effect of these differences on the PDF construction are minimal, which produces only 8 ppm difference in yield loss prediction.

Table 7 Predicted statistical moments and yield loss by MCS and EDR

| | Mean | Std. Dev. | Skewness | Kurtosis | Yield loss (ppm) |
|---|---|---|---|---|---|
| MCS* | 54.8721 | 7.9191 | -0.0006 | 2.9995 | 762 |
| EDR 2*N*+1 | 54.8720 | 7.9202 | 0.0000 | 3.0000 | 754 |
| EDR 4*N*+1 | 54.8720 | 7.9201 | 0.0000 | 3.0000 | 758 |

* Average value of 30 repetitions with sample size of 1,000,000

A quantitative comparison of the yield loss was made using the results of the MCS sample size of 1,000,000. The yield loss prediction by MCS is expressed as [11]:

$$\hat{p} = \left(1 - \frac{k}{N_{MCS}}\right) \quad (19)$$

where $N_{MCS}$ is the sample size used in MCS; and $k$ is the number of predicted coplanarity less than or equal to the coplanarity criterion. When $k$ and $N_{MCS}$ are sufficiently large, $\hat{p}$ can be approximated by the normal distribution with the mean of $p$ and the standard deviation of $\sqrt{p(1-p)/N_{MCS}}$ [11], where $p$ is the true yield loss.

Figure 17 shows the yield loss distribution of $N_{MCS}$ = 1,000,000. The yield loss predicted by EDR 2*N*+1 scheme is 754 ppm, which falls within the ±1% tolerance of the true yield loss (762 ± 7.62 ppm). The probability that the MCS predictions satisfy the ± 1% tolerance is theoretically 22.3% (the light green area). This theoretical probability is confirmed by the single MCS repetition results, which are also shown in Figure 17; only 7 out of 30 repetitions (23.3% probability) fall within the ± 1% error bound (± 7.62 ppm). In other words, for any yield loss predicted by MCS with a sample size of 1,000,000 has only 22.3% probability that the error will be smaller than or equal to than the yield loss predicted by EDR. When the tolerance is relaxed to

± 10% error of the true yield loss, the MCS has 99.4% probability to fulfill the tolerance. It is also confirmed by that all the 30 repetitions fall within the ± 10% error bounds (Figure 17).

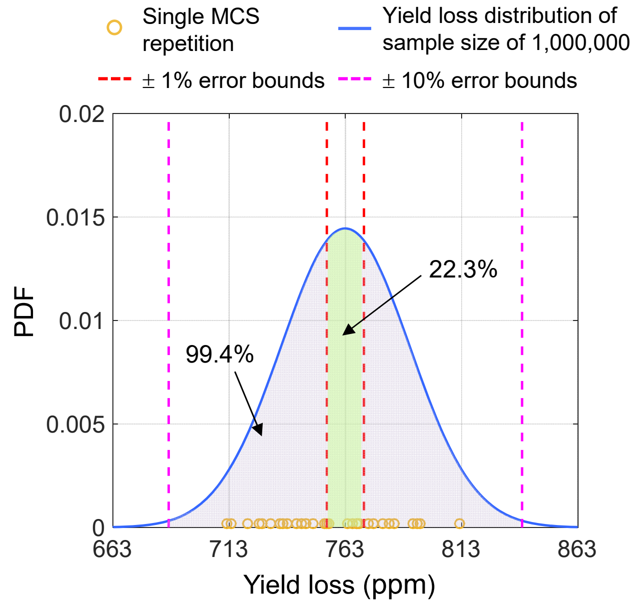



Figure 17. Yield loss distribution predicted by MCS with the sample size of 1,000,000

The yield loss by the MCS is affected obviously by the sample size. Table 8 summarizes the results obtained from different sample sizes. When the sample size is reduced to 10,000, the probability that the MCS prediction is to be comparable to the EDR is only 2.3%. Even the tolerance is relaxed to ± 10%, the sample size of 10,000 has only 21.7% probability.

Table 8 Effect of sample size on MCS predictions

| | Samples size = 10,000 | | Samples size = 100,000 | | Samples size = 1,000,000 | |
|---|---|---|---|---|---|---|
| | Theoretical probability | MCS repetitions | Theoretical probability | MCS repetitions | Theoretical probability | MCS repetitions |
| ± 1% error bounds (762 ± 7.62 ppm) | 2.3% | 0% (0/30) | 7.3% | 3.3% (1/30) | 22.3% | 26.7% (8/30) |
| ± 10% error bounds (762 ± 76.2 ppm) | 21.7% | 30% (9/30) | 61.6% | 70% (21/30) | 99.4% | 100% (30/30) |

Considering the fact that it takes approximately 1.5 hours to run the viscoelastic model used in this case study using an advanced workstation with 24 cores, the MCS with 10,000 model runs would take 2 months, which is impractical for most of the semiconductor packaging applications. Conversely, the EDR provides yield loss prediction with uncertainty less than 1% with only 25 model runs and results in 37.5 hours computational time.

Figure 18 shows the required modeling runs of the proposed approach, the random sampling method, RSM to illustrate the impact of the proposed approach. The yellow boxes show the required modeling runs for MCS for different level of tail-end probability and the green boxes show the case of Latin Hypercube sampling (LHS), which is a widely used advanced randomly sampling method. For the random sampling methods, the required number of modeling runs, $N_{MCS}$, is independent of the number of input variables $N$ of the problem. The number of modeling runs can be expressed as [38]:

$$N_{MCS} = \frac{1}{\varepsilon^2 p} \tag{20}$$

and

$$N_{LHS} = \frac{1}{\varepsilon p} \tag{21}$$

where $\varepsilon = \sigma_{\hat{p}} / p$ is statistical error and the standard deviation of $\sigma_{\hat{p}} = \sqrt{p(1-p)/N_{MCS}}$ or $\sigma_{\hat{p}} = \sqrt{p(1-p)/N_{LHS}}$ . The boxes in the figure are plotted using the statistical error of 10%. Moreover, the blue line represents the number of runs required by the EDR method with $4N+1$ sampling scheme and the red and pink lines represent the case of RSM using FFD and CCD, which require $s^N$ and $2^N + 2N + 1$, respectively, with $s$ as levels to be selected for each input variable and 5 levels are used here. Comparining to LHS for the tail-end propability prediciton of 1,000 ppm level, the proposed approach requreises the modeling runs 244 times and 123 times smaller

than LHS for 10 and 20 input variables, respectively. Meanwhile, the required modeling runs of the proposed apporach is 25 times and 13,000 times smaller than RSM with CCD for 10 and 20 input variables, respectively.

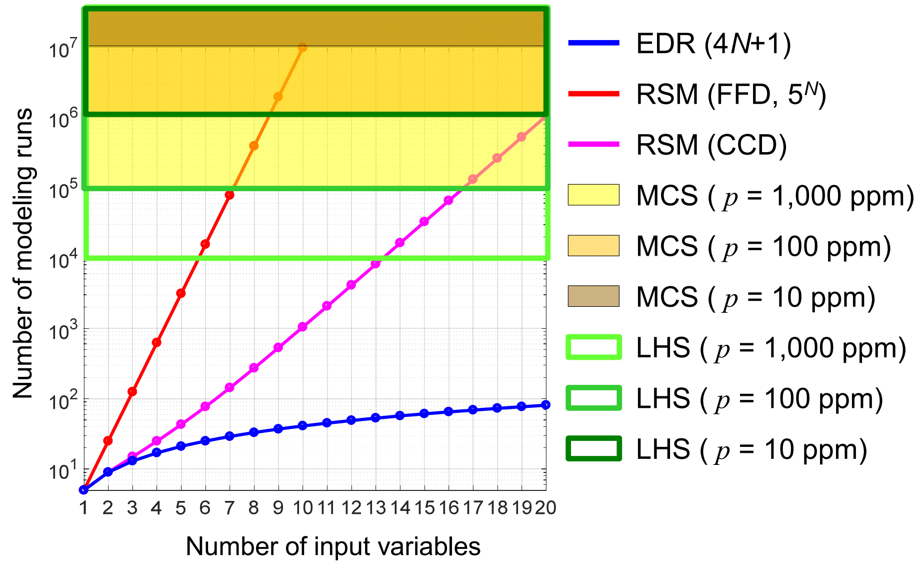



Figure 18. Number of required modeling runs for tail-end probability prediction

In this study, the accuracy of the proposed approach was confirmed for symmetric input distributions. It was confirmed in Ref. [22] that the EDR method is effective for both symmetric and asymmetric distributions. It is expected that the proposed approach will also work effectively with any asymmetric distributions as long as the input variables have linear-dependency.

### 2.1.5. Conclusion

The eigenvector dimension reduction (EDR) method was implemented to predict the assembly yield of a stacked die thin flat ball grid array (TFBGA) package. A total of 12 manufacturing input variables were considered during the yield prediction, among which two pairs of properties had the statistical correlation. The method calculated the statistical moments of warpage distribution first through dimensional reduction and eigenvector sampling. The probability density function (PDF) of the warpage was constructed from the statistical moments

by the Pearson system. The assembly yield was predicted from the PDF based on the JEDEC reflow flatness requirements.

In this case study, only 25 modeling runs were needed to predict the assembly yield with uncertainty less than 1% despite the fact that the prediction dealt with a tail-end probability (less than 1,000 ppm) with 12 input variables. The number of input variables was much larger than that has been conceived as the practical limit of the uncertainty propagation analysis. More applications of the EDR method are expected to the improve design and manufacturing processes of complex partially encapsulated components to avoid any early failure risks, in particular, tail-end probability related problems which have not been feasible due to a computational burden.

## 2.2. Stacking Yield Prediction of Package-on-Package Assembly Using Advanced Uncertainty Propagation Analysis: Part I Stochastic Model Development

### 2.2.1. Introduction

Package-on-Package (PoP) technology is widely used in advanced mobile applications. As illustrated in Figure 19, a top package (typically low I/O memory module) is stacked on a bottom package (typically high I/O logic module) using solder joints. Excessive warpages of the top and bottom packages during assembly often causes non-contact open, head-in-pillow, non-wet open, and solder bridging [39], which are directly related to PoP stacking yield. It has been known [2-4, 40-42] that the non-contact open is the most dominant factor to the yield loss.

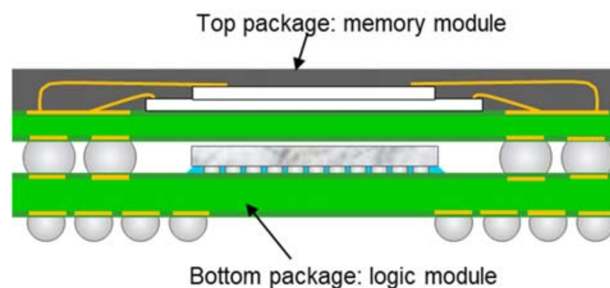



Figure 19. Schematic illustration of a typical Package-on-Package assembly

Figure 20 illustrates the non-contact open. In the illustration, solder joints are formed during the final assembly process when the top package solder balls touch the bottom package solder balls. The non-contact open occurs when a gap exists between the top solder ball and the bottom solder ball after the final assembly process.

The nominal warpage of packages can be optimized though material selections, designs and processes in the development phase. When moving into the production phase, however, the statistical distributions of package warpage and solder ball height, caused by the inherent variations in geometries and material properties, must be considered to predict a stacking yield loss accurately. Figure 20 illustrates schematically the statistical distributions (i.e., the probability

density functions) of five critical quantities that contribute to the gap formation. It is worth noting that the shape and sizes of solder joints are different because the solder balls have their own statistical distributions.

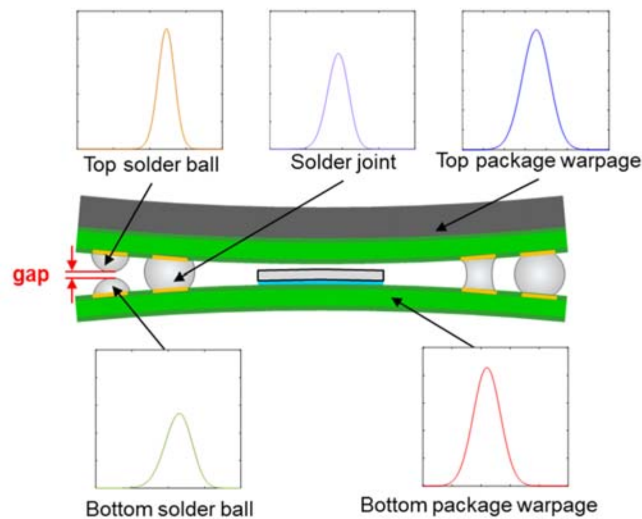



Figure 20. Schematic illustration of non-contact open

The previous statistical approaches for PoP stacking yield prediction mainly focused on a pad location that has the largest warpage difference [2-4]. This approach, namely considering a single pad location, can underestimate the yield loss because the non-contact open can occur at other pad locations if solder height is smaller than the warpage difference at those locations. In addition, the warpage distributions in the approach were obtained by experimental measurements. The limited sample size of the experiments added additional uncertainty about the estimation of statistical distributions.

More recently, a yield prediction model considering all pad locations in the stacking interface was proposed [41, 42]. The model was based on the assumption that the non-contact opens at different pads are "independent" events. This assumption is not always valid for PoP assemblies because the warpage values at different pads have statistical correlations among them as they come from the same package. In addition, a simple analytical model was used to predict

the warpage values of the top and bottom packages, which would be difficult to implement for advanced packages with complex geometries.

The objective of this section is to propose an advanced stochastic model for PoP stacking yield loss prediction. To cope with the above limitations, the proposed model takes into account all pad locations at the stacking interface while considering the statistical variations of warpages as well as the heights of solder balls and joints. The goal is achieved by employing three statistical approaches: (1) an advanced approximate integration-based method called eigenvector dimension reduction (EDR) method to conduct uncertainty propagation (UP) analyses, (2) the stress-strength interference (SSI) model to determine the non-contact probability at a single pad, and (3) the union of events considering the statistical dependence to calculate the final yield loss.

The section is divided into two parts. In this first part, theoretical development of the proposed stochastic model is presented. Implementation of the proposed model will be presented in the following section.

### 2.2.2. Conditions for Non-Contact Open

As illustrated in Figure 20, the non-contact open occurs when a gap exists between the top solder ball and the corresponding bottom solder ball. Two cases are considered based on the location of the maximum warpage difference: Case I for the corner and Case II for the center. Three possible scenarios (Scenario-1 to Scenario-3) of Case I are shown in Figure 21, where the gap formation of a single pad (say, $j^{th}$ pad) is illustrated on the cross section along the diagonal of a PoP assembly. Case II has only one scenario (Scenario-4), which is illustrated in Figure 22. It is to be noted that two peripheral rows are shown in Figure 21 and Figure 22 for the purpose of illustration, but these stacking scenarios can be applied for any ball patterns.

In Figure 21 and Figure 22, the warpages of the top and the bottom packages, the solder ball height of the top package, and the solder ball height of the bottom package at the single pad ($j^{th}$ pad) are denoted as $w_T^j$, $w_B^j$, $h_T^j$, and $h_B^j$, respectively. The thickness of a component on the bottom package plays an important role in all scenarios, and it is denoted as $t_{BC}$. In the illustrations, the component thickness is defined to be a sum of the chip thickness and the underfill/solder bump layer.

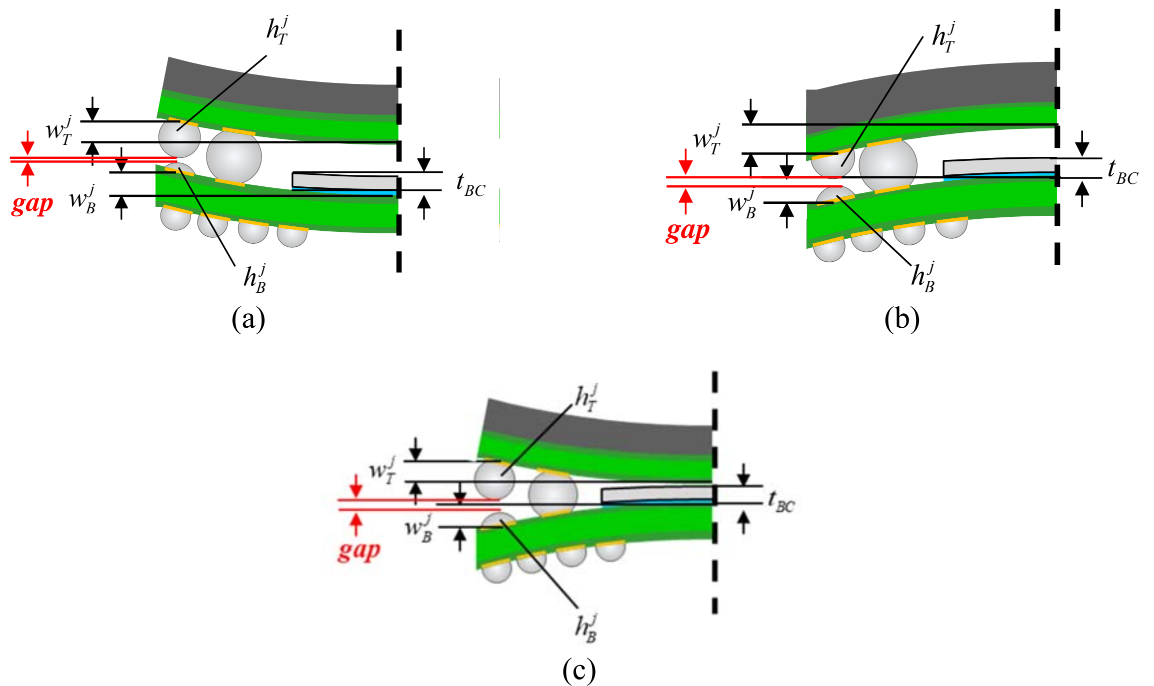



Figure 21. Three gap formation scenarios of Case I where the maximum warpage difference occurs at the corner of the package: (a) Scenario-1, (b) Scenario-2, and (c) Scenario-3

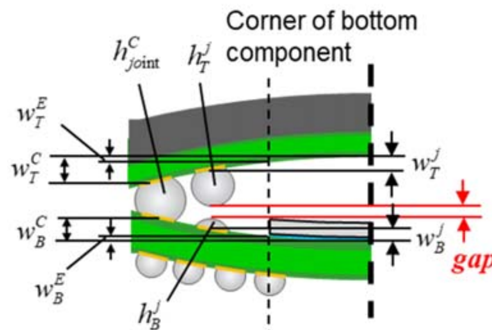



Figure 22. Gap formation scenario of Case II (Scenario-4) where the maximum warpage difference occurs at the center of the package

**Case I: Maximum warpage difference at the corner**

In Case I, the maximum warpage difference occurs at the corner of the packages. Thus, the component thickness of the bottom package ( $t_{BC}$ ) creates the minimum distance between the top package and the bottom package. Scenarios-1, 2 and 3 belong to this case. Using the geometrical parameters defined in Figure 21, the gap width of Case I at the $j^{th}$ pad, $gap^{j}_{case1}$, can be defined as:

$$gap^{j}_{case1} = \left|w^{j}_{T} - w^{j}_{B}\right| + t_{BC} - \left(h^{j}_{T} + h^{j}_{B}\right) \tag{22}$$

By defining the warpage difference between the top and bottom packages as $\hat{w}^{j} \equiv w^{j}_{T} - w^{j}_{B}$ and the total solder ball height (i.e., the sum of the heights of top and bottom solder balls) as $\tilde{h}^{j} \equiv h^{j}_{T} + h^{j}_{B}$, Eq. (22) can be rewritten as:

$$gap^{j}_{case1} = \left|\hat{w}^{j}\right| + t_{BC} - \tilde{h}^{j} \tag{23}$$

**Case II: Maximum warpage difference at the center**

In Case II, the maximum warpage difference occurs at the center of the packages. Scenario-4 belongs to this case. In this case, the distance between the top and bottom packages at the corner of the component on the bottom package should be considered in the gap calculation. This distance can be expressed as Figure 22:

$$d_{BC} = \left(w^{C}_{T} - w^{C}_{B}\right) - \left(w^{E}_{T} - w^{E}_{B}\right) + h^{C}_{joint} \tag{24}$$

where $w^{C}_{T}$ and $w^{C}_{B}$ are the warpages of the top and bottom packages at the outmost corner pad of the packages; $w^{E}_{T}$ and $w^{E}_{B}$ are the warpages of the top and bottom packages at the corner of the component on the bottom package; and $h^{C}_{joint}$ is the solder joint height at the outmost corner pad after assembly.

If $d_{BC}$ is larger than the thickness of the component, $t_{BC}$, the minimum distance between the top and bottom packages becomes $h^C_{joint}$, and the distance between the top and bottom packages the $j^{th}$ pad becomes $\left(w^C_T - w^C_B\right) + h^C_{joint} - \hat{w}^j$. Conversely, if $d_{BC}$ is smaller than $t_{BC}$, the component on the bottom package acts as a spacer, and the distance between the top and bottom packages the $j^{th}$ pad becomes $t_{BC} - \hat{w}^j$.

Based on above analysis, the gap width of Case II at the $j^{th}$ pad, $gap^j_2$, can be defined as:

$$gap^j_{case2} = \begin{cases} \left(w^C_T - w^C_B\right) + h^C_{joint} - \left(w^j_T - w^j_B\right) - \left(h^j_T + h^j_B\right) & \text{if } d_{BC} > t_{BC} \\ t_{BC} - \left(w^j_T - w^j_B\right) - \left(h^j_T + h^j_B\right) = t_{BC} - \hat{w}^j - \tilde{h}^j & \text{if } d_{BC} \leq t_{BC} \end{cases} \tag{25}$$

The distance between the top and bottom packages at the $j^{th}$ pad, $\Delta^j$, for $d_{BC} > t_{BC}$ can be expressed as:

$$\Delta^j = \left(w^C_T - w^C_B\right) + h^C_{joint} - \left(w^j_T - w^j_B\right) \tag{26}$$

Then, Eq. (25) can be rewritten as:

$$gap^j_{case2} = \begin{cases} \Delta^j - \tilde{h}^j & \text{if } d_{BC} > t_{BC} \\ t_{BC} - \hat{w}^j - \tilde{h}^j & \text{if } d_{BC} \leq t_{BC} \end{cases} \tag{27}$$

### 2.2.3. Stacking Yield Loss Prediction Model

The proposed yield loss prediction model is presented. A stacking yield loss prediction model for a single pad is described first, and it is extended to describe the stacking yield loss for multiple pads by considering the joint PDF of the five critical quantities. As illustrated in Figure 20, it is important to note that more than one non-contact open can occur at the stacking interface. This condition is incorporated during the final yield loss prediction of a PoP assembly using the concept of the union of events.

#### 2.2.3.1. Stacking Yield Loss Prediction Model for Single Pad

For a single pad ($j^{th}$ pad) on the stacking interface, the probability of non-contact open can be determined from the statistical interference between the PDFs of the warpage difference and the total solder ball height. When two PDFs overlap, statistical interference exists and the non-contact open occurs.

The stress-strength interference (SSI) model [43, 44] has been widely used for the reliability design of systems, where the statistical interference between the PDFs of stress (or load) and strength is considered. The SSI model represents the failure probability of a system as the probability that the load (*L*) exceeds the strength (*S*). The model can be expressed as:

$$\Pr\left(L-S>0\right)=\int_{-\infty}^{+\infty}\left(\int_{-\infty}^{L} f_S\left(S\right)dS\right)f_L\left(L\right)dL \tag{28}$$

where $f_L\left(L\right)$ and $f_S\left(S\right)$ are PDFs of the load and strength, respectively.

Figure 23 illustrates the interference between $f_L\left(L\right)$ and $f_S\left(S\right)$, where the inset shows a magnified view of the overlapping region. The red slashed area is the probability that a value of *L* occurs within a small interval of *dL* (i.e., $f_L\left(L\right)dL$). The green area is the probability that a value of $S$ is smaller than the given $L$. Thus, $\Pr\left(L-S>0\right)$ can be obtained by integrating the product of the red slashed area and the green area over the distribution of *L*.

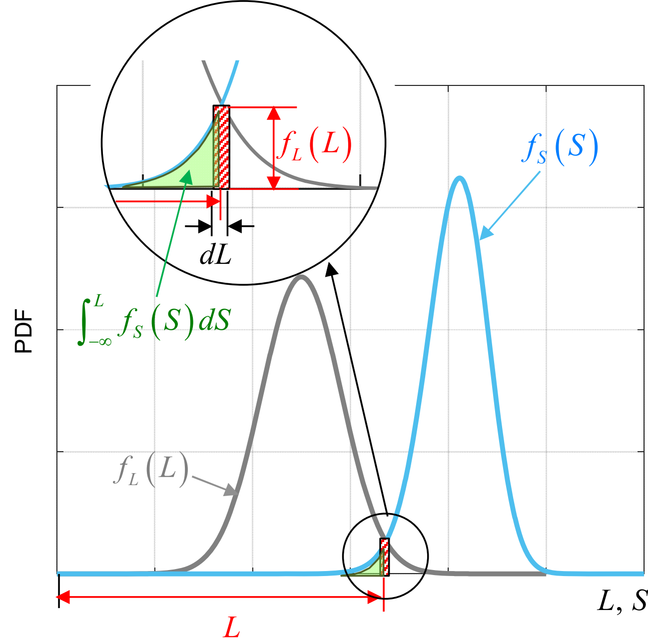



Figure 23. Statistical interference between the PDFs of load $f_L(L)$, and strength $f_S(S)$

A stochastic yield loss prediction model is proposed by adopting the SSI model. For both Cases I and II, the total solder ball height ($\tilde{h}^j$) can be regarded as strength (*S*) in Eq. (28). For Case I (Eq. (23)), Load (*L*) is the absolute value of the warpage difference between the top and bottom packages, $|\hat{w}^j|$. For Case II (Eq. (27)), the distance between the top and bottom packages at the $j^{th}$ pad, $\Delta^j$, or the warpage difference between the top and bottom packages, $\hat{w}^j$ becomes load (*L*) depending upon the condition between $d_{BC}$ and $t_{BC}$. Then, the probability of $gap^j_{case1}$ and $gap^j_{case2}$ larger than zero (i.e., a gap is formed) can be expressed as:

Case I: $$\Pr\left(gap^j_{case1} > 0\right) = \int_{-\infty}^{+\infty}\left\{\left[\int_{-\infty}^{|\hat{w}^j|+t_{BC}} f\left(\tilde{h}^j\right) d\tilde{h}^j\right] f\left(\left|\hat{w}^j\right|\right)\right\} d\left|w^j\right| \quad (29)$$

$$\text{Case II:} \quad \Pr\left(gap_{case2}^{j} > 0\right) = \begin{cases} \int_{-\infty}^{+\infty} \left\{ \left[ \int_{-\infty}^{\Delta^{j}} f\left(\tilde{h}^{j}\right) d\tilde{h}^{j} \right] f\left(\Delta^{j}\right) \right\} d\Delta^{j} & \text{if } d_{BC} > t_{BC} \\ \int_{-\infty}^{+\infty} \left\{ \left[ \int_{-\infty}^{t_{BC}-\hat{w}^{j}} f\left(\tilde{h}^{j}\right) d\tilde{h}^{j} \right] f\left(\hat{w}^{j}\right) \right\} dw^{j} & \text{if } d_{BC} \le t_{BC} \end{cases} \tag{30}$$

where $f\left(\left|\hat{w}^{j}\right|\right)$ is the PDF of the absolute value of the warpage difference between the top and bottom packages; $f\left(\Delta^{j}\right)$ is the PDF of distance between the top and bottom packages at the $j^{th}$ pad for $d_{BC} > t_{BC}$; $f\left(\hat{w}^{j}\right)$ is the PDF of the warpage difference between the top and bottom packages for $d_{BC} \le t_{BC}$; and $f\left(\tilde{h}^{j}\right)$ is the PDF of the total solder ball height.

The heights of solder balls of the top and bottom packages are two independent random variables because they are manufactured separately. Accordingly, the PDF of total solder ball height, $f\left(\tilde{h}^{j}\right)$, can be obtained by the convolution of $f\left(h_{T}^{j}\right)$ and $f\left(h_{B}^{j}\right)$, which is expressed as [45]:

$$f\left(\tilde{h}^{j}\right) = \int_{-\infty}^{+\infty} f\left(h_{T}^{j}\right) f\left[h_{B}^{j}\left(\tilde{h}^{j}, h_{T}^{j}\right)\right] dh_{T}^{j} \tag{31}$$

where $h_{B}^{j}\left(\tilde{h}^{j}, h_{T}^{j}\right) = \tilde{h}^{j} - h_{T}^{j}$ and $f\left[h_{B}^{j}\left(\tilde{h}^{j}, h_{T}^{j}\right)\right]$ is PDF of $h_{B}^{j}$ but the variable is transformed to $\tilde{h}^{j} - h_{T}^{j}$. The same expression will be used in the follows.

For Case I, two PDFs, $f\left(w_{T}^{j}\right)$ and $f\left(w_{B}^{j}\right)$, should be combined into one PDF, $f\left(\left|\hat{w}^{j}\right|\right)$. The top and bottom package are also manufactured separately, and thus, their warpages can be assumed as two independent random variables, and $f\left(\left|\hat{w}^{j}\right|\right)$ can be expressed as [46]:

$$\text{Case I:} \quad f\left(\left|\hat{w}^{j}\right|\right) = \begin{cases} \int_{-\infty}^{+\infty} f\left(w_{T}^{j}\right) f\left[w_{B}^{j}\left(w_{B1}^{j}\right)\right] dw_{T}^{j} \\ + \int_{-\infty}^{+\infty} f\left(w_{T}^{j}\right) f\left[w_{B}^{j}\left(w_{B2}^{j}\right)\right] dw_{T}^{j} & \text{for } \hat{w}^{j} > 0 \\ 0 & \text{for } \hat{w}^{j} \le 0 \end{cases} \tag{32}$$

where $w_{B1}^j = w_T^j + \hat{w}^j$ and $w_{B2}^j = w_T^j - \hat{w}^j$.

For the second condition of Case II, $f\left(\hat{w}^j\right)$ can be readily obtained from two PDFs $\left(f\left(w_T^j\right) \text{ and } f\left(w_B^j\right)\right)$. However, $f\left(\Delta^j\right)$ cannot be obtained directly from the individual PDFs because $\Delta^j$ contains variables that have statistical correlations. Eq. (26) can be rewritten as:

$$\Delta^j = \left(w_T^C - w_T^j\right) - \left(w_B^C - w_B^j\right) + h_{joint}^C = \hat{w}_T^j - \hat{w}_B^j + h_{joint}^C \tag{33}$$

where $\hat{w}_T^j$ and $\hat{w}_B^j$ are the warpage difference between the outmost corner pad and the $j^{th}$ pad for top and bottom package, respectively. It is important to note that $w_T^C$ and $w_B^C$ are correlated with $w_T^j$ and $w_B^j$, respectively, since they come from the same package. Therefore, the PDFs of $\hat{w}_T^j$ and $\hat{w}_B^j$ cannot be obtained directly by the convolution. Instead, they are obtained by the following two steps involving the joint PDFs of $\left(w_T^C, w_T^j \text{ and } w_B^C, w_B^j\right)$ [47]:

Step 1:

Perform the multivariate Fourier transform for the joint PDF as

$$\begin{cases} F_1\left[f\left(w_T^C,\, w_T^j\right)\right] = \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} e^{it_1 w_T^C + it_2 w_T^j} f\left(w_T^C,\, w_T^j\right) dw_T^C dw_T^j \\ F_2\left[f\left(w_B^C,\, w_B^j\right)\right] = \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} e^{it_1 w_B^C + it_2 w_B^j} f\left(w_B^C,\, w_B^j\right) dw_B^C dw_B^j \end{cases} \tag{34}$$

Step 2:

Let $t_1 = t_2 = t$ and evaluate the associated inversion integral to obtain the PDFs of $\hat{w}_T^j$ and $\hat{w}_B^j$ as:

$$\begin{cases} f\left(\hat{w}_T^j\right) = \dfrac{1}{2\pi}\int_{-\infty}^{\infty} e^{-it\hat{w}_T^j} F_1\left(t,\, t\right) dt \\ f\left(\hat{w}_B^j\right) = \dfrac{1}{2\pi}\int_{-\infty}^{\infty} e^{-it\hat{w}_B^j} F_2\left(t,\, t\right) dt \end{cases} \tag{35}$$

Using Eqs. (34) and (35), $f\left(\Delta^j\right)$ and $f\left(\hat{w}^j\right)$ can be expressed as [45]:

Case II: $$\begin{cases} f\left(\Delta^j\right) = \int_{-\infty}^{+\infty} f\left(h_{joint}^C\right) \int_{-\infty}^{+\infty} f\left(\hat{w}_T^j\right) f\left[\hat{w}_B^j\left(\Delta, \hat{w}_T^j, h_{joint}^C\right)\right] d\hat{w}_T^j dh_{joint}^C & \text{if } d_{BC} > t_{BC} \\ f\left(\hat{w}^j\right) = \int_{-\infty}^{+\infty} f\left(w_T^j\right) f\left[w_B^j\left(w_T^j, \hat{w}^j\right)\right] dw_T^j & \text{if } d_{BC} \le t_{BC} \end{cases} \tag{36}$$

where $\hat{w}_B^j\left(\Delta^j, \hat{w}_T^j, h_{joint}^C\right) = \Delta^j + \hat{w}_T^j - h_{joint}^C$ and $w_B^j\left(w_T^j, \hat{w}^j\right) = w_T^j - \hat{w}^j$.

Substituting Eqs. (31), (32) and (36) into Eqs. (29) and (30) yields:

Case I:

$$\Pr\left(gap_{case1}^j > 0\right) = \begin{cases} \int_{-\infty}^{+\infty} \left\{ \begin{array}{l} \left[ \int_{-\infty}^{\hat{w}^j + t_{BC}} \left[ \int_{-\infty}^{+\infty} \left[ f\left(h_T^j\right) f\left[h_B^j\left(\tilde{h}^j, h_T^j\right)\right]\right] dh_T^j \right] d\tilde{h}^j \right] \\ \times \left[ \int_{-\infty}^{+\infty} f\left(w_T^j\right) f\left[w_B^j\left(w_{B1}^j\right)\right] dw_T^j + \int_{-\infty}^{+\infty} f\left(w_T^j\right) f\left[w_B^j\left(w_{B2}^j\right)\right] dw_T^j \right] \end{array} \right\} d\hat{w}^j & \text{for } w_T^j - w_B^j > 0 \\ 0 & \text{for } w_T^j - w_B^j \le 0 \end{cases} \tag{37}$$

Case II:

$$\Pr\left(gap_{case2}^j > 0\right) = \begin{cases} \int_{-\infty}^{+\infty} \left\{ \begin{array}{l} \left[ \int_{-\infty}^{\Delta^j} \left[ \int_{-\infty}^{+\infty} f\left(h_T^j\right) f\left[h_B^j\left(\tilde{h}^j, h_T^j\right)\right] dh_T^j \right] d\tilde{h}^j \right] \\ \times \left[ \int_{-\infty}^{+\infty} f\left(h_{joint}^C\right) \int_{-\infty}^{+\infty} f\left(\hat{w}_T^j\right) f\left[w_T^j\left(\Delta^j, \hat{w}_T^j, h_{joint}^C\right)\right] dw_T^j dh_{joint}^C \right] \end{array} \right\} d\Delta^j & \text{if } d_{BC} > t_{BC} \\ \int_{-\infty}^{+\infty} \left\{ \begin{array}{l} \left[ \int_{-\infty}^{t_{BC} - \hat{w}^j} \left[ \int_{-\infty}^{+\infty} f\left(h_T^j\right) f\left[h_B^j\left(\tilde{h}^j, h_T^j\right)\right] dh_T^j \right] d\tilde{h}^j \right] \\ \times \left[ \int_{-\infty}^{+\infty} f\left(w_T^j\right) f\left[w_B^j\left(w_T^j, \hat{w}^j\right)\right] dw_T^j \right] \end{array} \right\} d\hat{w}^j & \text{if } d_{BC} \le t_{BC} \end{cases} \tag{38}$$

Finally, the probability of non-contact open at the $j^{th}$ pad can be obtained by adding the probabilities of $gap_{case1}^j$ and $gap_{case2}^j$, i.e., $\Pr\left(gap^j > 0\right) = \Pr\left(gap_{case1}^j > 0\right) + \Pr\left(gap_{case2}^j > 0\right)$.

#### 2.2.3.2. Stacking Yield Loss Prediction for Multiple Pads

The joint PDFs of the five critical quantities must be considered when the stacking yield loss model for a single pad is extended to predict the stacking yield loss of multiple pads. If the five critical quantities do not have any statistical dependence, the joint PDFs of the quantities are simply a product of their marginal PDFs. In this study, however the warpage values along the pads on the stacking interface have statistical dependences, and they should be incorporated when the joint PDF is determined.

Estimation of the statistical dependence can be very difficult or sometimes impractical if a large number of performance responses are involved [48-50] However, the joint PDF can be determined by MCS if the correlated performance responses can be transformed into a small set of uncorrelated input variables. First, mapping techniques [51] are used to transform the correlated variables into a smaller set of new uncorrelated variables.. Then, MCS is used to generate a large number of samples from the PDFs of the uncorrelated variables. For each sample, the original variables can be determined through inverse transformation. Finally, the joint PDF of original variables can be estimated using the large number of MCS samples. For example, in this study, if *n* dependent performance responses of warpage of top package, $y_1, \ldots, y_n$, can be transformed into independent performance responses, the warpages of top package of $i^{th}$ and $j^{th}$ pad, $y_i, y_j$. Then, the joint PDF can be expressed as:

$$f\left(y_1, \ldots, y_n\right) = f\left[y_1\left(y_i, y_j\right), \ldots, y_i, \ldots y_j, \ldots, y_n\left(y_i, y_j\right)\right] \tag{39}$$

where $f\left(y_1, \ldots, y_n\right)$ is the joint PDF of performance responses, $y_1, \ldots, y_n$ ; $y_1\left(y_i, y_j\right), \ldots, y_n\left(y_i, y_j\right)$ are the performance responses expressed by the independent performance responses, the warpages of top package of $i^{th}$ and $j^{th}$ pad, $y_i, y_j$.

Considering the joint PDFs, the probability of non-contact yield loss, Eqs. (29) and (30), can be written as:

Case I:

$$\Pr\left(gap_{case1}^{j} > 0\right) = \int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}\left\{\int_{-\infty}^{+\infty}\left\{\left[\int_{-\infty}^{\left|\hat{w}^{j}\right|+t_{BC}} f\left(\tilde{h}^{j}, \tilde{h}^{\sim j}\right)d\tilde{h}^{j}\right] f\left(\left|\hat{w}^{j}\right|, \left|w^{\sim j}\right|\right)\right\} d\left|\hat{w}^{j}\right|\right\} d\tilde{h}^{\sim j} d\left|w^{\sim j}\right| \tag{40}$$

and

Case II:

$$\Pr\left(gap_{case2}^{j} > 0\right) = \begin{cases} \int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}\left\{\int_{-\infty}^{+\infty}\left\{\left[\int_{-\infty}^{\Delta^{j}} f\left(\tilde{h}^{j}, \tilde{h}^{\sim j}\right)d\tilde{h}^{j}\right] f\left(\Delta^{j}, \Delta^{\sim j}\right)\right\} d\Delta^{j}\right\} d\tilde{h}^{\sim j} d\Delta^{\sim j} & \text{if } d_{BC} > t_{BC} \\ \int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}\left\{\int_{-\infty}^{+\infty}\left\{\left[\int_{-\infty}^{t_{BC}-\hat{w}^{j}} f\left(\tilde{h}^{j}, \tilde{h}^{\sim j}\right)d\tilde{h}^{j}\right] f\left(\hat{w}^{j}, w^{\sim j}\right)\right\} d\hat{w}^{j}\right\} d\tilde{h}^{\sim j} dw^{\sim j} & \text{if } d_{BC} \le t_{BC} \end{cases} \tag{41}$$

where $\left|\hat{w}^{\sim j}\right|$ is the absolute value of the warpage difference between the top and bottom packages at non-$j^{th}$ pads; $\Delta^{\sim j}$ is the distance between the top and bottom packages at the non-$j^{th}$ pads for $d_{BC} > t_{BC}$; $f\left(\left|\hat{w}^{j}\right|, \left|w^{\sim j}\right|\right)$ is the joint PDF of the absolute value of the warpage difference between the top and bottom packages; $\hat{w}^{\sim j}$ is the warpage difference between the top and bottom packages at the non-$j^{th}$ pads for $d_{BC} \le t_{BC}$; $\tilde{h}^{\sim j}$ is the total solder ball height at the non-$j^{th}$ pads; $f\left(\Delta^{j}, \Delta^{\sim j}\right)$ is the joint PDF of distance between the top and bottom packages for $d_{BC} > t_{BC}$; $f\left(\hat{w}^{j}, w^{\sim j}\right)$ is the joint PDF of the warpage difference between the top and bottom packages for $d_{BC} \le t_{BC}$; and $f\left(\tilde{h}^{j}, \tilde{h}^{\sim j}\right)$ is the joint PDF of the total solder ball height.

Subsequently, the Eqs. (37) and (38) can be written as:

Case I:

$$
\Pr\left(\boldsymbol{gap}_{\boldsymbol{case1}}^{J_k} > 0\right) = \begin{cases} \displaystyle\int_{-\infty}^{+\infty} \left\{ \begin{array}{l} \left[\displaystyle\int_{-\infty}^{\hat{w}^{J_k}+t_{BC}} \left[\int_{-\infty}^{+\infty} \left[ f\left(\boldsymbol{h}_{\boldsymbol{T}}^{J_k}\right) f\left[\boldsymbol{h}_{\boldsymbol{B}}^{J_k}\left(\tilde{\boldsymbol{h}}^{J_k}, \boldsymbol{h}_{\boldsymbol{T}}^{J_k}\right)\right]\right] d\boldsymbol{h}_{\boldsymbol{T}}^{J_k} \right] d\tilde{\boldsymbol{h}}^{J_k} \right] \\ \times \left[ \displaystyle\int_{-\infty}^{+\infty} f\left(\boldsymbol{w}_{T}^{J_k}\right) f\left[\boldsymbol{w}_{\boldsymbol{B}}^{J_k}\left(\boldsymbol{w}_{\boldsymbol{B1}}^{J_k}\right)\right] d\boldsymbol{w}_{\boldsymbol{T}}^{J_k} + \int_{-\infty}^{+\infty} f\left(\boldsymbol{w}_{\boldsymbol{T}}^{J_k}\right) f\left[\boldsymbol{w}_{\boldsymbol{B}}^{J_k}\left(\boldsymbol{w}_{\boldsymbol{B2}}^{J_k}\right)\right] d\boldsymbol{w}_{\boldsymbol{T}}^{J_k} \right] \end{array} \right\} d\hat{\boldsymbol{w}}^{J_k} & \text{for } \boldsymbol{w}_{\boldsymbol{T}}^{J_k} - \boldsymbol{w}_{\boldsymbol{B}}^{J_k} > 0 \\ 0 & \text{for } \boldsymbol{w}_{\boldsymbol{T}}^{J_k} - \boldsymbol{w}_{\boldsymbol{B}}^{J_k} \leq 0 \end{cases} \quad (42)
$$

Each component in Eq. (42) is defined as follows.

- Each variable $J_k$ can be any number between 1 to *n.*
- The symbols with bold font indicate that they are vectors.
- $\boldsymbol{gap}_{\mathbf{case1}}^{J_{1,\ldots,n}} = \left[gap_{case1}^{J_1}, \ldots, gap_{case1}^{J_n}\right]$ and $\boldsymbol{h}_{\boldsymbol{T}}^{J_{1,\ldots,n}} = \left[h_T^{J_1}, \ldots, h_T^{J_n}\right]$ are vectors of $gap_{case1}$ and top solder ball height from the 1st to $n^{th}$ pad, respectively.
- $f\left(\boldsymbol{h}_{\boldsymbol{T}}^{J_{1,\ldots,n}}\right)$ is the joint PDF of $\boldsymbol{h}_{\boldsymbol{T}}^{J_{1,\ldots,n}}$.
- $f\left(\boldsymbol{h}_{\boldsymbol{T}}^{J_k}\right) = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f\left(\boldsymbol{h}_{\boldsymbol{T}}^{J_{1,\ldots,n}}\right) d\boldsymbol{h}_{\boldsymbol{T}}^{J_{\sim k}}$ is the marginal joint PDF of $\boldsymbol{h}_{\boldsymbol{T}}^{J_k}$, where $\boldsymbol{h}_{\boldsymbol{T}}^{J_k}$ is a subset of $\boldsymbol{h}_{\boldsymbol{T}}^{J_{1,\ldots,n}}$, and $\boldsymbol{h}_{\boldsymbol{T}}^{J_{\sim k}}$ is the complement of $\boldsymbol{h}_{\boldsymbol{T}}^{J_k}$.

The same definitions are applied to other joint PDFs.

Case II:

$$
\Pr\left(\boldsymbol{gap}_{\textbf{case2}}^{J_k} > 0\right) =
$$

$$
\begin{cases}
\displaystyle\int_{-\infty}^{+\infty}\left\{\begin{aligned}&\left[\int_{-\infty}^{\Delta^{J_k}}\left[\int_{-\infty}^{+\infty} f\left(\boldsymbol{h}_{\boldsymbol{T}}^{J_k}\right) f\left[\boldsymbol{h}_{\boldsymbol{B}}^{J_k}\left(\tilde{\boldsymbol{h}}^{J_k}, \boldsymbol{h}_{\boldsymbol{T}}^{J_k}\right)\right] d\boldsymbol{h}_{\boldsymbol{T}}^{J_k}\right] d\tilde{\boldsymbol{h}}^{J_k}\right] \\ &\times\left[\int_{-\infty}^{+\infty} f\left(h_{joint}^{C}\right)\int_{-\infty}^{+\infty} f\left(\breve{\boldsymbol{w}}_{\boldsymbol{T}}^{J_k}\right) f\left[\boldsymbol{w}_{\boldsymbol{T}}^{J_k}\left(\boldsymbol{\Delta}^{J_k}, \breve{\boldsymbol{w}}_{\boldsymbol{T}}^{J_k}, h_{joint}^{C}\right)\right] d\boldsymbol{w}_{\boldsymbol{T}}^{J_k}\, dh_{joint}^{C}\right]\end{aligned}\right\} d\boldsymbol{\Delta}^{J_k} & \text{if } d_{BC} > t_{BC} \\[2ex]
\displaystyle\int_{-\infty}^{+\infty}\left\{\begin{aligned}&\left[\int_{-\infty}^{t_{BC}-\hat{\boldsymbol{w}}^{J_k}}\left[\int_{-\infty}^{+\infty} f\left(\boldsymbol{h}_{\boldsymbol{T}}^{J_k}\right) f\left[\boldsymbol{h}_{\boldsymbol{B}}^{J_k}\left(\tilde{\boldsymbol{h}}^{J_k}, \boldsymbol{h}_{\boldsymbol{T}}^{J_k}\right)\right] d\boldsymbol{h}_{\boldsymbol{T}}^{J_k}\right] d\tilde{\boldsymbol{h}}^{J_k}\right] \\ &\times\left[\int_{-\infty}^{+\infty} f\left(\boldsymbol{w}_{\boldsymbol{T}}^{J_k}\right) f\left[\boldsymbol{w}_{\boldsymbol{B}}^{J_k}\left(\boldsymbol{w}_{\boldsymbol{T}}^{J_k}, \hat{\boldsymbol{w}}^{J_k}\right)\right] d\boldsymbol{w}_{\boldsymbol{T}}^{J_k}\right]\end{aligned}\right\} d\hat{\boldsymbol{w}}^{J_k} & \text{if } d_{BC} \le t_{BC}
\end{cases}
\tag{43}
$$

The summation of Eqs. (42) and (43) can express as $\Pr\left(\left(gap^{J_1} > 0\right) \cap \cdots \cap \left(gap^{J_k} > 0\right)\right)$, which represents the intersection of the events that gaps occur simultaneously on *k* pads.

#### 2.2.3.3. Stacking Yield Loss Prediction Model

The final stacking yield loss can be calculated as the probability of the union of gap occurrence for all pads on the stacking interface. Considering a condition of $gap^j > 0$ as an event, the yield loss can be obtained using the general formula of the probability of the union of *n* events. Figure 24 illustrates the probability of the union of three events and the probabilities of intersections of events.

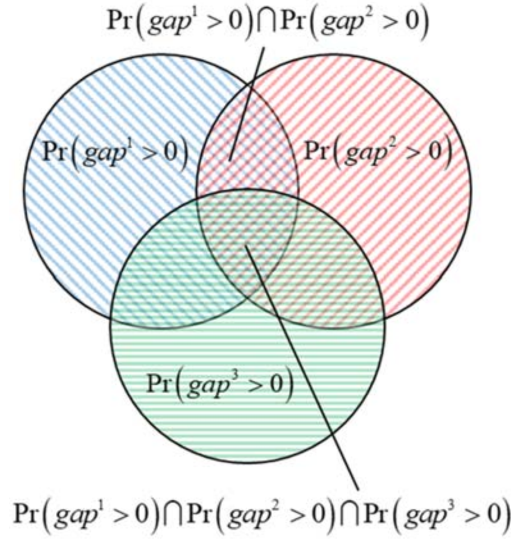



Figure 24. Schematic illustration of the union of three events

The probability of the union of *n* events can be expressed as [52]:

$$\begin{aligned}&\Pr\left(\bigcup_{j=1}^{n}\left(gap^{j}>0\right)\right)\\&=\sum_{k=1}^{n}(-1)^{k+1}\sum_{\substack{J_1,J_2,\ldots,J_k\\1\le J_1<J_2<\cdots<J_k\le n}}\Pr\left(\left(gap^{J_1}>0\right)\cap\left(gap^{J_2}>0\right)\cap\cdots\cap\left(gap^{J_k}>0\right)\right)\end{aligned}\tag{44}$$

where each variables $J_1, J_2..., J_k$ can be any number between 1 to *n*; the probabilities of the pair-wise intersections, $\Pr\left(\left(gap^{J_1}>0\right)\cap\left(gap^{J_2}>0\right)\right)$, to the *n*-tuple-wise intersections, $\Pr\left(\left(gap^{J_1}>0\right)\cap\cdots\cap\left(gap^{J_n}>0\right)\right)$, can be calculated by the summation of Eqs. (42) and (43). Finally, the total stacking yield loss after assembly can be obtained by substituting the summation of Eqs. (42) and (43) into the Eq. (44).

### 2.2.4. Conclusion

A comprehensive stochastic model for PoP stacking yield loss prediction was proposed. The model utilized the PDFs of five critical performance responses (the warpages and the solder ball heights of the top and bottom packages, and the solder joint height of the corner pad) to take into account their statistical variations. The approach used in the model was capable of handling a large number of input variables. The results can be used effectively to control the input uncertainties, and thus to achieve a yield goal for a given set of PoP designs.

## 2.3. Stacking Yield Prediction of Package-on-Package Assembly Using Uncertainty Propagation Analysis: Part II Implementation of Stochastic Model

### 2.3.1. Introduction

Section 2.2 described in detail a comprehensive stochastic model to predict the packaging-on-package (PoP) stacking yield loss. To cope with the limitations of the existing models, the proposed model took into account all pad locations at the stacking interface while considering the statistical variations of the warpages, heights of solder balls and corner solder joint. The goal was achieved by employing three statistical methods: (1) advanced approximate integration-based method called eigenvector dimension reduction (EDR) method to conduct uncertainty propagation (UP) analysis; (2) the stress-strength interference (SSI) model to determine the probability of non-contact open at a single pad; and (3) the union of events considering the statistical dependence to calculate the final yield loss.

In this companion section, the stochastic model is implemented for a PoP, which consists of a stacked die thin flat ball grid array (TFBGA) as the top package and a flip chip ball grid array (fcBGA) as the bottom package. The configuration is shown schematically in Figure 19. The top TFBGA package and the bottom fcBGA package are connected through 216 solder joints of 0.5 mm pitch in two peripheral rows.

The probability density functions (PDFs) of the top and bottom package warpages are determined first in Section 2.3.2. Determination of the heights and their PDFs of solder ball and corer solder joint are presented in Section 2.3.3. Stacking yield loss prediction from the PDFs is provided in Section 2.3.4.

### 2.3.2. Probability Density Functions of Warpages

The warpages of the top and bottom packages are predicted by a finite element analysis (FEA). The results are used to calculate the PDFs using the UP analysis.

### 2.3.2.1. Warpage Prediction

A quarter symmetry was used to build a finite element model using a commercial FEA package (ANSYS®). Figure 7 and Figure 25 show details of the models with the boundary conditions of top TFBGA package and bottom fcBGA package, respectively. The top TFBGA package contains two dies. The die stack configuration is shown using white dashed lines.

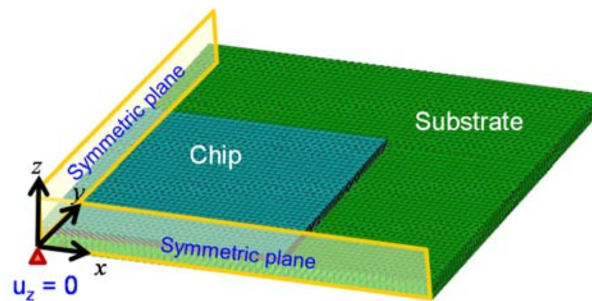



Figure 25. Details of the FEA model of bottom fcBGA package

The material properties and the nominal dimensions used in the models are summarized in Table 9 and Table 10. For the top package, the EMC was modeled as a linear viscoelastic material, and the temperature dependent Young's modulus was considered for the die attach film (DAF). Details about these nonlinear materials are described in Figure 8 and Figure 9. For the bottom package, the flip-chip solder bumps (SAC305) [53] and the underfill [54] were treated as a homogenous layer, and its effective material properties were calculated based on the modified rule of mixture [55].

Table 9  Material properties of top TFBGA and bottom fcBGA

| | Material | Young's modulus (GPa) | Poisson's Ratio | CTE (ppm/°C) | | $T_g$ (°C) |
|---|---|---|---|---|---|---|
| | | | | $\alpha_1$ (< $T_g$) | $\alpha_2$ (> $T_g$) | |
| TFBGA | Silicon die | 130 | 0.23 | 2.8 | | -- |
| | DAF | Temp. dependent | 0.3 | 65.3 | 162.9 | 138 |
| | Substrate | 46.794 | 0.3 | 16.2 (in-plane)<br>61.5 (out-of-plane) | | -- |
| | EMC | Viscoelastic | 0.21 | 9.12 | 35.13 | 137.5 |
| fcBGA | Silicon die | 130 | 0.23 | 2.8 | | -- |
| | Solder and UF (effective) | 28.767 @ 25°C<br>2.759 @ 260°C | 0.3425 | 23.2 | 38.9 | 80 |
| | Substrate | 32.078 @ 25°C<br>21.116 @ 260°C | 0.3 | 12.58 (in-plane)<br>30 (out-of-plane) | 10.69 (in-plane)<br>30 (out-of-plane) | 190 |

Table 10  Dimensions of top TFBGA and bottom fcBGA

| | | length × width × thickness |
|---|---|---|
| TFBGA | 1st Die (mm) | 13 × 11 × 0.575 |
| | 1st DAF (mm) | 13 × 11 × 0.025 |
| | 2nd Die (mm) | 11 × 9 × 0.575 |
| | 2nd DAF (mm) | 11 × 9 × 0.025 |
| | Substrate (mm) | 15 × 15 × 0.13 |
| | EMC (mm) | 15 × 15 × 0.59 |
| fcBGA | Die (mm) | 9.3 × 9.3 × 0.127 |
| | Solder bump + UF (mm) | 9.3 × 9.3 × 0.075 |
| | Substrate (mm) | 15 × 15 × 0.276 |

The top package was subjected to the EMC molding process at 175 °C, which was used as a stress free temperature. On the other hand, the underfill temperature (125 °C) was used as a stress free temperature for the bottom package. To analyze the non-contact open during stacking, the conventional lead-free solder reflow profile with the peak temperature as 260 °C was

considered [28]. Figure 26 shows the deformed configurations of the top and the bottom packages with the nominal design parameters at the peak reflow temperature. In (b), the white circles represent the solder ball locations at the stacking interface. Based on the sign convention of warpage values suggested by JEDEC [56], the maximum warpages of the top and bottom packages with the nominal design parameters occurred at the outmost corner pad with warpage value of 76.2 μm and −83.0 μm, respectively. The stress-free temperatures were confirmed by comparing them with the experimental warpage data of nearly identical packages found in the literature [57-59].

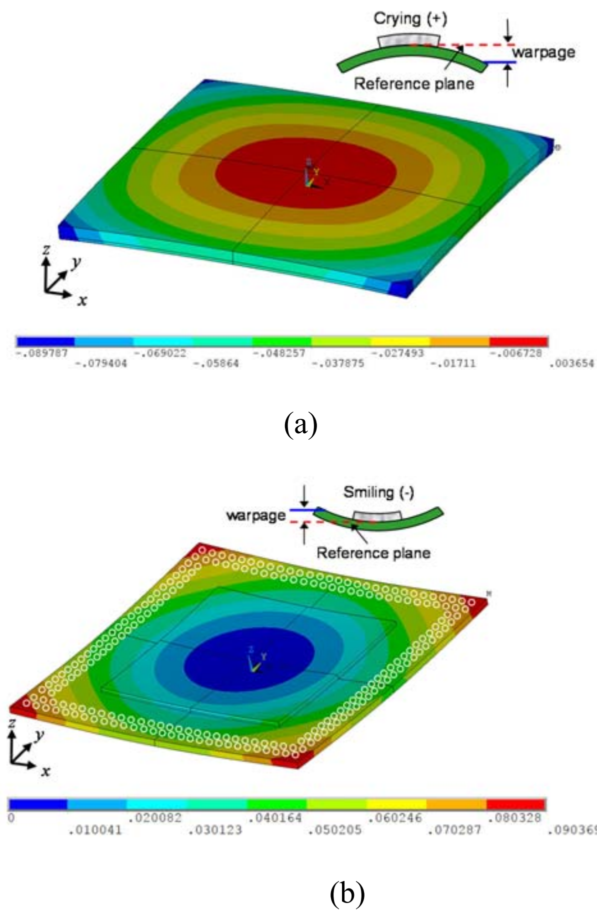



Figure 26. Deformed configurations (5x magnification) of (a) the top and (b) the bottom packages with the nominal design parameters at the peak reflow temperature

#### 2.3.2.2. Uncertainty Propagation Analysis for Warpage PDFs

A total of 12 input variables were considered for the top TFBGA package, and a total of 9 input variables for the bottom fcBGA package. Their means and standard deviations were obtained from the literature as well as the manufacturing specifications [31, 60-63], and they are summarized in Table 3 and Table 11.

Table 11 Input variables of bottom fcBGA

| Variables | Physical meaning | Mean | Std. Dev. | Distribution | Correlation coefficient |
|---|---|---|---|---|---|
| $x_{13}$ | PKG length (mm) | 15 | 0.067 | Normal | -- |
| $x_{14}$ | PKG width (mm) | 15 | 0.067 | Normal | -- |
| $x_{15}$ | Die thickness (mm) | 0.127 | 0.002 | Normal | -- |
| $x_{16}$ | Bump/underfill thickness (mm) | 0.075 | 0.0025 | Normal | -- |
| $x_{17}$ | Substrate thickness (mm) | 0.276 | 0.01 | Normal | -- |
| $x_{18}$ | Substrate CTE below $T_g$ (ppm/°C) | 12.58 | 0.63 | Bivariate Normal | 1 |
| $x_{19}$ | Substrate CTE above $T_g$ (ppm/°C) | 10.69 | 0.53 | | |
| $x_{20}$ | Substrate modulus @ 25°C (GPa) | 32.1 | 0.32 | Bivariate Normal | 1 |
| $x_{21}$ | Substrate modulus @ 260°C (GPa) | 21.1 | 0.21 | | |

Among the 12 input variables of the top package, the EMC thickness, $x_3$, and the substrate thickness, $x_4$, have statistical correlation because their sum should be equal to the cavity height of the transfer mold [7]. In addition, the CTEs of the EMC below and above $T_g$, $x_9$ and $x_{10}$, are measured from the same samples, and thus, they should have strong positive correlation, which can be presented by a correlation coefficient of “unity”. Similarly, the CTEs of the substrate above and below $T_g$ (i.e., $x_6$ and $x_7$) and the elastic moduli of the substrate at 25 °C and 260 °C (i.e., $x_8$ and $x_9$) of the bottom package are also given a correlation coefficient of “unity” [7].

Next, the PDFs of the warpages of the top and bottom packages were obtained using the EDR method with 4*N*+1 sampling scheme. A total of 49 and 37 modeling runs were conducted for the 12 input variables of the top package and the 9 input variables of the bottom package, respectively. The detailed procedure of constructing the PDFs can be found in Ref. [28] and [22], and it is illustrated below using the outmost corner pad of the top package.

**Step 1: Covariance Matrix**

The covariance matrix was constructed using the information in Table 3 and expressed as Eq. (13).

**Step 2: Eigenvalues and Eigenvectors of Covariance Matrix**

By solving the eigenvalue problem of the covariance matrix (i.e., $\Sigma X' = \lambda X'$), the eigenvalues, $\lambda$, and the corresponding eigenvectors, $X'$, were obtained. The results are listed in Table 12, which shows that the eigenvector directions of the two pairs $\left(\text{i.e.,} \left(x'_3,\, x'_4\right)\right.$ and $\left.\left(x'_9,\, x'_{10}\right)\right)$ were altered due to the correlations.

**Step 3: Sampling Points**

Using the 4*N*+1 sampling scheme, five sample points along each eigenvector direction were determined by using Eqs. (8) and (9) in Ref. [7]. Table 13 shows the sample points and the corresponding predicted warpage values along the $x'_3$ and $x'_{11}$ directions as examples for correlated input variable and uncorrelated input variable, respectively.

**Step 4: Integration Points**

In this study, 21 integration points were used for each one-dimensional integration to increase the accuracy. The locations and weights of the integration points were calculated by the moment based quadrature rule [21] (Eq. (8) in Ref. [22]). Table 14 lists the weights and locations

of integration points along the $x'_3$ and $x'_{11}$ directions, which represent the uncorrelated and correlated cases, respectively.

Table 12 Eigenvalues and eigenvectors of the covariance matrix

| $i^{th}$ eigenvalue | | $i^{th}$ eigenvector projected into the original directions of input variables $x'_i = [x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}, x_{12}]^T$ |
|---|---|---|
| $\lambda_1$ | $0.033^2$ | $[1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]^T$ |
| $\lambda_2$ | $0.033^2$ | $[0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]^T$ |
| $\lambda_3$ | 85.5 | $[0, 0, 0.99, -0.13, 0, 0, 0, 0, 0, 0, 0, 0]^T$ |
| $\lambda_4$ | 8.6 | $[0, 0, 0.13, 0.99, 0, 0, 0, 0, 0, 0, 0, 0]^T$ |
| $\lambda_5$ | 1 | $[0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0]^T$ |
| $\lambda_6$ | 1 | $[0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0]^T$ |
| $\lambda_7$ | $3.75^2$ | $[0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0]^T$ |
| $\lambda_8$ | $3.75^2$ | $[0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0]^T$ |
| $\lambda_9$ | $4.38^2$ | $[0, 0, 0, 0, 0, 0, 0, 0, 0.97, 0.25, 0, 0]^T$ |
| $\lambda_{10}$ | 0 | $[0, 0, 0, 0, 0, 0, 0, 0, -0.25, 0.97, 0, 0]^T$ |
| $\lambda_{11}$ | $0.81^2$ | $[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0]^T$ |
| $\lambda_{12}$ | $159^2$ | $[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1]^T$ |

Table 13 Locations and predicted warpages of sample points along the 3rd eigenvector and the 11th eigenvector

| $x'_i$ | Sample point | ${}^1V_i$ | ${}^2V_i$ | ${}^0V_i$ | ${}^4V_i$ | ${}^5V_i$ |
|---|---|---|---|---|---|---|
| | Performance response | $y({}^1V_i)$ | $y({}^2V_i)$ | $y({}^0V_i)$ | $y({}^4V_i)$ | $y({}^5V_i)$ |
| $x'_3$ | $V_3$ | $x_3 = 677$; $x_4 = 118$ | $x_3 = 634$; $x_4 = 124$ | $x_3 = 590$; $x_4 = 130$ | $x_3 = 547$; $x_4 = 136$ | $x_3 = 503$; $x_4 = 142$ |
| | $y(V_3)$ | 101.9 | 90.4 | 76.2 | 58.5 | 35.0 |
| $x'_{11}$ | $V_{11}$ | $x_{11} = 13.77$ | $x_{11} = 14.985$ | $x_{11} = 16.2$ | $x_{11} = 17.415$ | $x_{11} = 18.63$ |
| | $y(V_{11})$ | 102.5 | 89.4 | 76.2 | 63.1 | 50.0 |

Table 14 Locations, weights, and warpage values of integration points along the 3rd eigenvector and the 11th eigenvector

| | 3rd eigenvector, $x'_3$ | | | 11th eigenvector, $x'_{11}$ | | |
|---|---|---|---|---|---|---|
| | $x_{k,3'}$ | $w_{k,3'}$ | $y(x_{k,3'})$ | $x_{k,11'}$ | $w_{k,11'}$ | $y(x_{k,11'})$ |
| 1 | $x_3 = 366; x_4 = 160$ | 2.5E-14 | 116.0 | $x_{11} = 9.84$ | 2.5E-14 | 144.8 |
| 2 | $x_3 = 398; x_4 = 156$ | 5.0E-11 | 116.6 | $x_{11} = 10.73$ | 5.0E-11 | 135.2 |
| 3 | $x_3 = 424; x_4 = 152$ | 1.5E-08 | 115.2 | $x_{11} = 11.48$ | 1.5E-08 | 127.1 |
| 4 | $x_3 = 448; x_4 = 149$ | 1.2E-06 | 112.5 | $x_{11} = 12.15$ | 1.2E-06 | 119.9 |
| 5 | $x_3 = 470; x_4 = 146$ | 4.2E-05 | 108.9 | $x_{11} = 12.79$ | 4.2E-05 | 113.1 |
| 6 | $x_3 = 491; x_4 = 143$ | 0.00071 | 104.8 | $x_{11} = 13.39$ | 0.00071 | 106.6 |
| 7 | $x_3 = 512; x_4 = 141$ | 0.00644 | 100.0 | $x_{11} = 13.97$ | 0.00644 | 100.3 |
| 8 | $x_3 = 532; x_4 = 138$ | 0.03395 | 94.8 | $x_{11} = 14.54$ | 0.03395 | 94.2 |
| 9 | $x_3 = 551; x_4 = 135$ | 0.10839 | 89.1 | $x_{11} = 15.10$ | 0.10839 | 88.2 |
| 10 | $x_3 = 571; x_4 = 133$ | 0.21533 | 83.0 | $x_{11} = 15.65$ | 0.21533 | 82.2 |
| 11 | $x_3 = 590; x_4 = 130$ | 0.27026 | 76.2 | $x_{11} = 16.20$ | 0.27026 | 76.2 |
| 12 | $x_3 = 609; x_4 = 127$ | 0.21533 | 68.8 | $x_{11} = 16.75$ | 0.21533 | 70.3 |
| 13 | $x_3 = 629; x_4 = 125$ | 0.10839 | 60.5 | $x_{11} = 17.30$ | 0.10839 | 64.3 |
| 14 | $x_3 = 648; x_4 = 122$ | 0.03395 | 50.8 | $x_{11} = 17.86$ | 0.03395 | 58.3 |
| 15 | $x_3 = 668; x_4 = 119$ | 0.00644 | 39.7 | $x_{11} = 18.43$ | 0.00644 | 52.2 |
| 16 | $x_3 = 689; x_4 = 117$ | 0.00071 | 27.0 | $x_{11} = 19.01$ | 0.00071 | 45.9 |
| 17 | $x_3 = 710; x_4 = 114$ | 4.2E-05 | 12.8 | $x_{11} = 19.61$ | 4.2E-05 | 39.4 |
| 18 | $x_3 = 732; x_4 = 111$ | 1.2E-06 | -3.3 | $x_{11} = 20.25$ | 1.2E-06 | 32.6 |
| 19 | $x_3 = 756; x_4 = 108$ | 1.5E-08 | -21.7 | $x_{11} = 20.92$ | 1.5E-08 | 25.4 |
| 20 | $x_3 = 782; x_4 = 104$ | 5.0E-11 | -43 | $x_{11} = 21.67$ | 5.0E-11 | 17.4 |
| 21 | $x_3 = 814; x_4 = 100$ | 1.7E-14 | -69.9 | $x_{11} = 22.56$ | 1.7E-14 | 7.9 |
| First statistical moment | 75.4 | | | 16.2 | | |

To complete each one-dimensional integration, the performance responses at integration points were obtained by interpolation and extrapolation from the performance responses at the

sample points using the moving least square (MLS) method [22]. Table 14 also lists the performance responses at the integration points along the $x'_3$ and $x'_{11}$ directions.

**Step 5: Statistical moments of performance response**

The 1st statistical moment along the $i^{th}$ eigenvector direction was determined from the sum of the performance response ($y$) times the weight ($w$) at each integration point. Similarly, the higher order statistical moments were determined from the same procedure using $y^2$, $y^3$, and $y^4$. The 1st statistical moment along the $x'_3$ and $x'_{11}$ directions are shown in Table 14. Finally, the statistical moments of the warpage PDF were calculated by the recursive formula using all statistical moments (Eqs. (21) and (22) in Ref. [21]). The four statistical moments of the warpage PDF at the outmost corner pad are listed in Table 15.

Table 15 The 1st to the 4th statistical moments and coefficients of the stabilized Pearson distribution of the warpage at the outmost corner pad

| Statistical moments | | Mean | Std. Dev. | Skewness | Kurtosis |
|---|---|---|---|---|---|
| | | 39.68 | 19.48 | -0.0488 | 3.0053 |
| Coefficients of stabilized Pearson distribution | | $c_0$ | $c_1$ | $c_2$ | -- |
| | 1st hyper PDF | 1118.1745 | -0.8705 | -0.0091 | |
| | 2nd hyper PDF | 1068.3696 | -0.8192 | 0.0061 | |

**Step 6: Construction of PDF using stabilized Pearson system**

The four statistical moments were input to Eq. (17) in Ref.[7], and the coefficients of the Pearson system were determined ($c_0 = 1084.1438$, $c_1 = -0.8355$, and $c_2 = 0.0013$). These values were close to the boundary of the Pearson type I, IV and V distributions, which would cause numerical instability in constructing the PDF [22]. The stabilized Pearson system was subsequently employed to determine the PDF. Table 15 also lists the coefficients of two hyper-

PDFs obtained from the stabilized Pearson system. Figure 27 shows the PDFs of the warpages at the outmost corner pad for the top and bottom package, respectively.

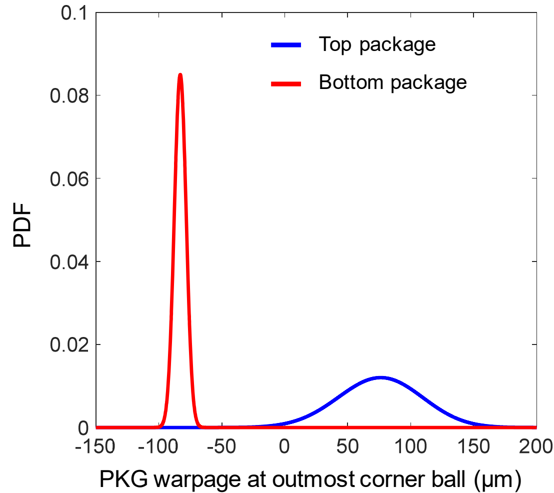



Figure 27. PDFs of warpage at the outmost corner pad

The above procedures were repeated to determine the warpage PDF of all pads of the top and bottom packages. Due to the quarter symmetry of the model, the PDFs of only 56 pads (out of 216 pads) had to be determined for each package (a total of 112 PDFs). It is important to recall that the warpages at different pads are statistically correlated. To avoid the overestimation of a yield loss, a joint PDF of the warpages at the 56 pads was determined using Eq. (19) in Ref. [64].

The statistical correlations were determined using the results of warpage predictions. In each modeling run, the out-of-plane displacements of pads were extracted. Figure 28 illustrates how the correlations are obtained, where the warpages at four pads shown in (a) are plotted against the warpage at the outmost corner pad for the top package (b) and the bottom package (c). Both top and bottom packages clearly show strong linear correlations; the correlation coefficients of the warpage at the $j^{th}$ pad to the warpage at the outmost corner pad are virtually unity. Therefore, the warpage at the $j^{th}$ pad can be calculated from the linear relationship with the warpage at the outmost corner pad.

For the condition of Case II (the maximum warpage difference at the center), the correlations between the warpage at the die edge and the warpage at the outmost corner pad were required. The results are also shown in Figure 28(b) and (c), and the correlation coefficients are virtually unity for both top and bottom packages.

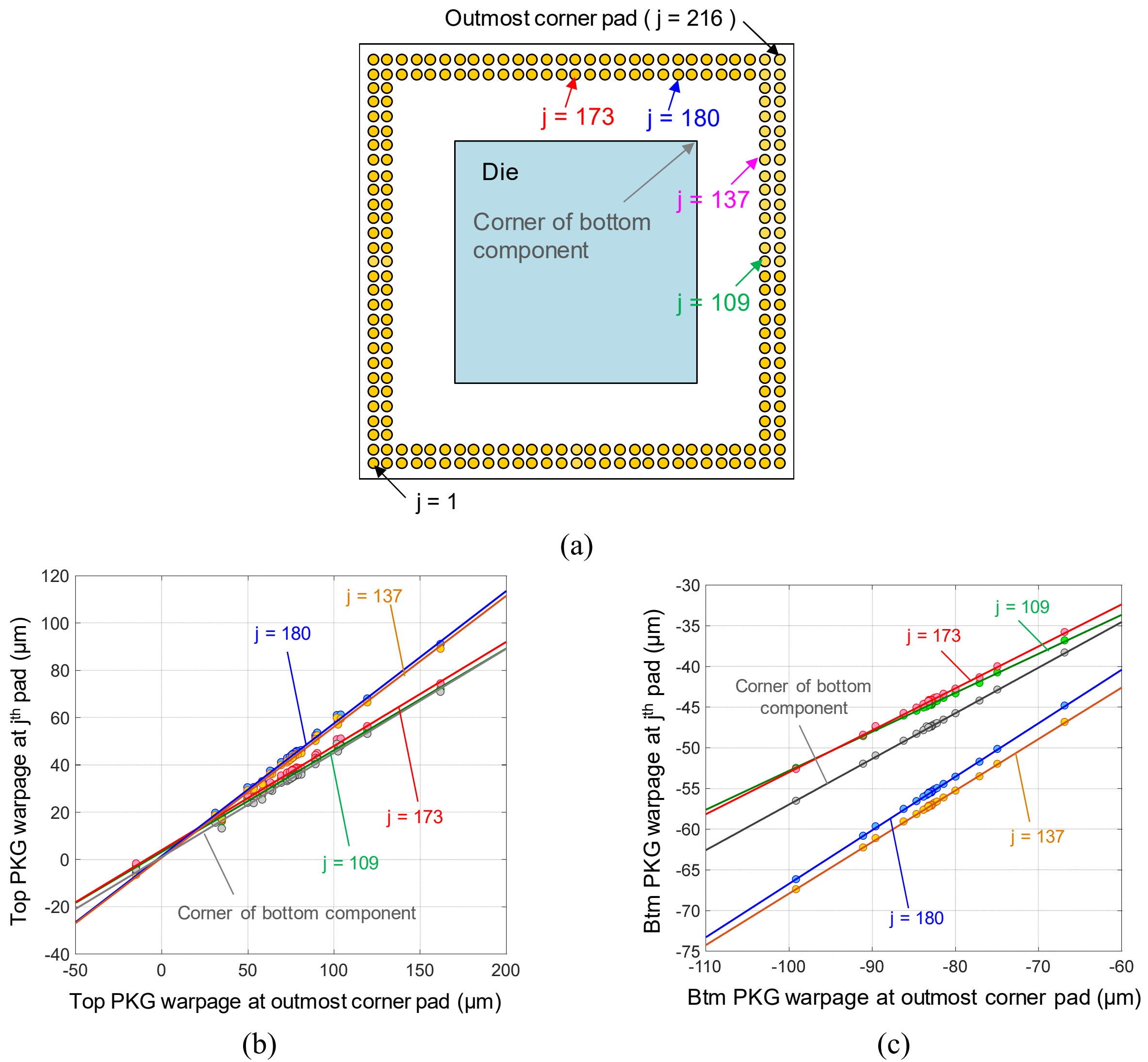



Figure 28. Statistical correlations of the warpages at the $j^{th}$ pad and the outmost corner pad: (a) pad locations, (b) top package, and (c) bottom package

### 2.3.3. Probability Density Functions of Solder Balls and Joint Heights

The final shapes of solder balls and joints are predicted by a program called Surface Evolver [65, 66]. The results are used to calculate the PDFs using the UP analysis.

#### 2.3.3.1. Prediction of Solder Ball Heights and Joint Heights at Corner Pad

For the fine pitch PoP used in this study, a technology called “ball on ball (BoB)” [67-69] is used for package stacking. The technology mounts solder balls on the topside of the bottom package (referred to as the bottom solder balls); these solder balls are aligned with the solder balls on the top package (referred to as the top solder balls). The bottom solder balls and the corresponding top solder balls form solder joints after stacking. The BoB method was developed to reduce the impact of excessive warpages on the stacking yield while providing a sufficient distance between the top and bottom packages to accommodate the component height of the bottom package. Figure 29 illustrates a PoP with the BoB approach just prior to reflow.

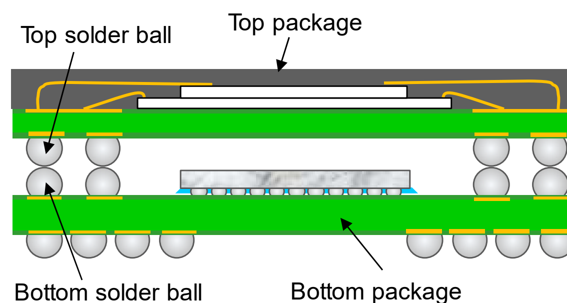



Figure 29. Cross-sectional view of a PoP using the BoB approach just prior to reflow

Figure 30(a) illustrates the solder balls after reflow. During the reflow process, the solder balls melt and form a final shape after filling the solder resist opening (SRO). The solder volume, pad dimensions, surface tension, component weights, etc. contribute to the final stand-off height of the solder ball [2]. During stacking, the top and bottom solder balls go through the second reflow process (stacking) and form a solder joint, as illustrated in Figure 30(b).

The final shapes of the solder balls and the solder joint were predicted by the Surface Evolver, which is based on the principle of surface energy minimization considering the energy contributed by the surface tension, gravitational energy and external loading.

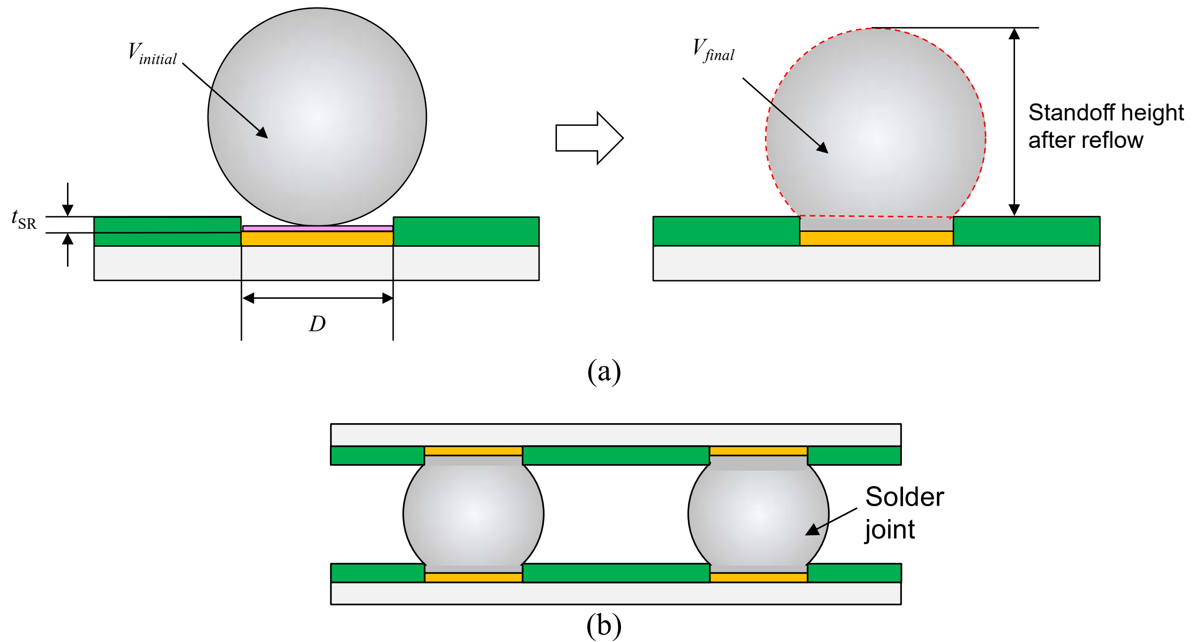


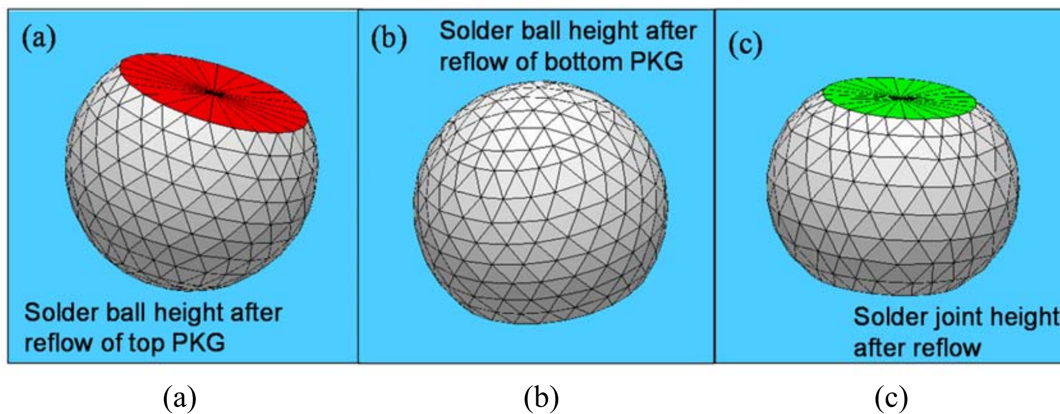

Figure 30. Solder balls after (a) the first reflow process and (b) the second reflow during stacking

For a PoP package with a given solder material, the surface tension and gravitational energy can be assumed to be constant, and they do not cause solder ball height variations [24]. Therefore, the solder ball standoff height after reflow is mainly affected by the geometrical parameters. In this case study, the solder ball pads of the top and bottom package are both solder mask defined (SMD). Figure 30(a) also shows the geometrical parameters considered for the solder ball standoff height prediction for the SMD design. Due to the constraint of the SR opening, the portion of the solder filled within the SR opening does not contribute to the solder ball stand-off height after reflow, and thus, the solder volume after reflow can be express as:

$$V_{final} = V_{initial} - \pi \left( \frac{D}{2} \right)^2 t_{SR} \tag{45}$$

where $V_{final}$ is the solder volume to form the solder ball stand-off height after reflow (red dashed area in Figure 30(b), $V_{initial}$ is the preformed solder ball volume, $D$ is the diameter of SR opening, and $t_{SR}$ is the SR thickness.

Figure 31(a) and (b) show the shape prediction results of the top and bottom solder balls after the 1st reflow. Figure 31(c) shows the solder joint after the 2nd reflow. The results were obtained using the nominal design parameters. The collapse heights of the top and bottom solder balls were 202.3 μm and 199.8 μm, respectively. The solder joint height after the 2nd reflow was 228.9 μm.




(a) (b) (c)

Figure 31. Shape prediction results of (a) the top and (b) bottom solder balls after the 1st reflow; and (c) the solder joint after the 2nd reflow

#### 2.3.3.2. Uncertainty Propagation Analysis for Solder Ball Heights and Joint Heights at Corner Pad

Table 16 and Table 17 list the input variables used to predict the top and bottom solder ball heights, which are uncorrelated. The means and standard deviations of these variables were obtained from the literature and the manufacturing specifications [8, 62, 70, 71]. The EDR method with $4N$+1 sampling scheme was used. For the solder balls, 13 modeling runs were conducted for each package. For the joint heights at the corner pad, all 6 input variables had to be considered

since both top and bottom solder balls affected the solder joint height, which required 25 modeling runs. Therefore, the total modeling runs for the UP analysis of solder balls and heights was 51.

Table 16 Input variables of top solder ball

| Variables | Physical meaning | Mean | Std. Dev. | Distribution |
|---|---|---|---|---|
| $x_{22}$ | Top solder ball diameter (μm) | 175 | 1.67 | Normal |
| $x_{23}$ | Top PKG SR open diameter (μm) | 230 | 16.67 | Normal |
| $x_{24}$ | Top PKG SR thickness (μm) | 15 | 1.05 | Normal |

Table 17 Input variables of bottom solder ball

| Variables | Physical meaning | Mean | Std. Dev. | Distribution |
|---|---|---|---|---|
| $x_{25}$ | Btm solder ball diameter (μm) | 175 | 1.67 | Normal |
| $x_{26}$ | Btm PKG SR open diameter (μm) | 230 | 16.67 | Normal |
| $x_{27}$ | Btm PKG SR thickness (μm) | 20 | 1.40 | Normal |

Since the input variables were not correlated, the first two steps of the EDR method (section 2.2) were not necessary. Steps 3 to 6 were repeated to determine the required PDFs. A total of 532 PDFs were calculated for the solder ball heights at all pads of the top and bottom packages (216 each), and one additional PDF was calculated for the solder joint height at the corner pad. It is to be noted that the quarter symmetry used in the warpage prediction was not applicable to solder ball cases. The joint PDF of solder ball heights at the 216 pads of the top package is the multiplication of the PDFs of the solder ball height of each pad, and so is the joint PDF of solder ball heights for the bottom package.

Table 18 summarizes the predicted statistical moments and the corresponding coefficients of the stabilized Pearson system of the heights of the solder balls and the corner solder joint. Figure

32 shows the PDFs of the top and bottom solder ball height of the $j^{th}$ pad and the PDF of the solder joint at the corner pad.

Table 18 Predicted statistical moments and the corresponding coefficients of the stabilized Pearson distribution for top and bottom solder balls as well as solder joint

<table>
<tr><th colspan="2">Statistical moments</th><th>Mean</th><th>Std. Dev.</th><th>Skewness</th><th>Kurtosis</th></tr>
<tr><td colspan="2">Top ball height</td><td>202.57</td><td>8.81</td><td>-0.13</td><td>3.03</td></tr>
<tr><td colspan="2">Bottom ball height</td><td>200.00</td><td>9.25</td><td>-0.14</td><td>3.03</td></tr>
<tr><td colspan="2">Solder joint height</td><td>228.57</td><td>8.35</td><td>-0.13</td><td>3.03</td></tr>
<tr><td colspan="2">Coefficients of stabilized Pearson distribution</td><td>$c_0$</td><td>$c_1$</td><td>$c_2$</td><td rowspan="7">--</td></tr>
<tr><td rowspan="2">Top ball height</td><td>1st hyper PDF</td><td>80.1262</td><td>-0.5974</td><td>-0.0105</td></tr>
<tr><td>2nd hyper PDF</td><td>76.5228</td><td>-0.5619</td><td>0.0050</td></tr>
<tr><td rowspan="2">Bottom ball height</td><td>1st hyper PDF</td><td>88.1799</td><td>-0.6547</td><td>-0.0105</td></tr>
<tr><td>2nd hyper PDF</td><td>84.2124</td><td>-0.6158</td><td>0.0049</td></tr>
<tr><td rowspan="2">Solder joint height</td><td>1st hyper PDF</td><td>71.9823</td><td>-0.5675</td><td>-0.0105</td></tr>
<tr><td>2nd hyper PDF</td><td>68.7456</td><td>-0.5338</td><td>0.0050</td></tr>
</table>

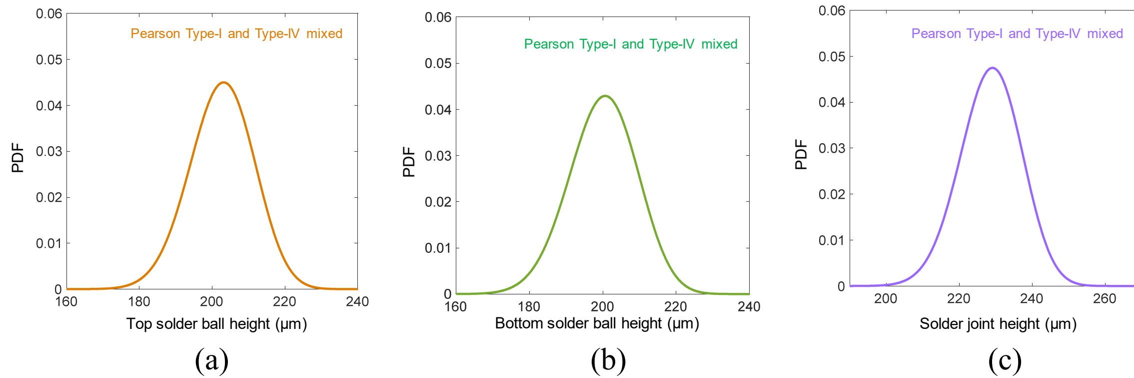



Figure 32. (a) PDF of the top solder ball height, (b) PDF of the bottom solder ball height, and (c) PDF of the solder joint height

### 2.3.4. Stacking Yield Loss Prediction

The procedure to solve the convolutions in the proposed yield model using a single MCS run is presented first. The estimation of true yield loss using multiple MCS runs is followed.

#### 2.3.4.1. Stacking Yield Loss Prediction using a Single MCS Run

Once the joint PDFs of the five critical quantities are obtained, the PDFs of load and strength can be formed by combining these PDFs. Then, the stacking yield loss can be predicted using Eq. (22), (23) and (24) in Ref. [64]. Analytical methods to solve the convolutions of the distribution forms (i.e., Pearson's distributions) in the proposed yield loss model were not available, and the convolutions were calculated using MCS. The procedure is listed as follows. It is also illustrated in Figure 33.

Step 1 Draw the $N_{MCS}$ number of samples: for each MCS sample, 216 sets of load and strength values are randomly generated using the joint PDFs of package warpages, solder ball heights and corner pad solder joint height. In the figure, the rows in the table represent the $k^{th}$ MCS sample.

Step 2 Form the PDFs of load and strength at the $j^{th}$ pad using all MCS samples along the $j^{th}$ column.

Step 3 Calculate the probability of non-contact open at the $j^{th}$ pad by the statistical interference between the PDFs of load and strength.

Step 4 Determine the stacking yield loss as the probability of the union of the gap occurrence in every pad across the stacking interface.

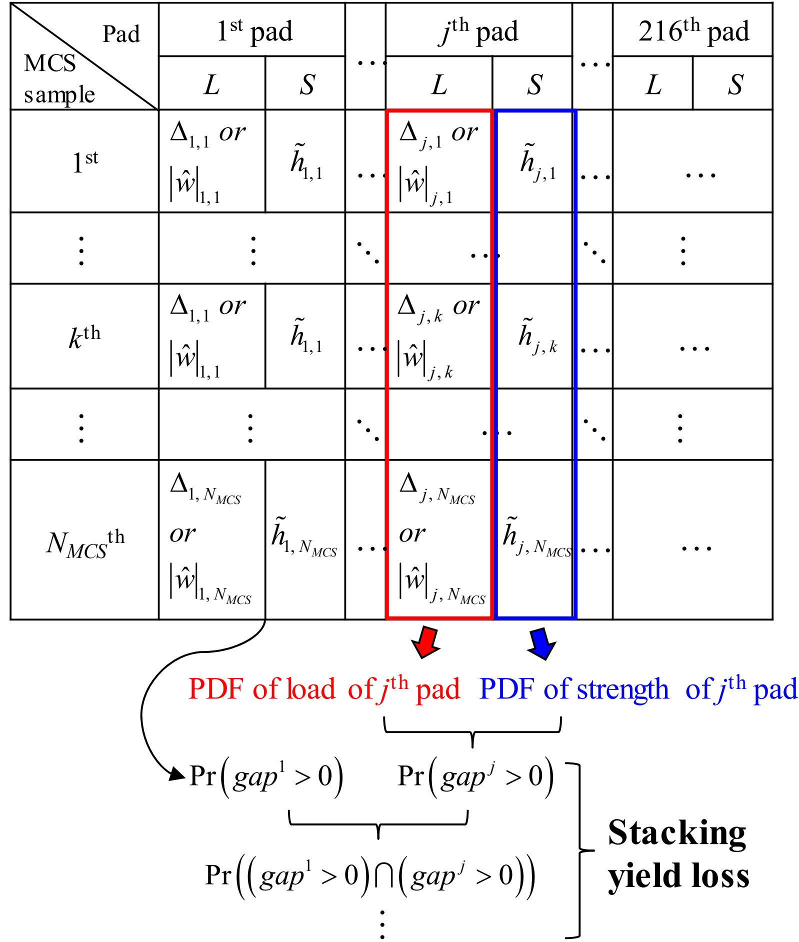



Figure 33. Procedure of stacking yield prediction by MCS

Figure 34 illustrates the PDFs of load and strength for the 109th pad ($j$ = 109) for Steps 2 and 3. Two stacking scenarios (scenario-1 and scenario-4) were observed for the PoP configuration of this study, and they are shown in Figure 34(a) and (b), respectively. In (a), the PDFs of load and strength (i.e., the PDF of the absolute value of the warpage difference, $f\left(|\hat{w}|\right)$, and the total solder ball height, $f\left(\tilde{h}\right)$) were obtained by 1,292 samples among the 100,000 samples.

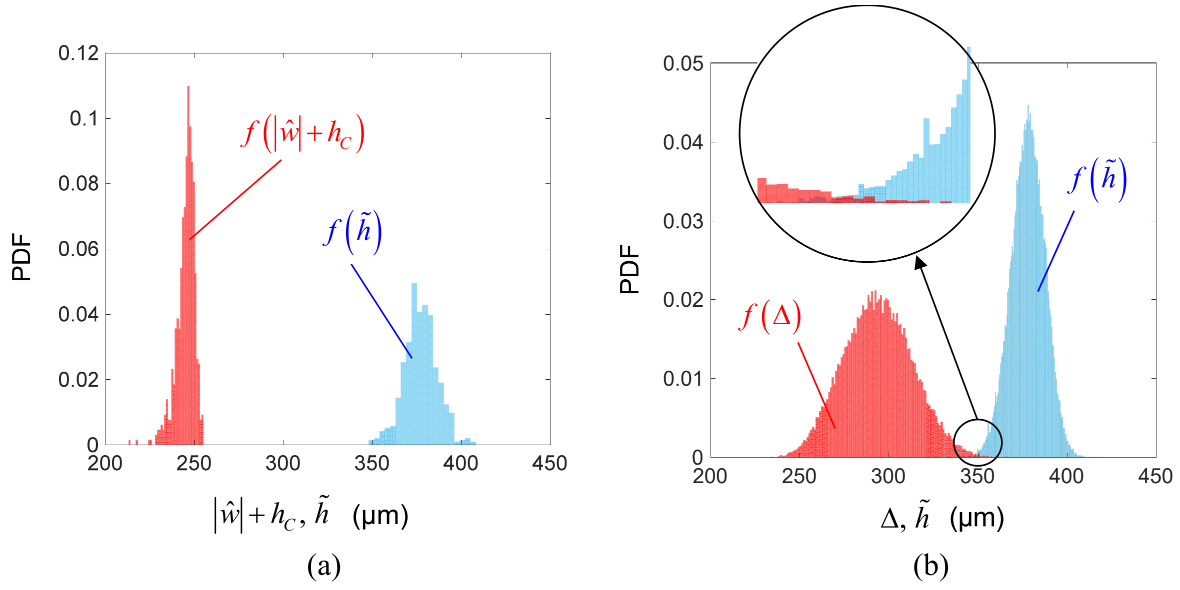



Figure 34. Interference of PDFs of load and strength for the 109th pad ($j$ = 109) for a single MCS run: (a) scenario-1 and (b) scenario-4

The two PDFs do not overlap, and thus, the probability of non-contact open for the Case I is zero, i.e, $\Pr\left(gap_1 > 0\right) = 0$. The other 98,708 samples among the 100,000 samples of the MCS run belong to the scenario-4. They were used to construct the PDF of distance between the top and bottom packages, $f\left(\Delta\right)$, and the PDF of the total solder ball height, $f\left(\tilde{h}\right)$, as shown in (b). In this scenario, the PDFs of load and strength overlap and the enlarged view of the overlapping area is shown in the inset of (b). Then, the probability of non-contact open for the Case II can be obtained by calculating the statistical interference of these two PDFs using numerical convolution methods, such as Monte Carlo convolution method [72] or Fast Fourier Transform (FFT) technique [73].

The Monte Carlo convolution method is used in this study, where a large number of sample points of the load and the strength are generated by MCS from the joint PDFs. The load and strength of each sample are subtracted to get the gap width. For example, in the $k^{\text{th}}$ MCS sample, the load, $\Delta^k$ for the scenario-4 and strength, $\tilde{h}^k$, are subtracted to get the gap width,

$gap_2^k = \Delta^k - \tilde{h}^k$. The result of convolution, i.e., the PDF of the $j^{th}$ gap width, is estimated by using the histogram of gap width from all the MCS samples.

Figure 35(a) shows the entire PDF of the gap at the 109th pad, and Figure 35(b) shows the enlarged view of the tail-end marked by the red box in (a). The probability of the non-contact open is the area of the histogram where *gap* > 0.

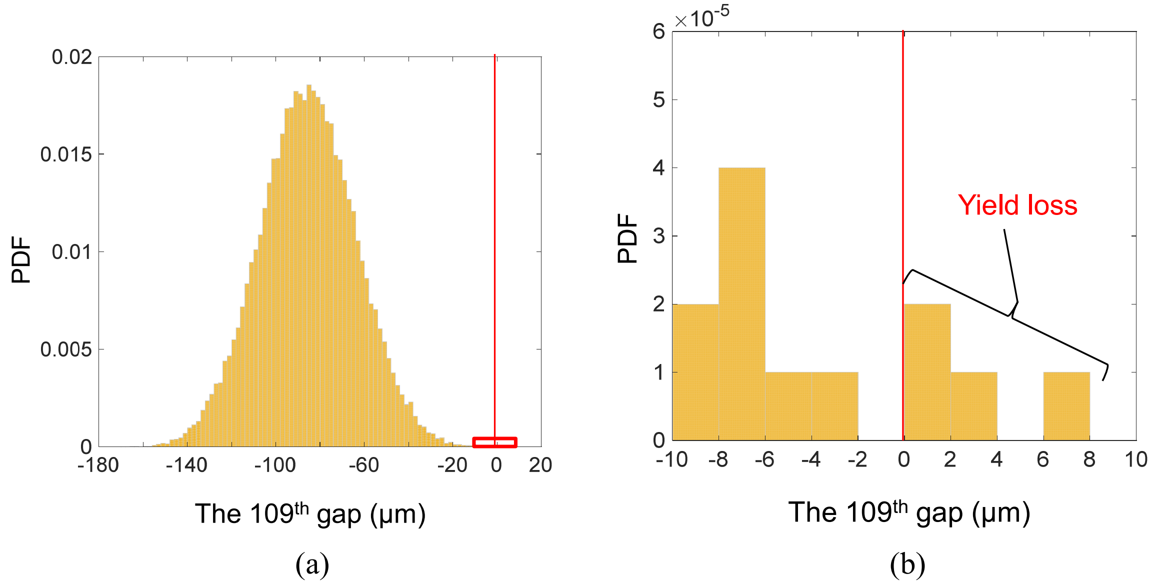



Figure 35. (a) PDF of gap at the 109th pad and (b) the enlarged view of the tail-end marked by the red box in (a)

To complete Step 4, the intersections of the probability of non-contact open for multiple pads (Eq. (22) and (23) of Ref. [64]) have to be calculated in addition to the probability of non-contact open of every single pad. This was also calculated using the table in Step 1.

The $n$-tuple-wise intersections, $\Pr\left(\left(gap^{J_1} > 0\right) \cap \cdots \cap \left(gap^{J_n} > 0\right)\right)$, can be expressed as:

$$\Pr\left(\left(gap^{J_1} > 0\right) \cap \left(gap^{J_2} > 0\right) \cap \cdots \cap \left(gap^{J_k} > 0\right)\right)\Big|_{1 \le J_1 < J_2 < \cdots < J_k \le n} = \frac{N_{\left(gap^{J_1} > 0\right) \cap \left(gap^{J_2} > 0\right) \cap \cdots \cap \left(gap^{J_k} > 0\right)}}{N_{MCS}} \tag{46}$$

where $N_{MCS}$ is the total number of MCS samples; and $N_{(gap^{J_1}>0)\cap(gap^{J_2}>0)\cap\cdots\cap(gap^{J_k}>0)}$ is the number of MCS samples where multiple pads simultaneously have gaps larger than zero; for instances, when $k = 2$, $N_{(gap^{J_1}>0)\cap(gap^{J_2}>0)}$ indicates the number of MCS samples where the gaps are larger than zero on the $J_1$th pad and the $J_2$th pad simultaneously; the $J_1$th pad and the $J_2$th pad are all pairs of pads which can be formed within the total 216 pads.

Once the probability of all intersections were determined, the stacking yield loss was defined as the probability of the union of that gap occurrence for all pads on the stacking interface using Eq. (24) in Ref. [64].

#### 2.3.4.2. True Stacking Yield Loss Estimation using Multiple MCS Runs

It can be seen from Section 2.3.4.1 that regardless of a single pad, an intersection of multiple pads, or the total yield loss for all pads, the estimated probability of non-contact open using the MCS can be expressed as the ratio of the number of samples of failure and the total MCS sample, $k/N_{MCS}$, where $k$ is the number of samples of failure. Due to the very nature of MCS, the estimation result using a single MCS run contains uncertainty.

Based on Ref. [11], when the MCS is done with a sufficiently large sample size, the predicted yield loss follows a binominal distribution, which can be expressed as:

$$f\left(\hat{p} = \frac{k}{N_{MCS}}\right) \sim \binom{N_{MCS}}{k} p^k (1-p)^{N_{MCS}-k}, \; k = 0, \ldots, N_{MCS} \tag{47}$$

where $\hat{p}$ is the predicted yield loss, $k$ is the number of samples having gap > 0, and $p$ is the true yield loss. The mean value of the binominal distribution equals to the true yield loss.

In this study, the MCS with sample size of 100,000 is used, which means the resolution of each MCS run is 10 ppm (i.e., each sample accounts for 10 ppm probability). The probability of non-contact open of each pad and the stacking yield loss were obtained from the average of 30

repetitions of MCS runs with 100,000 samples. The stacking yield loss considering all pads is 464ppm.

The results of the pads with failure probability larger than zero are summarized in Table 19, from which the most critical pad is identified as the 109$^{th}$ pad. Figure 36 shows how the ball pads listed in Table 19 distributed on the stacking interface. The non-contact opens occur on the pads around the package center line in parallel to the longitudinal direction of the silicon dies of the top package (± four pitches in the inner row), where a large warpage difference occurs.

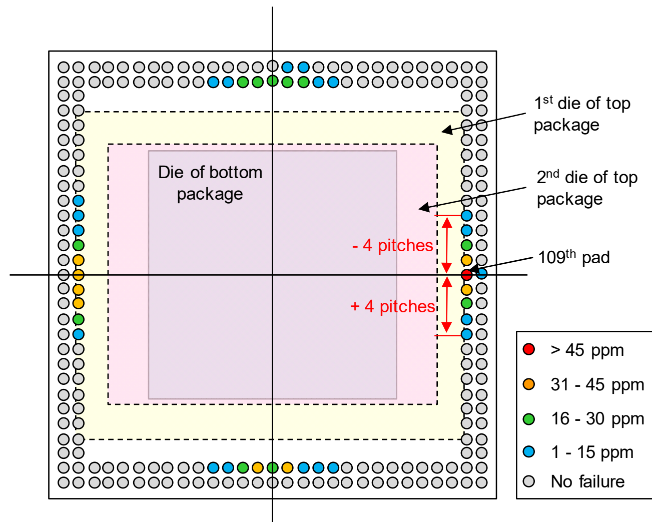



Figure 36. Regions of ball pads having non-contact opens for all of the MCS runs

Table 19 Rank of probability of non-contact open for the $j^{th}$ pad

| $j$ | Pr(gap > 0) (ppm) | $j$ | Pr(gap > 0) (ppm) | $j$ | Pr(gap > 0) (ppm) |
|---|---|---|---|---|---|
| 109 | 59 | 116 | 22 | 121 | 9 |
| 105 | 45 | 171 | 22 | 125 | 8 |
| 113 | 44 | 173 | 20 | 176 | 8 |
| 108 | 37 | 101 | 19 | 40 | 7 |
| 104 | 33 | 174 | 19 | 177 | 7 |
| 43 | 32 | 172 | 18 | 48 | 5 |
| 100 | 32 | 120 | 15 | 169 | 3 |
| 112 | 32 | 97 | 14 | 203 | 3 |
| 45 | 31 | 46 | 12 | 110 | 2 |
| 44 | 26 | 47 | 12 | 124 | 2 |
| 42 | 25 | 170 | 12 | 128 | 2 |
| 117 | 24 | 41 | 11 | 204 | 2 |
| 175 | 24 | 92 | 11 | Others | 0 |
| 96 | 22 | 93 | 10 | | |

### 2.3.5. Conclusion

The stochastic yield loss prediction model was implemented for a PoP assembly with a TFBGA as the top package and an fcBGA as the bottom package. The implementation considered 27 input variables for the uncertainty propagation analyses. Using the $4N + 1$ sample scheme of the EDR method, 137 modeling runs were conducted to produce 549 PDFs. The stacking yield loss (less than 1,000 ppm) was predicted from the PDFs while considering all 216 pads across the stacking interface.

The proposed model offers a more comprehensive PoP stacking yield loss prediction for non-contact open. The uncertainty propagation analysis method used in the study is able to handle the number of input variables much larger than that has been conceived as the practical limit of the UP analysis. More applications are expected to improve the design and manufacturing processes of advanced PoP assemblies, and thus to achieve a desired yield goal.

# CHAPTER 3. ADVANCED STATISTICAL MODEL CALIBRATION TO DETERMINE MANUFACTURING-INDUCED VARIATIONS OF EFFECTIVE ELASTIC PROPERTIES OF SAC SOLDER JOINTS IN LEADLESS CHIP RESISTOR ASSEMBLIES

## 3.1. Introduction

In recent years, Sn-rich solders, such as SnAgCu (SAC) alloys, have been used widely in electronics packaging to comply with the regulations. It has been known that the Sn-rich solders have large variations in grain sizes and orientations, intermetallic compound (IMC) sizes, and distributions of anisotropic Sn crystals [5]. These microstructural variations occur even under the same assembly condition, which results in inherent mechanical property variations of solder joints; not only package-to-package variations but also joint-to-joint variations in the same package [6, 74-79]. The mechanical properties of Sn-rich solder materials are attributed significantly to the statistical variation of a solder joint life.

The solder joint life is assessed typically by (1) testing a "statistically significant" number of components under accelerated loading conditions, and (2) determining the characteristic number of cycles-to-failure (CTF) from the test results, which have a statistical distribution due to the inherent manufacturing variability. Figure 37 illustrates the statistical distribution of CTF data and the characteristic life. The characteristic life is often referred to as "durability", which represents the length of time over which a product will provide its intended function under operating conditions, i.e., the reliability of nominal parts. Durability is important when competitive designs are compared for design-for-reliability. On the other hand, the distribution of test results can be used to calculate the "reliability", which is the probability that a product will perform its intended function under operating conditions, for a specific period of time. The reliability becomes critical when early failure probability has to be evaluated, e.g., the probability that the product will fail before a warranty period ends.

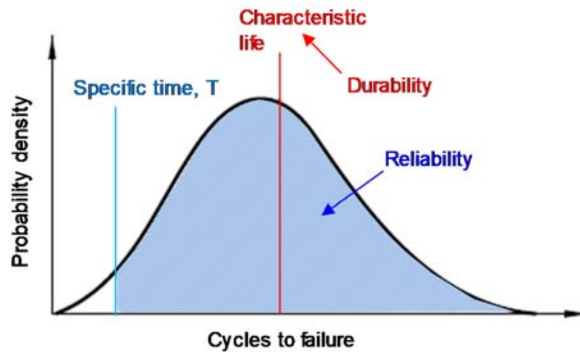



Figure 37. Durability vs. reliability

Numerous physics-based lifetime models have been developed for durability assessment. More recently, the models have been extended into a probabilistic domain to predict reliability while taking into account inherent manufacturing variability [7, 80-82] . The inherent variations can be categorized into "known input variables" and "unknown input variables". The "known input variables" are the variables whose statistical distributions are known or can be measured. The variables are called "unknown" when their statistical distributions are extremely difficult or impractical to be obtained experimentally. These unknown distributions can be obtained by "statistical model calibration", which identifies the statistical distribution of the unknown input variables that produces the best agreement between the experimental data and the predicted distribution through an optimization process.

In this section, the unknown statistical distributions of two effective elastic properties of Sn-3.0Ag-0.5Cu (SAC305) solder joint of leadless chip resistors (LCRs) assemblies are determined by the advanced statistical model calibration while considering the statistical variations of serval other known input variables including die thickness, solder joint height, termination length, and thickness and elastic moduli of a printed circuit board (PCB).

The background of the statistical model calibration is described first. The cyclic bending test results of the LCR assemblies are followed, and the results are subsequently used to obtain the

statistical distributions of the effective elastic properties of SAC305 solder. Validity of the calibrated model is corroborated by comparing the predicted probability density function (PDF) of cycles-to-failure of the identical LCR assemblies subjected to a different loading level with the cycles to failure distribution of actual testing data.

## 3.2. Background: Statistical Model Calibration

Statistical model calibration is an optimization process to inversely identify the statistical distributions of unknown input variables that produce the best agreement between experimental data and predicted distribution [83]. It utilizes a full set of experimental data containing multiple samples so that the calibration results can be obtained after considering the statistical variations of performance responses induced by the intrinsic variabilities of input variables. Figure 38 shows the calibration procedure. The uncertainties in the known and unknown input variables are propagated through the model to acquire the predicted PDF of performance responses. Then, the predicted PDF of performance response is compared with the experimental data. The degree of agreement is determined by a “calibration metric,” which acts as the objective function of the optimization process. The unknown input variables are iteratively updated until the calibration metric is minimized or maximized.

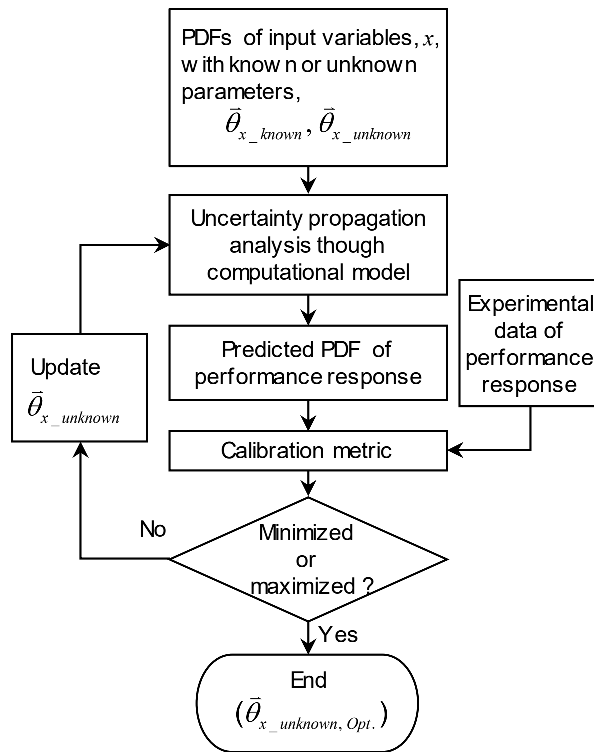



Figure 38. Procedure of statistical model calibration

The most critical element of the calibration procedure is the uncertainty propagation (UP) analysis. The models of semiconductor packages are usually computationally intensive. The EDR method is also implemented for the UP analysis in this study.

The other key element of the calibration procedure is a calibration metric, which has to account for the statistical comparison between the predicted PDF and the experimental data. The likelihood function is the most widely used calibration metric. Figure 39 illustrates the cases with the maximum likelihood value and a low likelihood value. In practice, the logarithm operation is applied to the likelihood function to convert the multiplication form of the likelihood function to a summation form. In addition, most of the commercial optimization algorithms minimize a function, and thus, a negative log-likelihood (NLL) is used; minimizing the NLL is equivalent to maximizing the log-likelihood.

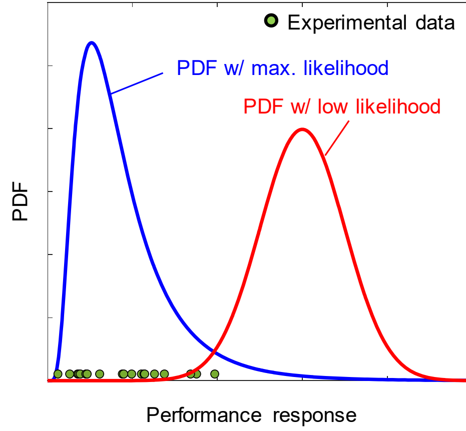



Figure 39. Illustration of likelihood function with the maximum and minimum likelihood values

The NLL function can be expressed as [84]:

$$L\left(y_1, \ldots, y_n \mid \vec{\theta}_{x_known}, \vec{\theta}_{x_unknown}\right) = -\sum_{i=1}^{n} \log\left[f_{\mathbf{pre}}\left(y_i \mid \vec{\theta}_{x_known}, \vec{\theta}_{x_unknown}\right)\right] \tag{48}$$

where $y_1, \ldots, y_n$ are the $n$ number of experimental data, $\vec{\theta}_{x_known}$ and $\vec{\theta}_{x_unknown}$ are the vectors of the parameters of PDFs of the known and the unknown input variables, respectively, and $f_{\mathbf{pre}}$ is the predicted PDF of performance response, $y$. The $\vec{\theta}_{x_unknown}$ is calibrated by minimizing the value of the NLL.

## 3.3. Implementation Using Vibration-Induced Solder Fatigue Failure

The experimental setup of a cyclic bend test is described in Section 3.3.1. The physics-based lifetime prediction is followed in Section 3.3.2. In Section 3.3.3, the statistical model calibration for the effective elastic properties of SAC305 solder is provided.

### 3.3.1. Setup of Four-Point Bend Test

Figure 40 shows the test coupon, which was fabricated by mounting multiple 6332-type LCRs on a PCB with SAC305 solder. The pad length was 1 mm.

A four-point bend test setup used in the study is shown in Figure 41. A stain gage was attached to the PCB, and it was aligned with the longitudinal side of the PCB to determine the

board level strain. Two sets of LCR assemblies were tested at a frequency of 2.5 Hz for the maximum board strains of 2000 με and 1200 με, respectively.

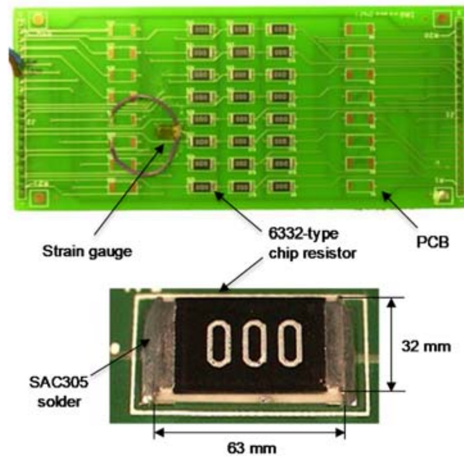



Figure 40. Detail of a test coupon and a single leadless chip resistor assembly

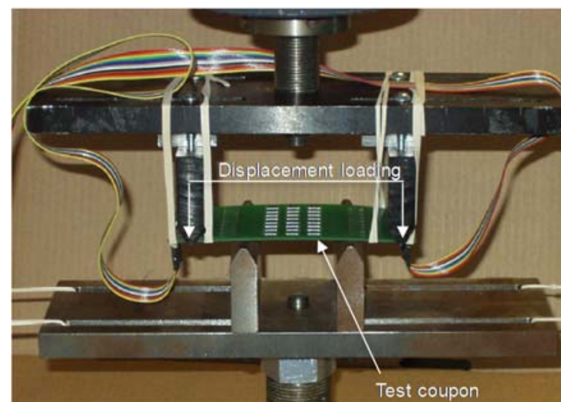
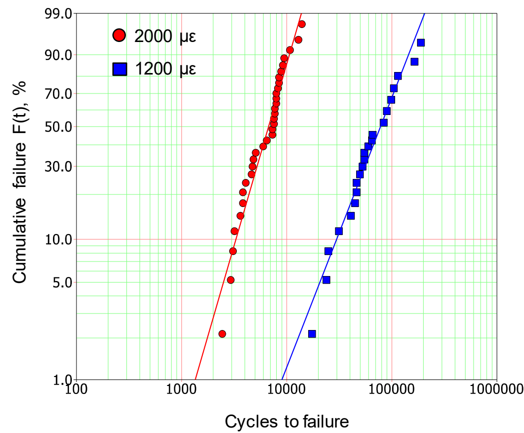
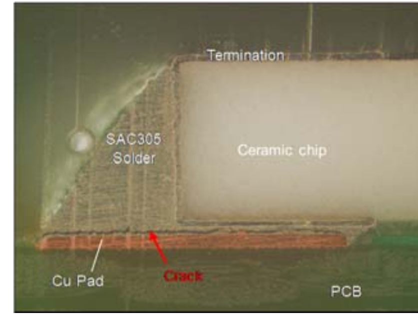



Figure 41. Test coupon mounted on 4 point bend fixture

Figure 42(a) shows the 2-parameter Weibull plots of the test results. The characteristic cycles-to-failure (CTFs) and the standard deviations of test results were (7,723, 2,847) and (78,618, 43,827) for the 2000 με stain level and the 1200 με stain level, respectively. Figure 42(b) shows the observed failure mode, i.e., the bulk solder failure along the copper pad.




(a)




(b)

Figure 42. Accelerated life testing results of LCR assemblies subjected to strain levels of 2000 and 1200 με. (a) Weibull plots and (b) Cross section of failed assembly

### 3.3.2. Physics-based Solder Joint Fatigue Life Modeling

The Basquin power law relation is widely used for the lifetime prediction of solder joints under mechanical vibrations, which can be expressed as [85]:

$$\sigma_a = \sigma_f \left(2N_f\right)^b \tag{49}$$

where $\sigma_a$ is the stress amplitude, $\sigma_f$ is the fatigue strength coefficient, $b$ is the fatigue strength exponent, and $N_f$ is the cycles to failure. The constants from Ref. [85] were used in this study: $\sigma_f$ = 64.8 MPa and $b = -0.1443$.

Figure 43(a) shows a finite element model constructed by ANSYS® using quarter symmetry. The condition of periodicity was imposed on the boundary to simulate the effect of multiple components. Table 20 lists the material properties used in the analysis. Since only stress magnitudes were required for Eq. (49), linear elastic material properties were used for the packaging materials.

The displacement loading was iteratively adjusted to find the accurate loading condition, which made the x-direction strain in the strain gage area reached 2000 με.

Table 20  Material Properties

| Material | Young's modulus (GPa) | Poisson's Ratio |
|---|---|---|
| Alumina chip | 295 | 0.22 |
| Termination (Ag) | 76 | 0.37 |
| Cu Pad | 132.4 | 0.34 |
| PCB | In-plane: 19.7<br>Out-of-plane: 9.1 | In-plane: 0.39<br>Out-of-plane: 0.17 |
| SAC305 | 51 | 0.36 |

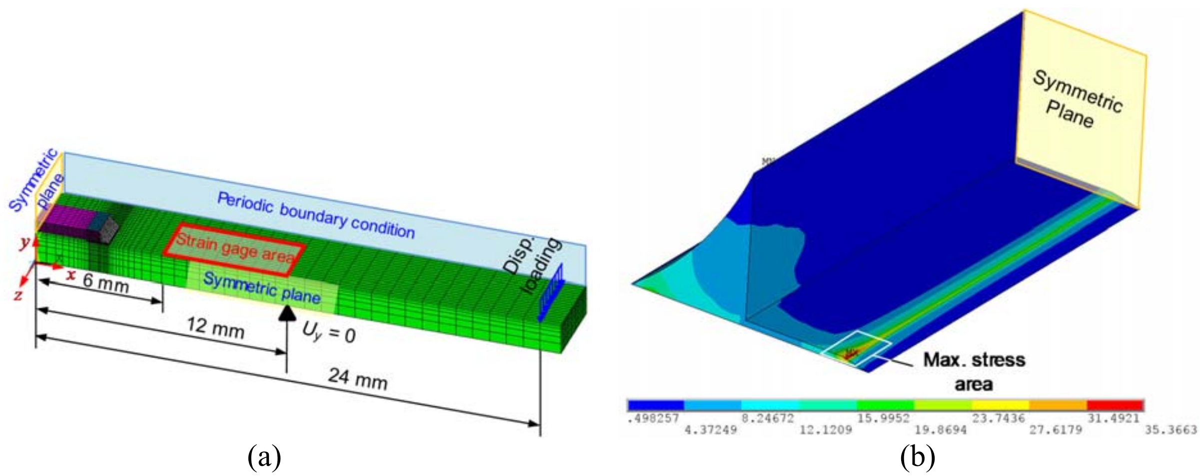



Figure 43.  Damage analysis (a) finite element model (b) von Mises stress distribution in the solder joint

Figure 43(b) shows the von Mises stress distribution on the solder joint.  The volume averaged von Mises stress over the critical region on the top layer of the solder joint was selected as the damage index, which can be expressed as [85]:

$$\sigma_a = \frac{\sum \sigma \cdot V}{\sum V} \tag{50}$$

### 3.3.3. Statistical Model Calibration

The known and unknown input variables are described first. Calibration results and discussion are followed.

#### 3.3.3.1. Known and Unknown Input Variables

The statistical calibration begins with the selection of known and unknown variables in the solder joint fatigue life model. Six known input variables were identified from the literature review [62, 81, 86, 87]: ceramic die thickness, solder joint height, termination length, as well as thickness and two effective moduli of PCB. Their variations are listed in Table 21. Among these variables, the effective moduli of PCB and PCB thickness have statistical correlations, because their variations come from the thickness variation of each layer in the PCB stack-up. Their correlation coefficients were estimated from the literature [62] and they are listed in Table 22.

Table 21 Input Variables

| Variables | Physical meaning | Mean | Std. Dev. | Distribution |
|---|---|---|---|---|
| $x_1$ | Alumina chip thickness (mm) | 0.55 | 0.033 | Normal |
| $x_2$ | Solder joint height (mm) | 0.0237 | 0.0068 | Normal |
| $x_3$ | Termination length (mm) | 0.405 | 0.053 | Normal |
| $x_4$ | PCB thickness (mm) | 1.59 | 0.036 | Trivariate Normal |
| $x_5$ | PCB in-plane Young's modulus (GPa) | 19.7 | 0.591 | |
| $x_6$ | PCB out-of-plane Young's modulus (GPa) | 9.1 | 0.273 | |
| $x_7$ | Solder joint modulus (GPa) | unknown | | Lognormal |
| $x_8$ | Solder joint Poisson's ratio | | | Lognormal |

Table 22 Correlation coefficient of correlated input variables

| | $x_4$ | $x_5$ | $x_6$ |
|---|---|---|---|
| $x_4$ | 1 | 0.84 | 0.38 |
| $x_5$ | 0.84 | 1 | 1 |
| $x_6$ | 0.38 | 1 | 1 |

Two unknown variables are the effective Young's modulus and Poisson's ratio of the SAC305 solder joint. An unknown number of tin grains and their orientations within the solder joint were attributed to their statistical variations. Both unknown variables were assumed to follow a lognormal distribution, since a negative value was not permitted.

Two statistical parameters are required to describe a lognormal distribution. The representative values of the literature [85] were used for the initial guess of the mean and standard deviation of two unknown input variables: (51 GPa, 2.55 GPa) for the Young's modulus and (0.36, 0.018) for the Poisson's ratio.

#### 3.3.3.2. Calibration Results and Discussion

Since the calibration metric was an implicit objective function, the finite difference method was used to determine the convergence direction by calculating the gradients of the objective function with respect to the mean ($\mu$) and standard deviation ($\sigma$) of the unknown variables.

Figure 44 shows the value of a negative log-likelihood (NLL) as a function of the iteration number. The calibration converged after 5 iterations. The iteration was terminated when the change of the input variables (the Young's modulus and the Poisson's ratio) to the next iteration was smaller than 0.1 GPa and 0.01, respectively.

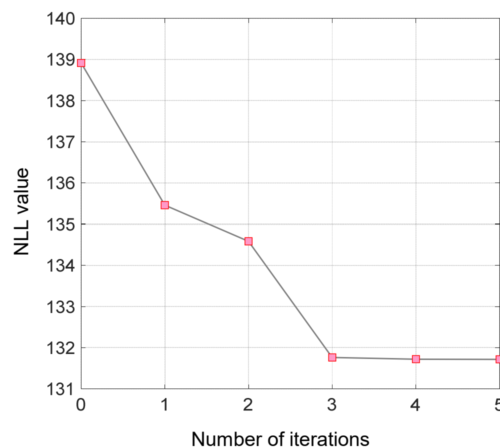
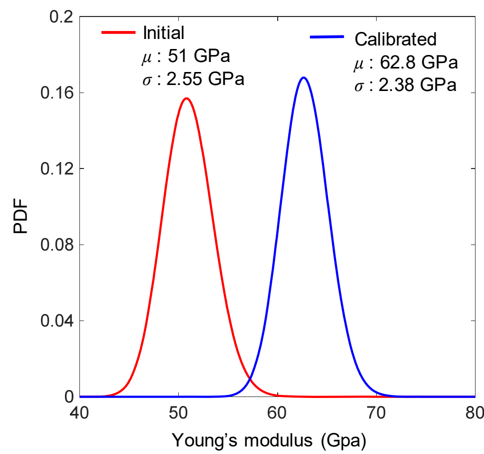
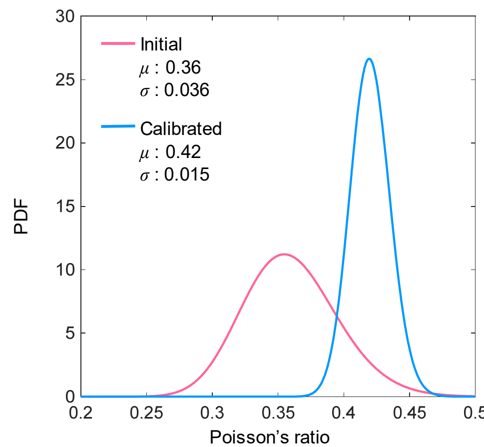



Figure 44. Negative log-likelihood (NLL) values as a function of iteration number

Figure 45 compares the initial PDFs of two unknown variables with the calibrated PDFs. The mean value of Young's modulus increases from 51 GPa to 62.8 GPa, and the standard deviation slightly decreases from 2.55 GPa to 2.38 GPa. Based on the calibrated PDF of Young's modulus, 99% of the Young's modulus values are distributed within the range of [56.9 GPa, 69.2 GPa], which is close to the variation reported in Ref. [88]. In (b), the Poisson's ratio also shows an increasing mean (0.36 to 0.42) and deceasing standard deviation (0.036 to 0.015). Accordingly, there are 99% of the Poisson's ratio values are located within [0.38, 0.46].




(a)




(b)

Figure 45. Initial guesses and calibrated values of (a) Young's modulus and (b) Poisson's ratio of SAC305 solder

Figure 46 compares the experimentally obtained CTF histogram with the predicted PDFs of CTF using the initial guesses of input variables as well as the calibrated input variables. The result clearly shows an improved agreement after calibration.

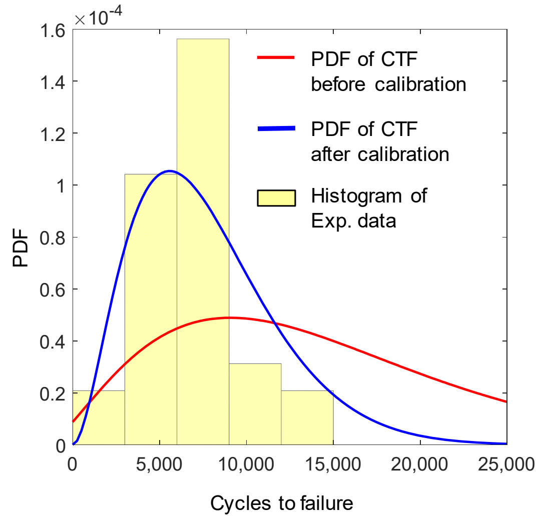



Figure 46. PDFs of cycles-to-failure before and after calibration are compared with the histogram of experimental data obtained at 2000 με strain level

To evaluate the validity of the calibration results, the CTF distribution of the identical LCR assemblies subjected to a strain level of 1200 με was predicted by the calibrated model. The results are compared with the actual test data in Figure 47. This confirms that the calibrated model can accurately predict the CTF distribution of the LCR assemblies subjected to other loading conditions, provided that the assemblies are fabricated by the same manufacturing condition.

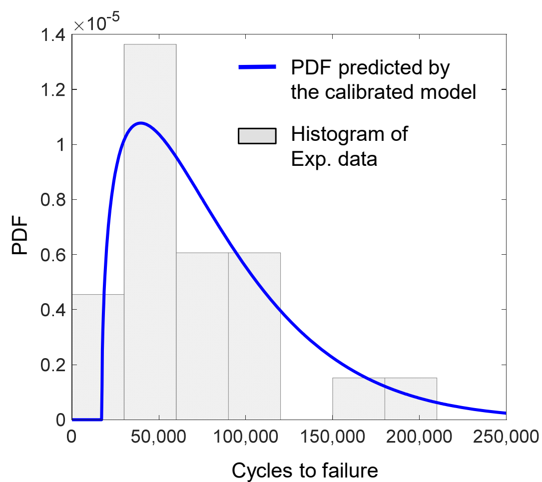



Figure 47. PDF of cycles-to-failure predicted by the calibrated model is compared with the histogram of experimental data obtained at 1200 με strain level

### 3.3.4. Discussions

#### 3.3.4.1. Modeling Runs

The number of modeling runs required to complete the iterations is described below:

(1) In each iteration, the gradients with respect to $\mu$ and $\sigma$ of each unknown variable had to be determined. For the gradient of each $\mu$, a total of $4N$+1 modeling runs (33 = 4 x 8 + 1) had to be performed to construct the PDF. For the gradient of each $\sigma$, however, only 4 additional modeling runs related to each unknown variable had to be performed because the center did not change. It is worth recalling that, with the EDR method, all of one dimensional integrations pass through the same cut center, which is a vector of the mean values of all input variables. The eigenvector sampling selects this point as the center, and as long as the mean does not change, only four additional points of $\mu \pm 1.5\sigma$ and $\mu \pm 3\sigma$ are needed to calculate the PDF. Thus, a total of **74** modeling runs (= 33 x 2 + 4 x 2) were performed for each iteration.

(2) After the gradients were determined, the values of $\mu$ and $\sigma$ of two unknown variables were updated, and an additional $4N + 1$ modeling runs were performed subsequently to calculate the calibration metric with the updated unknown input variables: **33** modeling runs.

(3) Considering the modeling runs (33) that were performed to evaluate the calibration metric with the initial guesses of two unknown variables, a total of **568** modeling runs (107 x 5 + 33) were required to complete the calibration.

A linear elastic analysis was used in this study, and the required modeling runs were handled readily by a high-performance workstation. If a more complex assembly that requires a non-linear analysis is considered, the required computational burden can exceed a practical limit in spite of the effectiveness provided by the EDR method. A more efficient calibration metric

and/or a more accurate metamodeling scheme that does not suffer from the curse of dimensionality are warranted for those applications.

#### 3.3.4.2. Reliability Prediction under Usage Condition

Reliability of the LCR assemblies under a usage condition can be predicted by the calibrated model. A typical key-press condition of the mobile phone was considered to illustrate the extension. The maximum board strain of about 400 με stain level was reported in JESD22B113 [89], and it was used in the calibrated model.

Figure 48 shows the left tail-end of the PDF of the solder joint fatigue life of the LCR assemblies under the above usage condition. The full PDF is shown in the inset. As an example, for a typical requirement of one million key-presses, only 0.001 ppm LCRs will fail before the requirement (i.e., the red area). If the requirement increases to ten million key-presses, the probability of early failure becomes 0.0092 ppm (i.e., the yellow area).

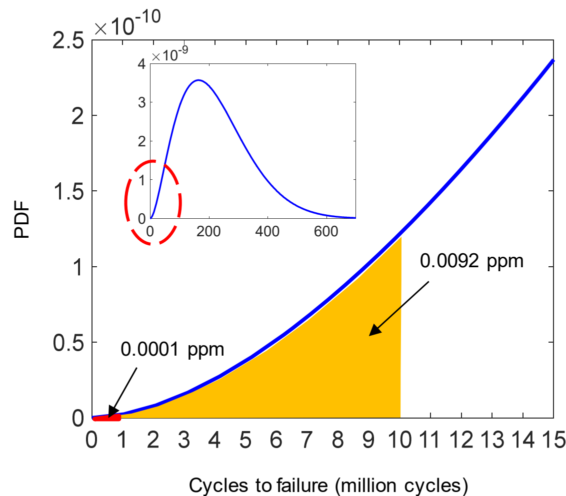



Figure 48. PDF of cycles-to-failure under a portable device usage condition, predicted by the calibrated model

The results indicate that the LCR assemblies will not likely to fail by pressing keys, but the calibration results obtained from the proposed method can be used effectively to assess drop reliability, which is more critical requirements for portable electronics.

### 3.3.5. Conclusions

The unknown statistical distributions of two effective elastic properties of SAC305 solder were determined by the advanced statistical model calibration using the cyclic bending test results of LCR assemblies. The calibration procedure employed the advanced approximate integration method for an UP analysis to be able to take into account the statistical variations of six additional known input variables. The cycle-to-failure distribution of the identical LCR assemblies subjected to a different loading level was predicted accurately by the calibrated model, which corroborated the validity of the proposed approach.

It is important to note that the conventional UP analysis would require a lot more modeling runs, which would make the statistical calibration impractical; for example, the MCS would need thousands to tens of thousands of modeling runs for the PDF construction of each gradient calculation, and the RSM would need $5^8$ modeling runs for the response surface construction with the same level of sampling resolution of the $4N+1$ scheme used in this study.

The calibrated model was further utilized to predict the reliability of the LCR assemblies under a usage condition - the key-press condition of mobile phones. It is anticipated that the proposed methodology can be used effectively to predict a field failure rate and/or a warranty period of electronics products using Sn-rich solders.

# CHAPTER 4. AN ADVANCED HDMR-BASED METAMODELING TECHNIQUE FOR SEMICONDUCTOR PACKAGES WITH A LARGE NUMBER OF INPUT VARIABLES

## 4.1. Introduction

Numerical modeling has been used widely for product development of semiconductor packages to understand mechanical behavior and to provide design guidelines [90-93]. When a large number of input variables has to be considered, the metamodel (or metamodeling) techniques, which are also known as response surface method (RSM), surrogate model, or reduced-order model, are used to cope with the challenges associated with the excessive computational burden [94-97].

The metamodeling techniques utilize simple analytical models to approximate the input/output relationship of numerical models, which involve two major tasks: (1) discrete sample points generation and (2) model choice to connect the discrete sample points [98]. These two tasks are combined to form the approximated response surface of a computationally expensive model. The studies on metamodeling have been conducted rigorously for a few decades [99-103]. It is clear from the literature that the metamodels have been implemented successfully for relatively low dimensional problems (i.e., the problems with a few number of input variables).

As the number of input variables increases, the required modeling runs of the classical metamodeling methods grow rapidly. For example, the full factorial design (FFD), as the most basic discrete sample points generation method, selects $s^N$ sample points, where s is the number of sample points along each input variable and $N$ is the total number of input variables. The central composite design (CCD) significantly alleviates the computational burden of FFD by taking 2 sample points along each variable and the center points of each plane of the two-variable pair, but CCD still requests the number sample points proportional to $2^N$. This is known as "curse-of-

dimensionality", which makes the computational cost of the metamodel construction unaffordable when the number of input variables is large.

More recently, another efficient metamodeling technique, termed "High Dimensional Model Representation (HDMR)" [104, 105], has been introduced to engineering design community. The HDMR decomposes an integrable multivariate function into multiple lower-order component functions based on the hierarchical structure of interaction effects of the input variables. For most well-defined physical systems, the high-order interactions are negligible [105, 106], and thus, the performance responses can be approximated by the sum of low-order component functions. A family of HDMRs have been developed to serve for different purposes using various component functions [74, 104, 107-109]. Among the HDMR family, the Cut-HDMR based methods are most widely used because of the simplicity of the component functions (i.e., the component functions do not require integrations). If the Cut-HDMR based methods use the component functions up to $L^{\text{th}}$ order, the computational cost reduces from the exponential scaling of $\sim s^N$ of the classical metamodels to a polynomially increasing of $\sim (Ns)^L / L!$ [101].

For semiconductor packaging products, it is often to have a large number of input variables and non-negligible second-order interaction effects. Even employing Cut-HDMR based methods, the required number of model runs is proportional to $(Ns)^2/2$, which is still impractical for many semiconductor packaging applications.

The other limitation of implementing the current metamodeling techniques for semiconductor package products is the requirement that input variables be independent from each other. In general, the input variables of semiconductor packages are statistically linearly dependent, which is also known as statistically correlated. Figure 49 illustrates the effect of the assumption of independency on the cases where the input variables are correlated. In the figure, the green

diamonds represent existing EMC materials obtained from data sheets [110]. A strong negative correlation exists because the higher filler percentage of low-CTE silica particles results in the lower CTE and the higher elastic modulus. On the other hand, the red dots illustrate the sample points without considering the correlation. It is clear that the sample points at the upper-right and lower-left corners contain unrealistic CTE and Young's modulus combinations. These sample points are wasted in the metamodel construction; in some cases, they can twist the metamodel. To the author's best knowledge, there are no existing metamodeling techniques that can handle the statistical correlations of the input variables.

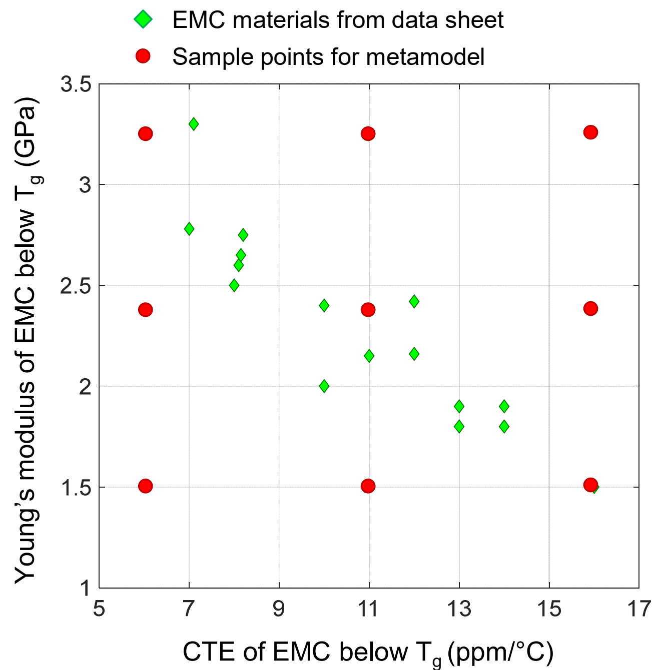



Figure 49. Illustration of the independence requirement of input variables on the sampling strategy

The objective of this paper is to propose an advanced Cut-HDMR based metamodeling technique, called partitioned bivariate Cut-HDMR (PB Cut-HDMR), to cope with the above-mentioned challenges for semiconductor package products.

Section 2 provides background of the current Cut-HDMR based methods. Section 3 describes the proposed PB Cut-HDMR. In Section 4, the proposed method is implemented to construct the metamodel for the warpage prediction of a thin flat ball grid array (TFBGA) assembly. Finally, conclusions are drawn in Section 5.

## 4.2. Background: Cut-HDMR Based Methods

The HDMR expansion is performed based on the interaction effects of the input variables. The term "interaction" employed here means the effect of variables on the performance response when more than one variables act together. This is distinctly different from the term "correlation" employed in statistics, which represents whether and how strongly pairs of random variables are related. A general form of HDMR is defined as [104]:

$$\begin{aligned} y(x_1, x_2, \ldots, x_N) \equiv y_0 + \sum_{i=1}^{N} y_i(x_i) + \sum_{1 \le i < j \le N} y_{ij}(x_i, x_j) + \cdots \\ + \sum_{1 \le i_1 < \cdots < i_l \le N} y_{i_1 i_2 \ldots i_l}(x_{i_1}, x_{i_2}, \ldots, x_{i_l}) + \cdots + y_{12\ldots N}(x_1, x_2, \ldots, x_N) \end{aligned} \quad (51)$$

where $y(x)$ is the performance response function; $y_0$ is a constant representing the mean of the performance response, which is called "zeroth-order effect"; $y_i(x_i)$ represents the effect when the variable $x_i$ acts independently on $y(x)$, which is called "first-order effect" or "main effect"; $y_{ij}(x_i, x_j)$ is the effect on $y(x)$ when the variables $x_i$ and $x_j$ act together, which is called "second-order effect" or "bivariate interaction effect". It should be noted that $y_{ij}(x_i, x_j)$ excludes the main effects of $x_i$ and $x_j$ as well as the mean effect. The subsequent terms indicate the higher order interaction effects of more variables acting together on $y(x)$. The last term $y_{12\ldots N}(x_1, x_2, \ldots, x_N)$ represents the residual influence.

The Cut-HDMR expresses $y(x)$ as a superposition of value of $y(x)$ on a point called "cut" center as well as lines, planes and hyper-planes (called cuts) passing through the "cut" center. For a chosen "cut" center, $\vec{x}_0$, the component functions of the Cut-HDMR are defined as [104]:

$$\begin{cases} y_0 = y(\vec{x}_0) \\ y_i(x_i) = y(x_i, \vec{x}_0^i) - y_0 \\ y_{ij}(x_i, x_j) = y(x_i, x_j, \vec{x}_0^{ij}) - y_i(x_i) - y_j(x_j) - y_0 \\ \vdots \\ y_{12\ldots N}(x_1, x_2, \ldots, x_N) = y(x_1, x_2, \ldots, x_N) - y_0 - \sum_i y_i(x_i) \\ \qquad - \sum_{ij} y_{ij}(x_i, x_j) - \ldots - \sum_{i\ldots,N-1} y_{i,\ldots,N-1}(x_i, \ldots, x_{N-1}) \end{cases} \tag{52}$$

where $\vec{x}_0 = \left[x_{1_0}, x_{2_0}, \ldots, x_{N_0}\right]$; $\vec{x}_0^i$ is $\vec{x}_0$ without the element $x_i$; $\vec{x}_0^{ij}$ is $\vec{x}_0$ without the elements, $(x_i, x_j)$, and so on; and $y(x_i, \vec{x}_0^i)$ is a 1-D performance response function along the $x_i$ direction that passes through $\vec{x}_0$; $y(x_i, x_j, \vec{x}_0^{ij})$ is a 2-D performance response function of the $(x_i, x_j)$ plane that passes through $\vec{x}_0$, and so on. The complicated *N*-dimensional performance response function is decomposed and expressed as a superposition of low dimensional performance response functions. The Cut-HDMR uses only arithmetic computation to determine the component functions, and thus, it has the least amount of computational cost comparing to other HDMRs [74, 104].

It has been shown that the mean values of the input variables are the optimal "cut" center when the terms only up to the second-order are considered [111]. Accordingly, the metamodel based on the bivariate Cut-HDMR can be obtained by substituting Eq. (52) into Eq. (51) with the "cut" center being the mean values of input variables, which is written as [112]:

$$y(x_1, x_2, \ldots, x_N) \cong \sum_{1 \le i < j \le N} y(x_i, x_j, \vec{\mu}^{ij}) - (N-2) \sum_{i=1}^{N} y(x_i, \vec{\mu}^i) + \frac{N(N-1)}{2} y_0 \tag{53}$$

where $\vec{\mu} = [\mu_1, \mu_2, \ldots, \mu_N]$ is the vector of mean values of $N$ input variables; $\vec{\mu}^i$ is $\vec{\mu}$ without the element $\mu_i$; $\vec{\mu}^{ij}$ is $\vec{\mu}$ without the elements, $(\mu_i, \mu_j)$; $y_i(x_i, \vec{\mu}^i)$ is the 1-D function indicating the underlying function along the $x_i$ direction that passes through $\vec{\mu}$; and $y_{ij}(x_i, x_j, \vec{\mu}^{ij})$ is a 2-D functions representing the underlying function on the $(x_i, x_j)$ plane that passes through $\vec{\mu}$.

Figure 50 illustrates the concept of the Cut-HDMR with a 2-D function, where the underlying function is decomposed into four component functions. Figure 50(a) shows the underlying 2-D function as the black meshed surface and the dot represents the zeroth-order effect, which is a constant. In (b), the blue curve is the performance response along $x_1$ direction while keeping $x_2$ as $\mu_2$, and the green line is the zeroth-order effect along $x_1$ direction. The main effect of $x_1$ is the red curve, which is obtained by subtracting the green line from the blue curve. The same procedure can be applied to obtain the main effect of $x_2$, as shown in (c). In (d), the blue surface is obtained by the superposition of the red curves in (b) and (c), which represents the performance response without interaction effects. The green plane is the zeroth-order effect. By subtracting the blue surface and the green plane from the black surface, the interaction effect of the $(x_1, x_2)$ pair is obtained, and it is shown as the red surface.

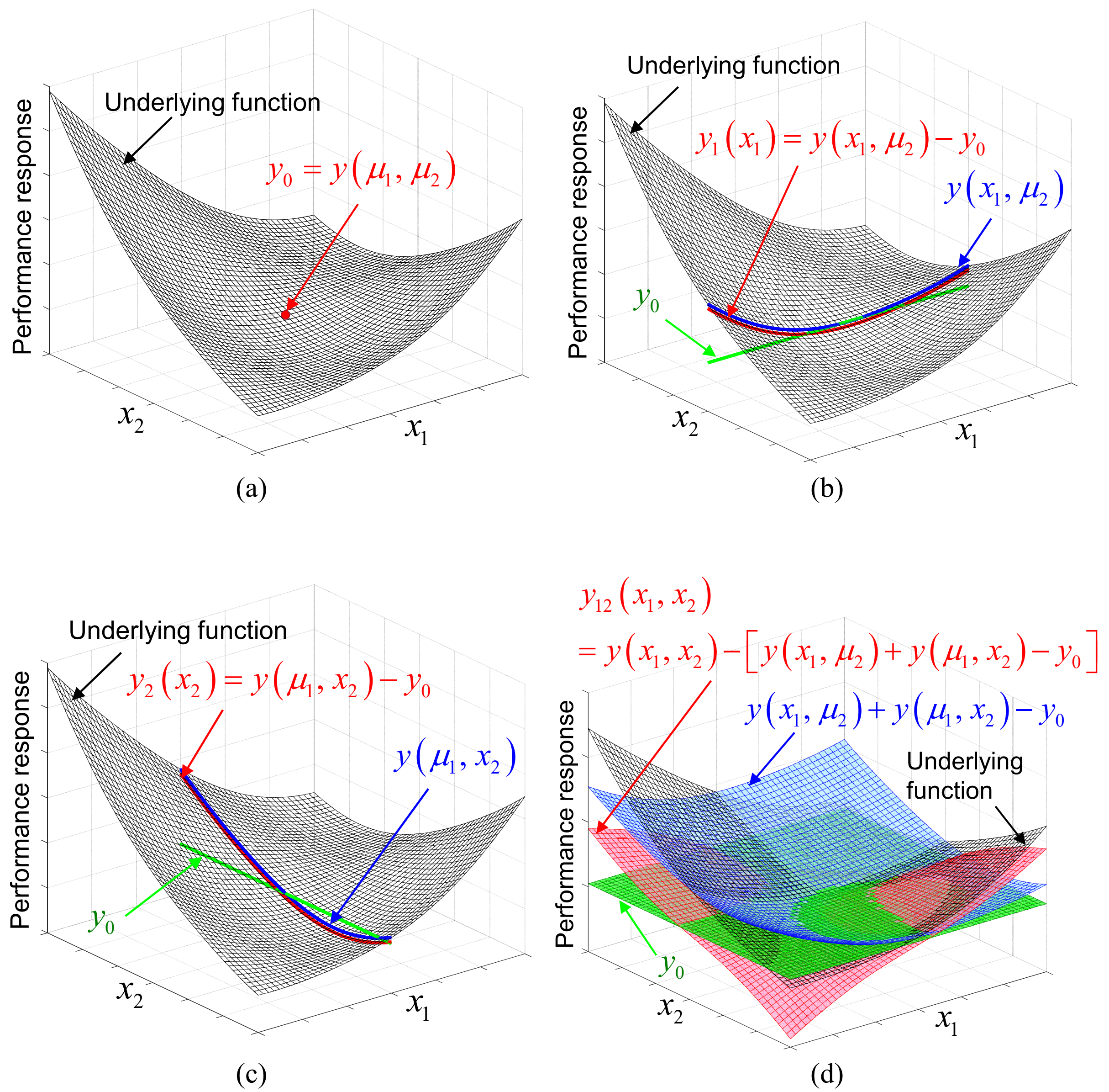



Figure 50. Illustration of Cut-HDMR using a 2-D function (a) underlying 2-D function and zeroth-order effect, (b) main effect of $x_1$, (c) main effect of $x_2$, and (d) interaction effect.

Several Cut-HDMR based methods, including the Cut-HDMR with *Lagrange interpolation polynomial* (LIP) [112], *moving least squares* (MLS) [74], and *radial basis function*-HDMR (RBF-HDMR) [109], have been developed to construct multiple 1-D and 2-D metamodels to estimate the 1-D functions and 2-D functions in Eq. (53) and combine them to form the *N*-dimensional metamodel.

In the original form, the Cut-HDMR [105] employees a look-up table but did not provide the guidelines of how to build it. To provide a more systematic sampling approach, the LIP and MLS were adopted to approximate the performance response using uniformly distributed sample points. For univariate terms, say a total of $s$ sample points are distributed along each input variable, the reference point is the center and the remaining ($s - 1$) samples are evenly located at both side with respect to the reference point. For the bivariate terms, the sample points form a uniform gird on a plane with the reference point as the center.

The number of modeling runs to construct a bivariate Cut-HDMR metamodel using the LIP or MLS, can be generally expressed as:

$$c = \frac{N(N-1)}{2}(s-1)^2 + N(s-1) + 1 \tag{54}$$

where $s$ is the number of sample points taken along the direction of each input variable.

Based on Eq. (54), it is clear that the majority of modeling runs comes from the bivariate terms. The effect sparsity principle of the factorial experiments states that the number of relatively significant bivariate terms is small [113, 114]. This principle is corroborated by numerous parametric studies for semiconductor packages in the literature [115-117].

If the bivariate terms in HDMR can be partitioned into significant terms and minor terms, the computational cost of construction of 2-D metamodels for the bivariate terms can be reduced by spending the expensive modeling runs on the significant terms only.

The RBF-HDMR is developed to express the Cut-HDMR with an explicit model form for high dimensional design optimization problems and propose an accompanying sampling strategy further reduces the required modeling runs of LIP or MLS bivariate Cut-HDMR by using less sample point for the linear univariate and bivariate terms as well as employing the concept of spending modeling runs on significant bivariate terms only. The RBF-HDMR, however,

determines the interaction effects and nonlinearity based on randomly selected sample points, which can possibly fail to identify some significant bivariate terms. In addition, there is lack of study on the effect of different RBFs on the prediction accuracy, and the metamodel is accurate only when the selected RBFs can represent the underlying functions.

### 4.3. Proposed Partitioned Bivariate Cut-HDMR (PB Cut-HDMR)

The proposed PB Cut-HDMR is presented. First, the challenge associated with correlated input variables is handled by transforming the correlated input variables into a new set of uncorrelated input variables. Then, the effects of univariate terms and bivariate terms of the transformed input variables are quantitatively ranked of partitioned for the reduction of computational cost. A detailed procedure of incorporating the two critical components to form a metamodel is followed. Finally, the validity of the proposed approach is checked using five typical mathematical examples.

#### 4.3.1. Transformation of Correlated Input Variables to Uncorrelated Input Variables

For the semiconductor package products, if exist, the statistical dependencies are generally linear. In the probability theory and statistics, the degree of linear dependency and the joint variability of two random variables is measured by covariance [118], which can be expressed as:

$$\mathrm{Cov}\left(x_i, x_j\right) = \Sigma_{ij} = E\left[\left(x_i - \mu_i\right)\left(x_j - \mu_j\right)\right] \tag{55}$$

where $E[\cdot]$ is the expectation operator; $\mu_i$ and $\mu_j$ are the mean values of the input variable $x_i$ and $x_j$, respectively.

For $N$ input variables, the covariance of each pair and the variance of each input variable can be summarized into a matrix, called the covariance matrix, which is expressed as [22]:

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \Sigma_{12} & \Sigma_{13} & \cdots & \Sigma_{1N} \\ \Sigma_{21} & \sigma_2^2 & \Sigma_{23} & \cdots & \Sigma_{2N} \\ \Sigma_{31} & \Sigma_{32} & \sigma_3^2 & \cdots & \Sigma_{3N} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \Sigma_{N1} & \Sigma_{N2} & \Sigma_{N3} & \cdots & \sigma_N^2 \end{bmatrix} \tag{56}$$

where $\sigma_1^2, \ldots, \sigma_N^2$ are the variance of the $N$ input variables and $\Sigma_{ij} = \Sigma_{ji}$ is the covariance of the input variables $x_i$ and $x_j$. When two input variables are uncorrelated, $\Sigma_{ij} = \Sigma_{ji} = 0$. When the input variables $x_i$ and $x_j$ are correlated, the covariance $\Sigma_{ij}$ should be determined through either measurements [119] or statistical analyses [7].

The covariance matrix can be used to convert the correlated variables into a set of linearly uncorrelated variables using the eigen-decomposition of covariance matrix. The eigen-decomposition of covariance matrix has been widely used in the design community. For example, the principal component analysis (PCA) [120] uses the eigenvalues and eigenvectors of the covariance matrix to explain the data with fewer variables, which best explain the variance in the data. An advanced uncertainty propagation analysis, called eigenvector dimension reduction (EDR) method [22], chooses sample points along the eigenvectors to calculate the statistical moments, because the eigenvectors are orthogonal and reveal the correlation of input variables.

The same concept is adopted in this study to handle the statistical correlation of input variables. The eigen-decomposition is performed by solving the eigenvalue problem (i.e., $\Sigma U = \lambda U$ ) of the covariance matrix to obtain the eigenvalues, $\lambda$ , and the corresponding eigenvectors, $U$ .

The matrix consists of $N$ eigenvectors indicate a new coordinate system such that $N$ new input variables are statistically independent, and thus, the transformation from correlated input

variables, $\vec{x} = \left[x_1, \ldots, x_N\right]^T$, to uncorrelated input variables, $\vec{z} = \left[z_1, \ldots, z_N\right]^T$, can be achieved by coordinates transformation as:

$$\begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ \vdots \\ z_N \end{bmatrix} = \begin{bmatrix} u_{11} & u_{12} & u_{13} & \cdots & u_{1N} \\ u_{21} & u_{22} & u_{23} & \cdots & u_{2N} \\ u_{31} & u_{32} & u_{33} & \cdots & u_{3N} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ u_{N1} & u_{N2} & u_{N3} & \cdots & u_{NN} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_N \end{bmatrix} \tag{57}$$

where $u_{ij} = 0$, if the $\left(x_i,\ x_j\right)$ pair does not have statistical correlation.

After the transformation, the new range of design space along each new input variables can be obtained as $\mu_{z_i} \pm 3\sqrt{\lambda_i}$, where $\mu_{z_i}$ is the mean vector the new input variables but is equal to the mean vector the original input variables, $\mu_i$; $\sqrt{\lambda_i}$ is the standard deviation of each new input variable.

The following ranking and partitioning of bivariate terms will be conducted using the new input variables and design space.

### 4.3.2. Ranking and Partitioning of Bivariate Terms using HDMR-Factorial Design Hybrid Method

As mentioned before, the bivariate terms in the Cut-HDMR representing the interaction effect of two input variables on the performance response. A factorial experimental designs is the most common method to evaluate interaction effects. The results from a factorial design are often analyzed using the effect model of analysis of variance (ANOVA) [113, 121]. In the two level factorial design with $N$ input variables (referred to as $2^N$ design), each input variable takes two discrete values referred to as "high" and "low" levels, which are often represented as "+1" and "−1" using an indicator variable $d_i$ for the $i^{\text{th}}$ input variable, $x_i$ [122]. Define $\vec{d}_k = \left[d_{1_k}, \ldots, d_{N_k}\right]$

is the $k^{th}$ combination of the indicator variables (referred to as the $k^{th}$ treatment), the ANOVA effect model can be expressed as [121, 123]:

$$y_k = \beta_0 + \sum_{i=1}^{N} \beta_i d_{i_k} + \sum_{i=1}^{j-1} \sum_{j=2}^{N} \beta_{ij} d_{i_k} d_{j_k} + \ldots + \beta_{i_1 \ldots i_N} d_{i_1_k} \cdots d_{i_N_k} \tag{58}$$

where $y_k$ is the performance response under the $k^{th}$ treatment; $\beta_0$ is the mean response; $\beta_i$ is the main effect of $x_i$,; $\beta_{ij}$ is the interaction effect of $x_i$ and $x_j$, and so on.

Eq. (58) can be rewritten as [123]:

$$y_k - \beta_0 = D = \sum_{i=1}^{N} \beta_i d_{i_k} + \sum_{i=1}^{j-1} \sum_{j=2}^{N} \beta_{ij} d_{i_k} d_{j_k} + \ldots + \beta_{i_1 \ldots i_N} d_{i_1_k} \cdots d_{i_N_k} \tag{59}$$

In the above equation, $D$ is the deviation of performance response from the mean response while subjected to the $k^{th}$ treatment (reffered to as the total variability), which can be decomposed into a sum of the deviations contributed by each input variable acting individually $(\beta_i)$, two input variables acting together $(\beta_{ij})$, and so on (refferred to as the partial variabilitys).

By normalizing the absolute values of partial variabilities using the absolute value of total variability, an ANOVA-based sensitivity indices are defined as [123]:

$$S_{i_1 \ldots i_N} = \frac{\left|\beta_{i_1 \ldots i_N}\right|}{|D|} \tag{60}$$

where $S_i = |\beta_i| / |D|$ is the first-order sensitivity index that provides a measure about the ratio of the variability contributed by $x_i$ alone to the total variability; $S_{ij} = |\beta_{ij}| / |D|$ is the second-order sensitivity index representing the ratio of the partial variability contributed by the interaction between $x_i$ and $x_j$ to the total variability.

It is worth noting that the ANOVA model is conceptually the same as the HDMR approach, which can be seen as an HDMR expansion considering the linear parts of every component

function. Therefore, the sensitivity indiecs can be used to partition the bivariate terms of Cut-HDMR in two steps: (1) ranking the first-order and second-order sensitivity indices from the maximum to minimum, and (2) converting the maximum allowable prediction error to the percentage of the total variability as the partition criteria. For example, if the total variability is 30 μm and the maximum allowable prediction error is set as 1.5 μm, then, the partition criterion becomes 5%, and thus, the sum of ranked effects up to (100% − 5%) are the significant terms.

In general, $2^N$ sample points are required to estimate every coefficient in Eq. (58) [113, 122], which makes the computational cost increase rapidly as the number of input variables increases. When the goal is to distinctively evaluate the two-variable interaction effects, one can select a fraction of the $2^N$ sample points to estimate the corresponding coefficients, called the fractional factorial design with resolution V [122]. It is, however, not always available for high dimensional problems.

Due to the hierarchical structure of Cut-HDMR, the univariate terms must to be obtained before the construction of the bivariate terms, and thus, this paper proposes an approach to estimate the coefficients of the ANOVA effect model using the existing univariate terms with only one additional modeling run per each bivariate pair.

Along each input variable, five sample points are located at the mean value, $\mu_i$, upper bound, $z_{iU}$, lower bound, $z_{iL}$, and the points at $\mu_i + \left(z_{iU} - \mu_i\right)/2$ and $\mu_i - \left(\mu_i - z_{iL}\right)/2$ of the input variable, $z_i$. Five sample points with MLS interpolation has been proven to provide sufficient accuracy for structural analysis [74], and thus, the five sample points are adopted in this study. After obtaining the performance responses of the sample points of the univariate terms, the main effects can be estimated as:

$$\left|\hat{\beta}_i\right| = \boldsymbol{max}\left(\left|y\left(z_{iU}, \bar{\mu}^i\right) - y\left(\bar{\mu}\right)\right|, \left|y\left(z_{iL}, \bar{\mu}^i\right) - y\left(\bar{\mu}\right)\right|\right) \tag{61}$$

where $y(\bar{\mu})$, $y(z_{iU}, \bar{\mu}^i)$ and $y(z_{iL}, \bar{\mu}^i)$ are the performance responses at the corresponding sample points. Eq. (61) determines which value between the effect caused by $z_{iU}$ and $z_{iL}$ is larger. one of which causes a larger effect is denoted as $z_{i_main}$.

For each pair of $x_i$ and $x_j$, an additional modeling run is performed at the $(z_{i_main}, z_{j_main})$, because at least one of the parent main effects should be sufficiently large to make an interaction effect signigicant, which is called the effect heredity principle [113, 114]. Then, the interaction effect can be estimated by:

$$\hat{\beta}_{ij} = y(z_{i_main}, z_{j_main}, \bar{\mu}^{ij}) - \left[ y(z_{i_main}, \bar{\mu}^i) + y(z_{j_main}, \bar{\mu}^j) - y(\bar{\mu}) \right] \tag{62}$$

where $\left[ y(z_{i_main}, \bar{\mu}^i) + y(z_{j_main}, \bar{\mu}^j) - y(\bar{\mu}) \right]$ is the superposition of the main effects and the mean effects.

Figure 51 illustrates the proposed HDMR-factorial design (HDMR-FD) hybrid method. In this figure, the red dots represent the modeling runs for the univariate terms, which are used to produce the blue meshed surface by using superposition and interpolation. The black meshed surface is the underlying function on the $(x_i, x_j)$ plane, and the green diamond is the perfomance reseponse obtained from the additional modeling run. The differece between the green diamond and the blue surface is the interaction effect, which is equvilent to the $\beta_{ij}$ of the ANOVA effect model.

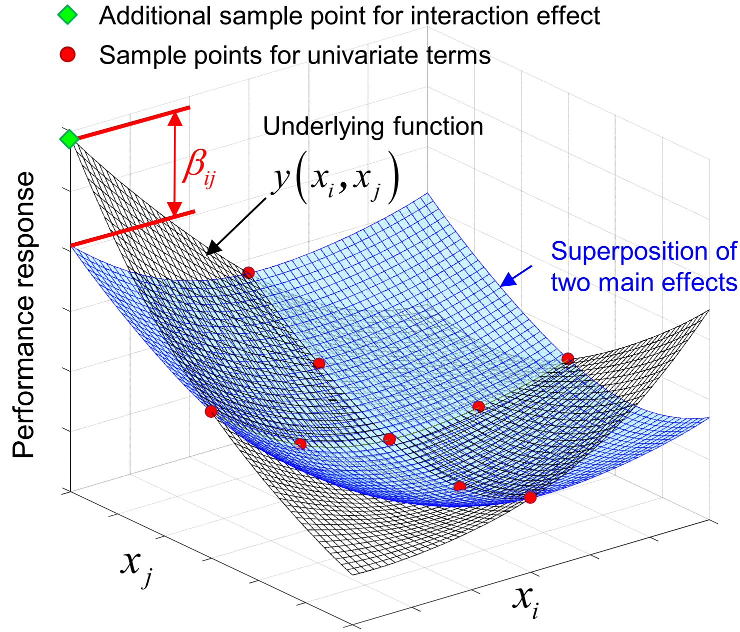



Figure 51. Illustration of HDMR-FD hybrid method

The total variability, $D$, can be obtained by performing one additional run, which can be expressed as:

$$D = y\left(z_{i_1_main}, \ldots, z_{i_N_main}\right) - y\left(\vec{\mu}\right) \tag{63}$$

Finally, the sensitivity indices can be calculated uisng Eq. (60) to (63) for ranking and partitioning the bivariate terms.

### 4.3.3. Procedure of PB Cut-HDMR

A detailed procedure of PB Cut-HDMR is described below. A numerical example with 10 input variables [109] accompanies the procedure to provide a better understanding. The response function of this example is expressed as:

$$\begin{aligned} y(x) = {} & x_1^2 + x_2^2 + x_1 x_2 - 14x_1 - 16x_2 + (x_3 - 10)^2 + 4(x_4 - 10)^2 + (x_5 - 3)^2 \\ & + 2(x_6 - 1)^2 + 5x_7^2 + 7(x_8 - 11)^2 + 2(x_9 - 10)^2 + (x_{10} - 7)^2 + 45 \end{aligned} \tag{64}$$

where all variables are independent and ranged between -10 to 11, and only the $x_1x_2$ pair has an interaction effect.

(1) Check if any statistical correlation exists among the input variables. If yes, perform the eigen-decomposition of the covariance matrix to transform the correlated input variables, $x_i$, to the uncorrelated input variables, $z_i$. There is no statistical correlations among the 10 input variables in the example, and thus, $z_i = x_i$.

(2) Perform numerical modeling at the point, $\bar{\mu} = [\mu_1, \mu_2, \ldots, \mu_N]$, to obtain $y(\bar{\mu})$, where each input variable takes its mean value. In the example, $y(\bar{\mu})$ = 1204.5 at the point, $\bar{\mu} = [0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5]$.

(3) Perform numerical modeling at the total of 4*N* sample points for the *N* univariate terms and estimate the main effects using Eq. (61). Table 23 lists $z_{i_main}$ and the corresponding performance responses and main effects.

Table 23 $z_{i_main}$ and corresponding main effects

| Variables | $z_{i_main}$ | Performance response | Main effects |
|---|---|---|---|
| $x_1$ | -10 | 1446.00 | 241.50 |
| $x_2$ | -10 | 1467.00 | 262.50 |
| $x_3$ | -10 | 1514.25 | 309.75 |
| $x_4$ | -10 | 2023.50 | 819.00 |
| $x_5$ | -10 | 1367.25 | 162.75 |
| $x_6$ | -10 | 1446.00 | 241.50 |
| $x_7$ | 11 | 1808.25 | 603.75 |
| $x_8$ | -10 | 3519.75 | 2315.25 |
| $x_9$ | -10 | 1824.00 | 619.50 |
| $x_{10}$ | -10 | 1451.25 | 246.75 |

(4) For each pair of $x_i$ and $x_j$, perform numerical modeling at the $\left[z_{i_main}, z_{j_main}, \vec{\mu}^{ij}\right]$ and estimate the interaction effect using Eq. (62). A total of 45 bivariate terms exist for 10 input variables. For illustration, Table 24 lists the results of a pair having interaction effect ($x_1x_2$ pair) and a pair having no interaction ($x_1x_3$ pair). Other pairs have the same interaction effect as the $x_1x_3$ pair.

Table 24 Illustration of the results of the pair having interaction effect and the pair having no interaction

| Bivariate pair | $\left[z_{i_main}, z_{j_main}\right]$ | Performance response | Superposition of main effects | Interaction effects |
|---|---|---|---|---|
| $x_1x_2$ | [-10, -10] | 1818.75 | 1708.50 | 110.25 |
| $x_1x_3$ | [-10, -10] | 1755.75 | 1755.75 | 0 |

(5) Perform numerical modeling at $\left[z_{i_1_main}, \ldots, z_{i_N_main}\right]$ to estimate the total variability, $D$, using Eq. (63). In this example, $z_{7_main} = 11$ and all other nine values are equal to $-10$. The performance response, $y\left(z_{i_1_main}, \ldots, z_{i_N_main}\right)$ = 7137.

(6) Calculate the sensitivity indices and rank the main effects and the interaction effects. The results from the example are listed in Table 25.

Table 25 First order and second order sensitivity indices

| Variable | Sensitivity index | Variable | Sensitivity index | Variable | Sensitivity index |
|---|---|---|---|---|---|
| $x_8$ | 0.390 | $x_3$ | 0.052 | $x_6$ | 0.041 |
| $x_4$ | 0.138 | $x_2$ | 0.044 | $x_5$ | 0.027 |
| $x_9$ | 0.104 | $x_{10}$ | 0.042 | $x_1x_2$ | 0.019 |
| $x_7$ | 0.102 | $x_1$ | 0.041 | $x_1x_3, \ldots, x_9x_{10}$ | 0 |

(7) Partition the bivariate terms based on the application specific tolerance. In this example, 100% of the total variability is contributed by the main effect plus the only one significant bivariate term.

(8) Perform numerical modeling at the uniformly spaced sample points for the significant bivariate terms. The approximated response surface (the green surface) using the sample points (the red dots) with MLS interpolation is compared with the underlying function (the black meshed surface) of the $x_1x_2$ pair in Figure 52. The results clearly show that they are virtually identical.

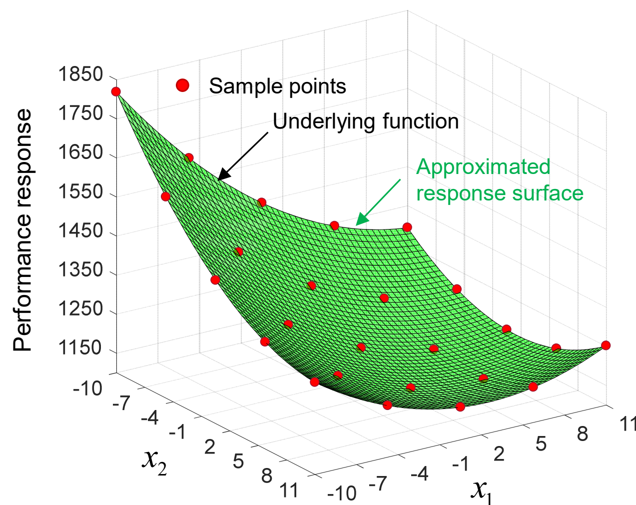



Figure 52. Comparison of true response and approximated response surface with 2-D MLS interpolation for significant bivariate terms

(9) Approximate the bivariate terms with minor interaction by superposing the corresponding univariate terms. In Figure 53, the true response surface of the $x_1x_3$ pair (the black meshed surface) is compared with the approximated surface (the light blue surface). The approximated surface is obtained by the superposition of the univariate terms of $x_1$ and $x_3$, which are interpolated through the nine sample points (red dots). It is clear from the results

that the 2-D metamodel can be accurately constructed even when the underlying function is nonlinear as long as the interaction effect is minor.

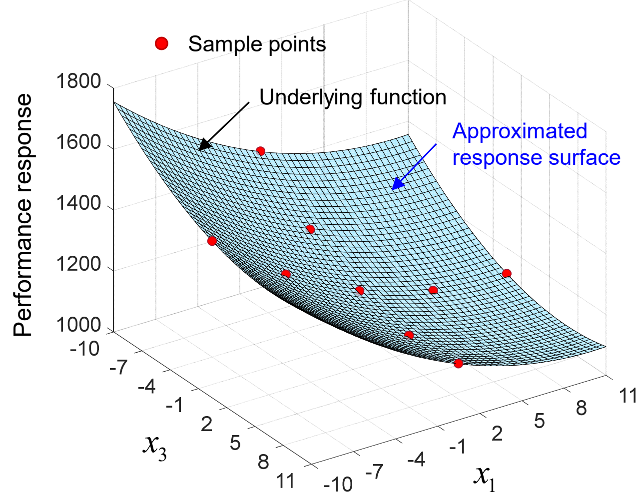



Figure 53. Comparison of true response and approximated response surface from superposition of the corresponding main effects for minor bivariate terms

(10) Construct the metamodel using Eq. (53).

Based on the procedure, the overall computational cost of the PB Cut-HDMR can be expressed as:

$$c = \left[ \frac{N(N-1)}{2} + 1 \right] + N_s \left[ (s-1)^2 - 1 \right] + N(s-1) + 1 \tag{65}$$

where $s$ is the number of sample points taken along the direction of each input variable, $N$ is the number of input variables, and $N_s$ is the number of significant bivariate terms.

A simplified DoE case with 2 levels was conducted considering three variables, $x_1$, $x_2$, and $x_8$ as the design variables to evaluate the accuracy of the metamodel. Table 26 shows the prediction results of three approaches: the univariate and bivariate Cut-HDMR as well as the PB Cut-HDMR. As expected, the predictions of the univariate Cut-HDMR deviated from the true

value since the interaction effect was ignored. On the other hand, the PB Cut-HDMR and the bivariate Cut-HDMR predicted the true response in this numerical example. It is important to note that the PB Cut-HDMR used only 102 modeling runs without compromising accuracy, which was substantially smaller than 761 modeling runs required for the bivariate Cut HDMR.

Table 26 Comparison of the prediction results of univariate Cut-HDMR, bivariate Cut-HDMR and PB Cut-HDMR

| DoE Leg | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| $x_1$ | -5 | -5 | -5 | -5 | 6 | 6 | 6 | 6 |
| $x_2$ | -9 | -9 | 8 | 8 | -9 | -9 | 8 | 8 |
| $x_8$ | -6 | 10 | -6 | 10 | -6 | 10 | -6 | 10 |
| Performance responses | | | | | | | | |
| True | 2,835 | 2,593 | 2,461 | 2,406 | 819 | 577 | 445 | 390 |
| Uni. Cut-HDMR | 2,783 | 2,645 | 2,502 | 2,365 | 767 | 629 | 486 | 349 |
| Bi. Cut-HDMR | 2,835 | 2,593 | 2,461 | 2,406 | 819 | 577 | 445 | 390 |
| PB Cut-HDMR | 2,835 | 2,593 | 2,461 | 2,406 | 819 | 577 | 445 | 390 |

### 4.3.4. Validity Check of PB Cut-HDMR

The performance metrics widely used to evaluate the efficiency and accuracy of a metamodel is described first. The validity check results are followed.

#### 4.3.4.1. Performance Metric

As the critical requirement of metamodeling for high dimensional problems, the model efficiency is commonly represented by the number of modeling runs (i.e., sample points).

For the model accuracy, three performance metrics, namely R square, relative average absolute error (RAAE) and relative maximum absolute error (RMAE), are widely used to verify the validity of metamodeling techniques [100, 109]. The metrics are described below.

**(1) R square**

$$R^2 = 1 - \frac{\sum_{i=1}^{m}\left[y(x_i) - \hat{y}(x_i)\right]^2}{\sum_{i=1}^{m}\left[y(x_i) - \bar{y}(x_i)\right]^2} \tag{66}$$

where $y(x_i)$ is a performance response at the $i^{th}$ new sample point used for validity check; $\hat{y}(x_i)$ is an approximated performance response at the $i^{th}$ new sample point; and $\bar{y}(x_i)$ is the mean of all $y(x_i)$. R square indicates the overall accuracy of a metamodel. Its value is a quantitative indication of metamodel accuracy, and the maximum value is 1.

**(2) RAAE**

$$RAAE = \frac{\sum_{i=1}^{m}\left|y(x_i) - \hat{y}(x_i)\right|}{m \times STD} \tag{67}$$

where STD is the standard deviation of all $y(x_i)$. Similar to R square, RAAE represents the overall accuracy, but its value closer to zero indicates better accuracy.

**(3) RMAE**

$$RMAE = \frac{\boldsymbol{max}\left(\left|y(x_1) - \hat{y}(x_1)\right|, \left|y(x_2) - \hat{y}(x_2)\right|, \ldots, \left|y(x_m) - \hat{y}(x_m)\right|\right)}{STD} \tag{68}$$

The RMAE represents the local accuracy by measuring the maximum error among the sample points. Thus, the smaller value is desired.

To be consistent with Ref. [109], 1000 new sample points were generated by MCS to evaluate the accuracy of the PB Cut-HDMR.

#### 4.3.4.2. Numerical Examples

In Ref. [109] and [124], a total of 20 mathematical test functions were selected to test the effectiveness of metamodel for high dimensional problems. The test functions that were not applicable to the objective of this paper were excluded: (1) equally important two-variable

interactions, (2) no interactions, and (3) dominant high-order interactions. The five test functions used for the comparison study are listed in Table 27.

Table 27 Test functions

| No. | Function | Design space |
|---|---|---|
| 1 | $y(x)=\sum_{i=1}^{10} x_i\left[c_i+\ln\left(\frac{x_i}{x_1+\cdots+x_{10}}\right)\right]$, where $c_{i=1,\ldots,5}=[-6.089, -17.164, -34.054, -5.914, -24.721]$ $c_{i=5,\ldots,10}=[-14.986, -24.100, -10.708, -26.662, -22.179]$ | $1E-6\le x_i\le 10$, $i=1,\ldots,10$ |
| 2 | $y(x)=x_1^2+x_2^2+x_1x_2-14x_1-16x_2+(x_3-10)^2+4(x_4-10)^2+(x_5-3)^2+2(x_6-1)^2+5x_7^2+7(x_8-11)^2+2(x_9-10)^2+(x_{10}-7)^2+45$ | $-10\le x_i\le 11$, $i=1,\ldots,10$ |
| 3 | $y(x)=(x_1-1)^2+(x_{10}-1)^2+10\sum_{i=1}^{9}(10-i)\left(x_i^2-x_{i+1}\right)^2$ | $-3\le x_i\le 2$, $i=1,\ldots,10$ |
| 4 | $y(x)=\sum_{i=1}^{16}\sum_{j=1}^{16} a_{ij}\left(x_i^2+x_i+1\right)\left(x_j^2+x_j+1\right)$, where $a_{ij}=\begin{bmatrix} 1&0&0&1&0&0&1&1&0&0&0&0&0&0&0&1\\ 0&1&1&0&0&0&1&0&0&1&0&0&0&0&0&0\\ 0&0&1&0&0&0&1&0&1&1&0&0&0&1&0&0\\ 0&0&0&1&0&0&1&0&0&0&1&0&0&0&1&0\\ 0&0&0&0&1&1&0&0&0&1&0&1&0&0&0&1\\ 0&0&0&0&0&1&0&1&0&0&0&0&0&0&1&0\\ 0&0&0&0&0&0&1&0&0&0&1&0&1&0&0&0\\ 0&0&0&0&0&0&0&1&0&1&0&0&0&0&1&0\\ 0&0&0&0&0&0&0&0&1&0&0&1&0&0&0&1\\ 0&0&0&0&0&0&0&0&0&1&0&0&0&1&0&0\\ 0&0&0&0&0&0&0&0&0&0&1&0&1&0&0&0\\ 0&0&0&0&0&0&0&0&0&0&0&1&0&1&0&0\\ 0&0&0&0&0&0&0&0&0&0&0&0&1&1&0&0\\ 0&0&0&0&0&0&0&0&0&0&0&0&0&1&0&0\\ 0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&0\\ 0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1 \end{bmatrix}$ | $0\le x_i\le 5$, $0\le x_j\le 5$, $i, j=1,\ldots,16$ |
| 5 | $y(x)=\sum_{i=1}^{5}\left[(x_i+10x_{i+5})^2+5(x_{i+10}-x_{i+15})^2+(x_{i+5}-2x_{i+10})^4+10(x_i-x_{i+15})^4\right]$ | $-2\le x_i\le 5$, $i=1,\ldots,20$ |

Table 28 summarizes the values of three performance metrics and the number of sample points (denoted as NoP) for PB Cut-HDMR and the full bivariate Cut-HDMR. The results indicate

that PB Cut-HDMR provides virtually the same accuracy as the full bivariate Cut-HDMR, but offers significant improvement in efficiency.

Table 28 Comparison of performance metrics for full bivariate Cut-HDMR and PB Cur-HDMR

| Full bivariate Cut-HDMR | | | | |
|---|---|---|---|---|
| Function | R square | RAAE | RMAE | NoP |
| 1 | 1.00 | 0.002 | 0.014 | 761 |
| 2 | 1.00 | 0.000 | 0.000 | 761 |
| 3 | 1.00 | 0.034 | 0.125 | 761 |
| 4 | 1.00 | 0.002 | 0.006 | 1,985 |
| 5 | 1.00 | 0.026 | 0.074 | 3,121 |
| PB Cut-HDMR | | | | |
| Function | R square | RAAE | RMAE | NoP |
| 1 | 1.00 | 0.002 | 0.013 | 87 |
| 2 | 1.00 | 0.000 | 0.000 | 102 |
| 3 | 1.00 | 0.033 | 0.105 | 222 |
| 4 | 1.00 | 0.002 | 0.005 | 636 |
| 5 | 1.00 | 0.028 | 0.122 | 422 |

## 4.4. Implementation of PB Cut-HDMR: Warpage Prediction of Thin Flat Ball Grid Array

In this section, the PB Cut-HDMR was implemented to construct the metamodel for warpage prediction of a thin flat ball grid array (TFBGA) involving 12 input variables. The finite element model and the design space are described first. The metamodel construction and the accuracy check results are followed.

### 4.4.1. Problem Description

The thin flat ball grid array (TFBGA) package is often used as the top package of a Package-on-Package (PoP). The warpage of TFBGA at the solder reflow temperature is a critical

factor for a successful PoP stacking. Figure 54 shows details of the finite element model with the boundary conditions build by using a commercial FEA package (ANSYS®). A quarter symmetry was applied. The top TFBGA package contains two dies and the die stack configuration is shown the white dashed lines.

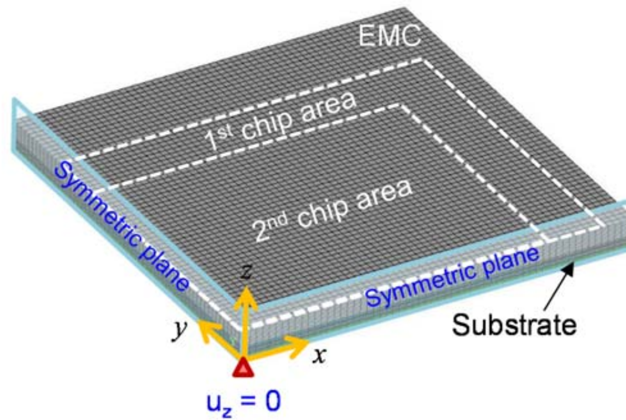



Figure 54. Configuration of TFBGA assembly and details of the FEA model

The material properties and the nominal dimensions used in the models are summarized in Table 29 and Table 30. All materials are assumed to be linear elastic in order to conduct the accuracy check for demonstration purpose.

Table 29 Material properties

| Material | Young's modulus (GPa) | Poisson's Ratio | CTE (ppm/°C) | | $T_g$ (°C) |
|---|---|---|---|---|---|
| | | | $\alpha_1$ (< $T_g$) | $\alpha_2$ (> $T_g$) | |
| Silicon die | 130 | 0.23 | 2.8 | | -- |
| DAF | 2.2 @ 25°C<br>1.0 @ 100°C<br>0.008 @ 100°C | 0.3 | 65.3 | 162.9 | 138 |
| Substrate | 46.794 | 0.3 | 16.2 (in-plane)<br>61.5 (out-of-plane) | | -- |
| EMC | 29.237 @ 25°C<br>14.030 @ 125°C<br>1.932 @ 175°C<br>1.498 @ 235°C | 0.21 | 9.12 | 35.13 | 137.5 |

Table 30  Dimensions

| | length × width × thickness |
|---|---|
| 1st Die (mm) | 13 × 11 × 0.575 |
| 1st DAF (mm) | 13 × 11 × 0.025 |
| 2nd Die (mm) | 11 × 9 × 0.575 |
| 2nd DAF (mm) | 11 × 9 × 0.025 |
| Substrate (mm) | 15 × 15 × 0.13 |
| EMC (mm) | 15 × 15 × 0.59 |

The top package was subjected to the EMC molding process at 175 °C, which was used as a stress free temperature.  The conventional lead-free solder reflow profile with the peak temperature as 260 °C was considered [28].

In this study, a total of 12 input variables were considered for the TFBGA package.  Their means and standard deviations were obtained from the literature as well as the manufacturing specifications [31, 60-63].  The design space is defined as the mean ± three standard deviations and summarized in Table 31.  Among these 12 input variables, the EMC thickness, $x_3$, and the substrate thickness, $x_4$, has a statistical correlation because their sum should be equal to the cavity height of the transfer mold [7].  In addition, the CTEs of the EMC below and above $T_g$, $x_9$ and $x_{10}$, are measured from the same samples, and thus, they should have strong positive correlation, which can be presented by a correlation coefficient of "unity".

Table 31  Input variables and design space

| Variable | Physical meaning | Mean | Std. Dev. | Design space | Correlation coefficient |
|---|---|---|---|---|---|
| $x_1$ | PKG length (mm) | 15 | 0.033 | [14.9, 15.1] | -- |
| $x_2$ | PKG width (mm) | 15 | 0.033 | [14.9, 15.1] | |
| $x_3$ | EMC thickness (mm) | 0.59 | 0.029 | [0.502, 0.678] | -0.35 |
| $x_4$ | PCB thickness (mm) | 0.13 | 0.01 | [0.102, 0.158] | |
| $x_5$ | 1st Chip thickness (mm) | 0.0575 | 0.001 | [0.0545, 0.0605] | -- |
| $x_6$ | 2nd Chip thickness (mm) | 0.0575 | 0.001 | [0.0545, 0.0605] | -- |
| $x_7$ | 1st DAF thickness (mm) | 0.025 | 0.00375 | [0.01375, 0.3625] | -- |
| $x_8$ | 2nd DAF thickness (mm) | 0.025 | 0.00375 | [0.01375, 0.3625] | -- |
| $x_9$ | EMC CTE above $T_g$ (ppm/°C) | 35.13 | 4.24 | [21.989, 48.271] | 1 |
| $x_{10}$ | EMC CTE below $T_g$ (ppm/°C) | 9.12 | 1.1 | [9.12, 9.12] | |
| $x_{11}$ | PCB CTE (ppm/°C) | 16.2 | 0.81 | [13.77, 18.63] | -- |
| $x_{12}$ | PCB modulus (MPa) | 46794 | 159 | [46317, 47271] | -- |

### 4.4.2. Metamodel Construction by PB Cut-HDMR

The same procedure as shown in Section 4.3.3 was carried out. The difference is that the TFBGA assembly has two pairs of correlated input variables. Figure 55(a) and (b) show the transformed variables after eigen-decomposition of the covariance matrix for the $x_3x_4$ pair and $x_9x_{10}$ pair, respectively, where the red and blue lines indicate the directions before and after transformation, respectively. Moreover, in the figure, the light red box indicates the design space before transformation, while the light blue box is the design space after transformation. The green dots are the possible design points generated while considering the statistical correlation. It is clear that the design spaces after transformation have much less waste area than the design spaces ignoring the statistical correlation.

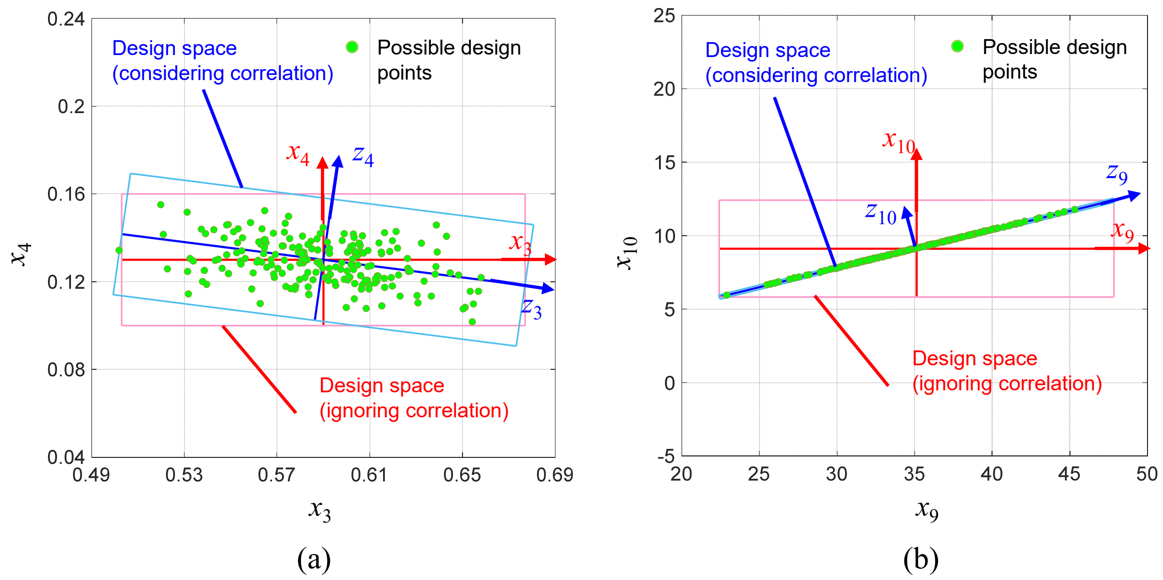



Figure 55. Transformation of input variable directions and design space change after eigen-decompostion of covariance matrix. (a) $x_3x_4$ pair and (b) $x_9x_{10}$ pair.

For 12 input variables, simulations at 49 sample points were performed for the univariate terms while taking 5 sample points along each input variable direction. Figure 56 plots the performance responses along each input variable. A low degree of nonlinearity is clearly seen, which is beneficial to the metamodel construction.

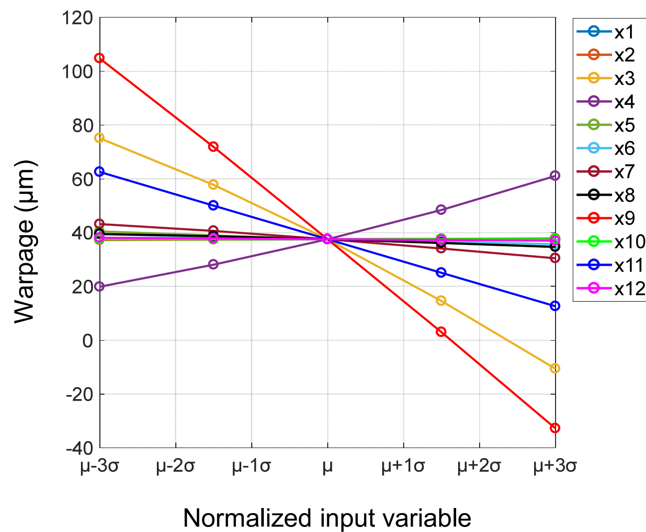



Figure 56. Warpages at the sample points along each input variable

Then, 67 modeling runs were conducted to calculate the sensitivity indices. The result is listed in Table 32. It is worth noting that there is no clear cut of the sensitivity indices. Setting the maximum allowable error is 5 µm, which is about 2.5% of the total variability of 198.6 µm. Accordingly, the 12 main effects plus the first 16 pairs of the bivariate terms are expected to explain 97.5% of the total variability and meet the criterion. Each of the 16 terms needs 15 additional modeling runs to construct the 2-D metamodel, and thus, 240 modeling runs are required. In summary, the PB Cut-HDMR took a total of 356 (49 + 67 + 240) modeling runs.

Table 32 Sensitivity indices of the TFBGA warpage prediction

| Variable | $S_i$ | Variable | $S_{ij}$ | Variable | $S_{ij}$ |
|---|---|---|---|---|---|
| $z_9$ | 0.3407 | $z_4z_9$ | 0.0273 | $z_3z_5$ | 0.0032 |
| $z_3$ | 0.1903 | $z_3z_{11}$ | 0.0223 | $z_5z_6$ | 0.0015 |
| $z_{11}$ | 0.1270 | $z_3z_9$ | 0.0219 | $z_3z_7$ | 0.0015 |
| $z_4$ | 0.1186 | $z_4z_7$ | 0.0117 | $z_2z_3$ | 0.0012 |
| $z_7$ | 0.0283 | $z_7z_9$ | 0.0094 | $z_8z_{11}$ | 0.0009 |
| $z_5$ | 0.0146 | $z_8z_9$ | 0.0063 | $z_1z_3$ | 0.0009 |
| $z_8$ | 0.0099 | $z_4z_{11}$ | 0.0062 | $z_2z_4$ | 0.0008 |
| $z_6$ | 0.0093 | $z_3z_4$ | 0.0044 | $z_5z_{11}$ | 0.0008 |
| $z_{12}$ | 0.0028 | $z_5z_9$ | 0.0043 | $z_7z_{11}$ | 0.0008 |
| $z_2$ | 0.0014 | $z_4z_8$ | 0.0038 | $z_1z_4$ | 0.0008 |
| $z_1$ | 0.0005 | $z_6z_9$ | 0.0035 | $z_{11}z_{12}$ | 0.0008 |
| $z_{10}$ | 0 | $z_9z_{11}$ | 0.0033 | *rest terms* | $\leq 0.0007$ |

Besides the sample points for metamodel construction, 500 new sample points were generated by MCS to assess the accuracy of PB Cut-HDMR in this real semiconductor application. The R square was 1.00, RAAE was 0.009, and the RMAE is 0.039. These results show that the PB Cut-HDMR maintains the same level of accuracy in the real application and the numerical

examples. Moreover, the standard deviation of the warpages at the 500 new sample points is 51 μm. Back calculated from RMAE, the maximum absolute error of these 500 sample points is 1.99 μm, which is, as expected, less than the maximum allowable error (5 μm). Figure 57 shows the histogram of the errors, where 91.4 % of the sample points has the absolute error smaller than 1 μm.

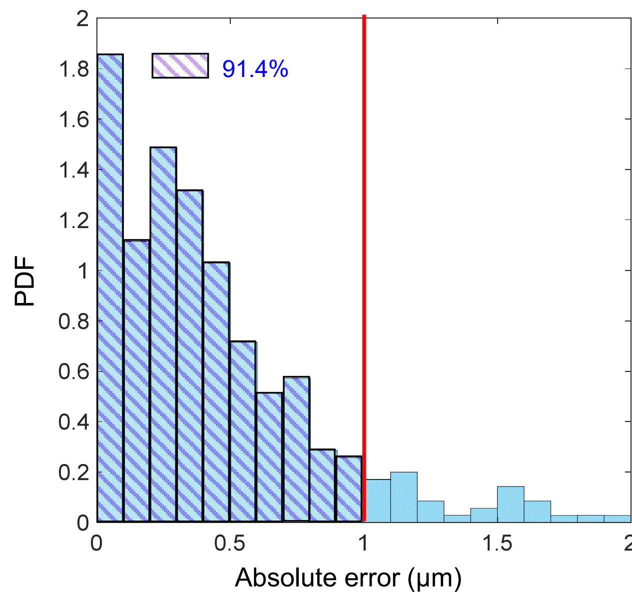



Figure 57. Histogram of errors at the sample points for validity check

### 4.5. Conclusion

This paper proposed an advanced metamodeling technique, called PB Cut-HDMR, for the semiconductor packaging products, which has a large number of input variables with both statistical correlation among input variables and interaction effects on the performance response. The PB Cut-HDMR utilizes the eigen-decomposition of the covariance matrix to handle the correlated input variables, which has not been solved by the current metamodeling techniques. In addition, a HDMR factorial design hybrid method is developed to effectively and deterministically partition the bivariate terms of HDMR into significant pairs and minor pairs and spending the

expensive numerical modeling on the significant bivariate terms only, and thus, the "curse of dimensionality" is alleviated or circumvented.

The performance of PB Cut-HDMR was evaluated though five typical numerical examples and compared with full bivariate Cut-HDMR and confirmed to have virtual identical accuracy but substantially improvement was made for efficiency.

The proposed method was implemented into a real semiconductor package problems for prediction of warpage at reflow temperature. A total of 12 input variables and two pairs of correlated input variables are considered. The accuracy and efficiency of the proposed method in real application was revealed through this demonstration.

Future research aims to effectively evaluate the nonlinearity of each terms and improve the sampling strategy based on nonlinearity for further reduction of computational cost. In addition, an extension of the HDMR-FD method to handle an even larger number of input variables (say, > 20) is also part of the future research.

# CHAPTER 5. CONTRIBUTIONS AND FUTURE WORKS

## 5.1. Dissertation Contributions

Stochastic reliability modeling capabilities are developed and implemented for semiconductor packaging problems with a very large number of input variables (> 10 input variables) in three critical areas for semiconductor packaging product development: (1) prediction of tail-end probability (i.e., assembly yield loss) by advanced uncertainty propagation (UP) analyses, (2) determination of the statistical distributions of unknown design and/or manufacturing parameters by advanced statistical model calibrations, and (3) determination of the performance response of high-dimensional problems by developing an advanced metamodeling scheme. The most significant contributions made in this dissertation are summarized below:

- An comprehensive stochastic model is proposed and implemented to predict PoP stacking yield loss. The major contributions are (1) to cope with the number of input variables which has been conceived as the practical limit with 25 times to thousands times reduction of modeling runs for the tail-end probability prediction of semiconductor package products and (2) to propose a model utilizing the PDFs of five critical performance responses (the warpages and the solder ball heights of the top and bottom packages, and the solder joint height of the corner pad) at hundreds pad locations at the stacking interface to take into account their statistical variations and correlations for accurate yield loss prediction. The model can be used effectively to control the input uncertainties, and thus to achieve a yield goal for a given set of PoP designs.
- The unknown statistical distributions of two effective elastic properties of SAC305 solder joint of leadless chip resistors (LCRs), induced by an assembly condition, are inversely determined from the full set of cyclic bending test results of LCR assemblies using the advanced statistical model calibration. During the calibration, the effects of other known input variables are taken

into account by employing the EDR method for UP analysis. The cycles-to-failure distribution of the identical LCR assemblies subjected to a different loading level is predicted accurately by the calibrated model, which corroborates the validity of the proposed approach.

- An advanced metamodeling scheme, called partitioned bivariate Cut-high dimensional model representation (PB Cut-HDMR), is developed to fulfill unique feature of the semiconductor package products, i.e., a large number of input variables and both statistical correlations among input variables and some of two-variable interaction effects are significant. The statistical correlation is handled by eigen-decomposition of a covariance matrix. The latter is achieved by the HDMR-factorial design (HDMR-FD) hybrid method. The validity of the proposed scheme is verified by comparing the performance of the proposed scheme with the full bivariate Cut-HDMR and the successful implementation to construct an accurate metamodel for a problem with 12 input variables among which 2 pairs are correlated. The proposed metamodel is expected to provide a useful tool for the design optimization and statistical model calibration for other semiconductor package products.

## 5.2. Future Works

As a nature extension of the advancing capabilities of numerical modeling, there will be an increased need to incorporate the statistical analysis into the modeling for semiconductor package development. The contributions by this dissertation can be extended in many directions. Some of them are described below.

**Part I:**

Extension of capability of EDR to handle nonlinear correlations

The EDR is developed for the cases with linear correlations among input variables. Most of the input variables for semiconductor package products fall in this regime. Extending the

capability of EDR to handle the nonlinear correlations, however, can further increase the application domain of EDR.

Extension of PoP yield loss prediction model with consideration of solder bridging

For PoP assembly, in addition to the yield loss caused by non-contact open. The other critical failure mode is solder joint bridging, which causes short circuits. By extending the yield loss prediction model to include the failures due to solder bridging, a more comprehensive design tool can be provided.

Development of sampling strategy considering linearity/nonlinearity of response functions

The current sampling scheme of EDR was determined based on the finding that $2N+1$ for linear performance response and $4N+1$ for nonlinear performance response can provide sufficient accuracy in numerous numerical examples. Quantitative evaluation of the linearity/nonlinearity of the performance response and establishment the sampling strategies for different degree of linearity/nonlinearity can provide a more systematic guideline of selecting the sample points.

**Part II:**

Quantitative measure for the result of statistical model calibration

The current calibration result was compared with another set of data using graphical comparison, which provides a rough measure of the performance of the calibration result. From the perspective of semiconductor package development, it is desired to quantitatively describe the error associated with the calibrated model, so that, a more precise judgement can be made before applying the calibrated model to other loading conditions. Challenges associate with this approach consist of quantifying the effects of uncertainties involved in the statistical calibration process,

such as the variation from the calibrated data, and then representing these effects as the variation of calibration result. Since the calibration result will be a statistical distribution but contains uncertainty, the UP analysis of handle the input variables with uncertainty would also need to be taken into consideration.

<u>Extension of model calibration capability to more than two unknown input variables</u>

During the calibration, the PDF of performance response has to be constructed several times per iteration for the evaluation of objective function and the derivative of the objective function. The number of PDF constructions increases as the number of unknown input variables grows, where each of the PDF construction will need $2N+1$ or $4N+1$ modeling runs. Therefore, even EDR can become impractical when the number of unknown input variables is more than two. A solution is needed to extend the capability of statistical model calibration, which can be done by the incorporating the proposed PB Cut-HDMR into the calibration procedure.

**Part III:**

<u>Extension of PB Cut-HDMR to reduce the number of modeling runs in ranking and partitioning bivariate terms</u>

In this dissertation, the PB Cut-HDMR method relies ranking and partitioning the bivariate terms. Although the proposed HDMR-FD hybrid method uses only one additional run per bivariate pair, the computational cost is equal to $N(N-1)/2$, which can prohibit PB Cut-HDMR when the number of input variables getting very large. This situation could happened when a system level reliability to be considered, for example, the board level solder joint reliability of PoP. A ranking and partitioning method with even better efficiency should be investigated.

Extension of sampling strategy of PB Cut-HDMR to consider linear/nonlinearity of the component functions

The proposed PB Cut-HDMR uses five sample points along each input variable based on the study done in the literature. In reality, the component functions along each variable or on the plane of each pair of input variables can have different degree of linear/nonlinearity. Five sample points for a linear component functions may not be necessary, while they could be insufficient for highly nonlinear component functions. Consideration of linear/nonlinearity of component functions can avoid the waste sample points and/or help to increase accuracy. Challenges of achieving this extension include the lack of proper quantitative measure of "linear/nonlinearity of functions" and the method of checking the nonlinearity with minimum additional computational cost.

## REFERENCE


[1] Y. Sun, H.-S. Lee, and B. Han, "Measurement of the Comprehensive Viscoelastic Properties of Advanced EMC Using FBG Sensor," in *Electronic Components and Technology Conference (ECTC), 2016 IEEE 66th*, 2016, pp. 531-537.

[2] K. Ishibashi, "PoP (package-on-package) stacking yield loss study," in *Electronic Components and Technology Conference, 2007. ECTC'07. Proceedings. 57th*, 2007, pp. 1403-1408.

[3] D. Xie, D. Geiger, D. Shangguan, B. Hu, and J. Sjoberg, "Yield Study of Inline Package on Package (PoP) Assembly," in *Electronics Packaging Technology Conference, 2008. EPTC 2008. 10th*, 2008, pp. 1202-1208.

[4] D. Xie, D. Shangguan, D. Geiger, D. Gill, V. Vellppan, and K. Chinniah, "Head in pillow (HIP) and yield study on SIP and PoP assembly," in *Electronic Components and Technology Conference, 2009. ECTC 2009. 59th*, 2009, pp. 752-758.

[5] T. R. Bieler, H. Jiang, L. P. Lehman, T. Kirkpatrick, E. J. Cotts, and B. Nandagopal, "Influence of Sn grain size and orientation on the thermomechanical response and reliability of Pb-free solder joints," *IEEE Transactions on Components and Packaging Technologies,* vol. 31, pp. 370-381, 2008.

[6] S. Mukherjee, B. Zhou, A. Dasgupta, and T. R. Bieler, "Multiscale modeling of the anisotropic transient creep response of heterogeneous single crystal SnAgCu solder," *International Journal of Plasticity,* vol. 78, pp. 1-25, 2016.

[7] H.-P. Wei, B. Han, B. D. Youn, H. Shin, I. Kim, and H. Moon, "Assembly yield prediction of plastically encapsulated packages with a large number of manufacturing variables by advanced approximate integration method," *Microelectronics Reliability,* vol. 78, pp. 319-330, 2017.

[8] T. Katahira, J. Scanlan, J. Park, and K. Oh, "0.3 mm pitch CSP/BGA development for mobile terminals," in *40th International Symposium on Microelectronics (IMAPS 2007) Proceedings*, 2007, pp. 393-401.

[9] K. Yang and B. S. El-Haik, *Design for six sigma*: McGraw-Hill New York, 2003.

[10] D. Kececioglu, *Robust Engineering Design-by-reliability with Emphasis on Mechanical Components & Structural Reliability* vol. 1: DEStech Publications, Inc, 2003.

[11] D. Vose, *Risk analysis: a quantitative guide*: John Wiley & Sons, 2008.

[12] D.-G. Yang, J. Liang, Q.-Y. Li, L. J. Ernst, and G. Zhang, "Parametric study on flip chip package with lead-free solder joints by using the probabilistic designing approach," *Microelectronics Reliability,* vol. 44, pp. 1947-1955, 2004.

[13] R. H. Myers, D. C. Montgomery, and C. M. Anderson-Cook, *Response surface methodology: process and product optimization using designed experiments*: John Wiley & Sons, 2016.

[14] P. Rajaguru, H. Lu, and C. Bailey, "Sintered silver finite element modelling and reliability based design optimisation in power electronic module," *Microelectronics Reliability,* vol. 55, pp. 919-930, 2015.

[15] B. Öztürk, P. Gromala, C. Silber, K. Jansen, and L. Ernst, "Finite element based design of a new dogbone specimen for low cycle fatigue testing of highly filled epoxy-based adhesives for automotive applications," in *Thermal, Mechanical and Multi-Physics Simulation and Experiments in Microelectronics and Microsystems (EuroSimE), 2013 14th International Conference on*, 2013, pp. 1-4.

[16] J. W. Evans, J. Y. Evans, and B. K. Yu, "Designing and building-in reliability in advanced microelectronic assemblies and structures," *IEEE Transactions on Components, Packaging, and Manufacturing Technology: Part A,* vol. 20, pp. 38-45, 1997.

[17] L. Marretta and R. Di Lorenzo, "Influence of material properties variability on springback and thinning in sheet stamping processes: a stochastic analysis," *The International Journal of Advanced Manufacturing Technology,* vol. 51, pp. 117-134, 2010.

[18] S. Stoyanov, C. Bailey, N. Strusevich, and J.-M. Yannou, "Computational approach for reliable and robust system-in-package design," in *2007 30th International Spring Seminar on Electronics Technology (ISSE)*, 2007, pp. 40-45.

[19] S. Reh, J.-D. Beley, S. Mukherjee, and E. H. Khor, "Probabilistic finite element analysis using ANSYS," *Structural Safety,* vol. 28, pp. 17-43, 2006.

[20] H. Seo and B. Kwak, "An improved reliability analysis using design of experiments and an application to tolerance design," in *The 5th world congress of structural and multidisciplinary optimization, Lodo do Jesolo, Italy, May*, 2003, pp. 19-23.

[21] S. Rahman and H. Xu, "A univariate dimension-reduction method for multi-dimensional integration in stochastic mechanics," *Probabilistic Engineering Mechanics,* vol. 19, pp. 393-408, 2004.

[22] B. D. Youn, Z. Xi, and P. Wang, "Eigenvector dimension reduction (EDR) method for sensitivity-free probability analysis," *Structural and Multidisciplinary Optimization,* vol. 37, pp. 13-28, 2008.

[23] J. Zhao, Y. Luo, Z. Huang, and R. Ma, "Effects of package design on top PoP package warpage," in *2008 58th Electronic Components and Technology Conference*, 2008, pp. 1082-1088.

[24] D. Xie, D. Shangguan, D. Geiger, D. Gill, V. Vellppan, and K. Chinniah, "Head in pillow (HIP) and yield study on SIP and PoP assembly," in *2009 59th Electronic Components and Technology Conference*, 2009, pp. 752-758.

[25] W. Lin and M. W. Lee, "PoP/CSP warpage evaluation and viscoelastic modeling," in *2008 58th Electronic Components and Technology Conference*, 2008, pp. 1576-1581.

[26] JEDEC, "JEP 95 SPP-024A standard," *Reflow Flatness Requirements for Ball Grid Array Packages,* 2009.

[27] JEDEC, "Coplanarity Test for Surface-Mount Semiconductor Devices," *JESD22-B108A,* 2003.

[28] IPC/JEDEC J-STD-020C, "Moisture/Reflow Sensitivity Classification for Nonhermetic Solid State Surface Mount Devices," JEDEC Solid State Technology Association2004.

[29] C. C. Zhang and H.-P. B. Wang, "Robust design of assembly and machining tolerance allocations," *IIE transactions,* vol. 30, pp. 17-29, 1997.

[30] Micron Technology Inc., "16Gb: 216-Ball, Dual-Channel Mobile LPDDR3 SDRAM Features," *EDFA164A1PB, EDFA164A1PK Datasheet,* 2014.

[31] Amkor Technology Inc., "Package on Package (PoP) Family," *DS586G Datasheet,* 2012.

[32] N. L. Johnson, S. Kotz, and N. Balakrishnan, "Continuous univariate distributions , vol. 1John Wiley & Sons," *New York,* p. 163, 1994.

[33] R. V. Hogg, E. Tanis, and D. Zimmerman, *Probability and statistical inference*: Pearson Higher Ed, 2014.

[34] J. M. Ferrin, M. Bishop, T. N. Tansey, M. Frain, E. A. Swett, and F. J. Lane, "Conceptual and practical implications for rehabilitation research: Effect size estimates, confidence intervals, and power," *Rehabilitation Education,* vol. 21, pp. 87-100, 2007.

[35] M. Holický, *Reliability analysis for structural design*: AFRICAN SUN MeDIA, 2009.

[36] JEDEC Publication 95 (JEP95), "Design Guide 4.5, Fine Pitch Square Ball Grid Array (FBGA) Package " January 2009.
[37] JEDEC Pubilication 95 (JEP95), "Design Registration 4.14, Ball Grid Array Package," April 2011.
[38] D. Roos, U. Adam, and V. Bayer, "Design reliability analysis," *Proceedings Weimarer Optimierungs-und Stochastiktage,* vol. 3, 2006.
[39] W. K. Loh, R. Kulterman, T. Purdie, H. Fu, and M. Tsuriya, "Recent trend of package warpage characteristic," in *Electronics Packaging and iMAPS All Asia Conference (ICEP-IACC), 2015 International Conference on*, 2015, pp. 233-238.
[40] N. Vijayaragavan, F. Carson, and A. Mistry, "Package on Package warpage-impact on surface mount yields and board level reliability," in *Electronic Components and Technology Conference, 2008. ECTC 2008. 58th*, 2008, pp. 389-396.
[41] C.-Y. Huang, "Package-on-package assembly yield assessment in the ODM/EMS environment using monte carlo simulation," *IEEE Transactions on Components, Packaging and Manufacturing Technology,* vol. 3, pp. 1611-1620, 2013.
[42] C.-Y. Huang, "Applying Monte Carlo Simulation to Analyze the Open Scenario in the Assembly of Electronic Components," *IEEE Transactions on Components, Packaging and Manufacturing Technology,* vol. 7, pp. 1911-1919, 2017.
[43] C. Sundararajan and F. Witt, "Stress-strength interference method," *Probabilistic Structural Mechanics Handbook: Theory and Industrial Applications,* pp. 8-26, 1995.
[44] A. P. Basu, "Stress–Strength Model," *Encyclopedia of Statistics in Quality and Reliability,* 2007.
[45] V. V. Petrov, *Sums of independent random variables* vol. 82: Springer Science & Business Media, 2012.
[46] P. S. Puri, "Probability generating functions of absolute difference of two random variables," *Proceedings of the National Academy of Sciences of the United States of America,* pp. 1059-1061, 1966.
[47] M. D. Springer, "The algebra of random variables," 1979.
[48] D. R. Cox and N. Wermuth, *Multivariate Dependencies- Models, analysis and interpretation*: Chapman & Hall, 1996.
[49] P. Kumar, "Probability distributions and estimation of Ali-Mikhail-Haq copula," *Applied Mathematical Sciences,* vol. 4, 2010.
[50] C. R. Taylor, "Two practical procedures for estimating multivariate nonnormal probability density functions," *American Journal of Agricultural Economics,* vol. 72, 1990.
[51] S. Shan and G. G. Wang, "Survey of modeling and optimization strategies to solve high-dimensional design problems with computationally-expensive black-box functions," *Structural and Multidisciplinary Optimization,* vol. 41, 2010.
[52] K. Bogart and C. Stein, "Discrete Math in Computer Science," ed: Dept. Of Computer Mathematics and Dept. Of Computer Science. Dartmouth College, 2002.
[53] S. G. Chow, W. K. Choi, R. Emigh, and E. Ouyang, "Board level solder joint reliability modeling of embedded wafer level BGA (eWLB) packages under temperature cycling test conditions," in *Electronics Packaging Technology Conference (EPTC), 2011 IEEE 13th*, 2011, pp. 674-680.
[54] J. Kwak, "Strain behaviors of solder bump with underfill for flip chip package under thermal loading condition," *Journal of Mechanical Science and Technology,* vol. 28, pp. 4899-4906, 2014.

[55] S. Park, H. Lee, B. Sammakia, and K. Raghunathan, "Predictive model for optimized design parameters in flip-chip packages and assemblies," *IEEE Transactions on Components and Packaging Technologies,* vol. 30, pp. 294-301, 2007.
[56] J. JEDEC, "SPP-024A standard," *Reflow Flatness Requirements for Ball Grid Array Packages,* 2009.
[57] F. Carson, S. M. Lee, and N. Vijayaragavan, "Controlling Top Package Warpage for POP Applications," in *Electronic Components and Technology Conference, 2007. ECTC'07. Proceedings. 57th*, 2007, pp. 737-742.
[58] M. J. Yim, R. Strode, R. Adimula, and C. Yoo, "Effects of material properties on PoP top package warpage behaviors," in *Electronic Components and Technology Conference (ECTC), 2010 Proceedings 60th*, 2010, pp. 1071-1076.
[59] M. Miyatake, H. Murai, S. Takanezawa, S. Tsuchikawa, M. Takekoshi, T. Kotake, and M. Ose, "Newly developed ultra low CTE materials for thin core PKG," in *Electronic Components and Technology Conference (ECTC), 2012 IEEE 62nd*, 2012, pp. 1588-1592.
[60] M. T. Inc., "16Gb: 216-Ball, Dual-Channel Mobile LPDDR3 SDRAM Features," *EDFA164A1PB, EDFA164A1PK Datasheet,* 2014.
[61] I. Corporation, "Ch 14, An Introduction to Plastic Ball Grid Array (PBGA) Packaging," *Packaging Databook [online],* pp. 1-16, 2000.
[62] T.-Y. Wen and S.-C. Ku, "Validation of warpage for small form factor flip-chip BGA by experimental and numerical methodology," in *Electronic Components and Technology Conference, 2008. ECTC 2008. 58th*, 2008, pp. 1587-1592.
[63] H. Mori, S. Kohara, K. Okamoto, H. Noma, and K. Toriyama, "Effects of low CTE materials on thermal deformation of organic substrates in flip chip package application," in *ASME 2015 International Technical Conference and Exhibition on Packaging and Integration of Electronic and Photonic Microsystems collocated with the ASME 2015 13th International Conference on Nanochannels, Microchannels, and Minichannels*, 2015, pp. V002T01A016-V002T01A016.
[64] H.-P. Wei and B. Han, "Stacking Yield Prediction of Package-on-Package Assembly Using Advanced Uncertainty Propagation Analysis: Part I Stochastic Model Development," 2018.
[65] K. A. Brakke, "The surface evolver," *Experimental mathematics,* vol. 1, pp. 141-165, 1992.
[66] K.-N. Chiang and C.-A. Yuan, "An overview of solder bump shape prediction algorithms with validations," *IEEE transactions on advanced packaging,* vol. 24, pp. 158-162, 2001.
[67] M. Dreiza, A. Yoshida, K. Ishibashi, and T. Maeda, "High density PoP (package-on-package) and package stacking development," in *Electronic Components and Technology Conference, 2007. ECTC'07. Proceedings. 57th*, 2007, pp. 1397-1402.
[68] P. Sun, V. Leung, D. Yang, R. Lou, D. Shi, and T. Chung, "Development of a new package-on-package (PoP) structure for next-generation portable electronics," in *Electronic Components and Technology Conference (ECTC), 2010 Proceedings 60th*, 2010, pp. 1957-1963.
[69] W. C. Wang, F. Lee, G. Weng, W. Tai, M. Ju, R. Chuang, and W. Fang, "Platform of 3D Package Integration," in *Electronic Components and Technology Conference, 2007. ECTC'07. Proceedings. 57th*, 2007, pp. 743-747.
[70] B. Roggeman, D. Vicari, L. Smith, and A. Syed, "Material selection and parameter optimization for reliable TMV PoP assembly," *Global SMT and Packaging,* vol. 12, p. 48, 2012.

[71] N. Chen, K. Chiang, Y. Wang, and C. Hsiao, "Controlled S-Parameters on Multi-Layer Package," in *ASME 2003 International Electronic Packaging Technical Conference and Exhibition*, 2003, pp. 11-15.

[72] F. N. Chowdhury, Z. S. Kolber, and M. D. Barkley, "Monte Carlo convolution method for simulation and analysis of fluorescence decay data," *Review of scientific instruments,* vol. 62, pp. 47-52, 1991.

[73] C. Baker and M. Derakhshan, "FFT techniques in the numerical solution of convolution equations," *Journal of Computational and Applied Mathematics,* vol. 20, pp. 5-24, 1987.

[74] D. Mukherjee, B. Rao, and A. Prasad, "Cut-HDMR-based fully equivalent operational model for analysis of unreinforced masonry structures," *Sadhana,* vol. 37, pp. 609-628, 2012.

[75] T.-K. Lee, B. Zhou, L. Blair, K.-C. Liu, and T. R. Bieler, "Sn-Ag-Cu solder joint microstructure and orientation evolution as a function of position and thermal cycles in ball grid arrays using orientation imaging microscopy," *Journal of electronic materials,* vol. 39, pp. 2588-2597, 2010.

[76] T. R. Bieler, B. Zhou, L. Blair, A. Zamiri, P. Darbandi, F. Pourboghrat, T.-K. Lee, and K.-C. Liu, "The role of elastic and plastic anisotropy of Sn in recrystallization and damage evolution during thermal cycling in SAC305 solder joints," *Journal of Electronic Materials,* vol. 41, pp. 283-301, 2012.

[77] H. Chen, B. Yan, M. Yang, X. Ma, and M. Li, "Effect of grain orientation on mechanical properties and thermomechanical response of Sn-based solder interconnects," *Materials Characterization,* vol. 85, pp. 64-72, 2013.

[78] J. J. Sundelin, S. T. Nurmi, and T. K. Lepistö, "Recrystallization behaviour of SnAgCu solder joints," *Materials Science and Engineering: A,* vol. 474, pp. 201-207, 2008.

[79] D. A. Shnawah, M. F. M. Sabri, and I. A. Badruddin, "A review on thermal cycling and drop impact reliability of SAC solder joint in portable electronic products," *Microelectronics reliability,* vol. 52, pp. 90-99, 2012.

[80] A. Micol, C. Martin, O. Dalverny, M. Mermet-Guyennet, and M. Karama, "Reliability of lead-free solder in power module with stochastic uncertainty," *Microelectronics Reliability,* vol. 49, pp. 631-641, 2009.

[81] Y. Aoues, A. Makhloufi, P. Pougnet, and A. El-Hami, "Robustness study of solder joints of different compositions by using Stochastic Finite Element Modeling," in *Integrated Power Systems (CIPS), 2014 8th International Conference on*, 2014, pp. 1-7.

[82] Y. Hsu, C.-Y. Su, and W.-F. Wu, "Quantitative reliability analysis of flip–chip packages under thermal–cyclic loading and in consideration of parameter uncertainties," *Microelectronics reliability,* vol. 51, pp. 2284-2289, 2011.

[83] L. E. Schwer, "Guide for verification and validation in computational solid mechanics," 2009.

[84] B. C. Jung, H. Yoon, H. Oh, G. Lee, M. Yoo, B. D. Youn, and Y. C. Huh, "Hierarchical model calibration for designing piezoelectric energy harvester in the presence of variability in material properties and geometry," *Structural and Multidisciplinary Optimization,* vol. 53, pp. 161-173, 2016.

[85] D. Yu, A. Al-Yafawi, T. T. Nguyen, S. Park, and S. Chung, "High-cycle fatigue life prediction for Pb-free BGA under random vibration loading," *Microelectronics Reliability,* vol. 51, pp. 649-656, 2011.

[86] C. E. Jih, "Thermal reliability of surface mount leadless solder joints," *International Journal of Materials and Product Technology,* vol. 16, pp. 370-378, 2001.

[87] Y. Corporation, "RC 2512 (RoHS Compliant) " *DATA SHEET GENERAL PURPOSE CHIP RESISTORS,* 2009.
[88] M. Hasnine, J. Suhling, B. Prorok, M. Bozack, and P. Lall, "Anisotropic Mechanical Properties of SAC Solder Joints in Microelectronic Packaging and Prediction of Uniaxial Creep Using Nanoindentation Creep," *Experimental Mechanics,* vol. 57, pp. 603-614, 2017.
[89] J. S. S. T. Association, "JESD22B113 Board level cyclic bend test method for interconnect reliability characterization of components for handled electronic products," ed: Virginia, USA:[sn], 2006.
[90] T. Lee, J. Lee, and I. Jung, "Finite element analysis for solder ball failures in chip scale package," in *Physical & Failure Analysis of Integrated Circuits, 1997., Proceedings of the 1997 6th International Symposium on*, 1997, pp. 39-43.
[91] G. Wang, C. Merrill, J.-H. Zhao, S. K. Groothuis, and P. S. Ho, "Packaging effects on reliability of Cu/low-k interconnects," *IEEE Transactions on Device and Materials Reliability,* vol. 3, pp. 119-128, 2003.
[92] Y. L. Tzeng, N. Kao, E. Chen, J. Y. Lai, Y. P. Wang, and C. Hsiao, "Warpage and stress characteristic analyses on package-on-package (pop) structure," in *Electronics Packaging Technology Conference, 2007. EPTC 2007. 9th*, 2007, pp. 482-487.
[93] S.-S. Yeh, P.-Y. Lin, S.-P. Jeng, W.-Y. Lin, M.-C. Yew, K.-C. Lee, S.-T. Leu, and K.-C. Liu, "A novel integrated warpage prediction model based on characterization of viscoelasticity in time domain and chemical shrinkage for molded underfill," in *Microsystems, Packaging, Assembly and Circuits Technology Conference (IMPACT), 2015 10th International*, 2015, pp. 79-82.
[94] A. Dasgupta, M. G. Pecht, and B. Mathieu, "Design-of-experiment methods for computational parametric studies in electronic packaging," *Finite Elements in Analysis and Design,* vol. 30, pp. 125-146, 1998.
[95] M. Yew, C. Yuan, C. Han, C. Huang, W. Yang, and K. Chiang, "Factorial analysis of chip-on-metal WLCSP technology with fan-out capability," in *Physical and Failure Analysis of Integrated Circuits, 2006. 13th International Symposium on the*, 2006, pp. 223-228.
[96] S. Rzepka, A. Müller, and B. Michel, "Virtual prototyping advanced by statistic and stochastic methodologies," in *Thermal, Mechanical & Multi-Physics Simulation, and Experiments in Microelectronics and Microsystems (EuroSimE), 2010 11th International Conference on*, 2010, pp. 1-9.
[97] A. Sasi, A. Yadur, and P. Gromala, "Simulation driven design of novel integrated circuits-Part 1: Selection of the materials based on the Virtual DoE," in *Thermal, Mechanical and Multi-Physics Simulation and Experiments in Microelectronics and Microsystems (EuroSimE), 2015 16th International Conference on*, 2015, pp. 1-7.
[98] T. W. Simpson, D. K. Lin, and W. Chen, "Sampling strategies for computer experiments: design and analysis," *International Journal of Reliability and Applications,* vol. 2, pp. 209-240, 2001.
[99] T. W. Simpson, J. Poplinski, P. N. Koch, and J. K. Allen, "Metamodels for computer-based engineering design: survey and recommendations," *Engineering with computers,* vol. 17, pp. 129-150, 2001.
[100] R. Jin, W. Chen, and T. W. Simpson, "Comparative studies of metamodelling techniques under multiple modelling criteria," *Structural and multidisciplinary optimization,* vol. 23, pp. 1-13, 2001.

[101] S. Shan and G. G. Wang, "Survey of modeling and optimization strategies to solve high-dimensional design problems with computationally-expensive black-box functions," *Structural and Multidisciplinary Optimization,* vol. 41, pp. 219-241, 2010.

[102] S. E. Gano, H. Kim, and D. E. Brown, "Comparison of three surrogate modeling techniques: Datascape, kriging, and second order regression," in *Proceedings of the 11th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference, AIAA-2006-7048, Portsmouth, Virginia*, 2006.

[103] S. W. Doebling and F. M. Hemez, "A metamodel-based approach to model validation for nonlinear finite element simulations," Los Alamos National Lab., NM (US)2001.

[104] H. Rabitz and Ö. F. Aliş, "General foundations of high-dimensional model representations," *Journal of Mathematical Chemistry,* vol. 25, pp. 197-233, 1999.

[105] G. Li, C. Rosenthal, and H. Rabitz, "High dimensional model representations," *The Journal of Physical Chemistry A,* vol. 105, pp. 7765-7777, 2001.

[106] Ö. F. Alış and H. Rabitz, "Efficient implementation of high dimensional model representations," *Journal of Mathematical Chemistry,* vol. 29, pp. 127-142, 2001.

[107] G. Li, S.-W. Wang, C. Rosenthal, and H. Rabitz, "High dimensional model representations generated from low dimensional data samples. I. mp-Cut-HDMR," *Journal of Mathematical Chemistry,* vol. 30, pp. 1-30, 2001.

[108] G. Li, M. Artamonov, H. Rabitz, S. w. Wang, P. G. Georgopoulos, and M. Demiralp, "High-dimensional model representations generated from low order terms—lp-RS-HDMR," *Journal of computational chemistry,* vol. 24, pp. 647-656, 2003.

[109] S. Shan and G. G. Wang, "Metamodeling for high dimensional simulation-based design problems," *Journal of Mechanical Design,* vol. 132, p. 051009, 2010.

[110] B. Vandevelde, I. Konstantinou, D. Moens, and D. Vandepitte, "Dealing with IC package material and design uncertainties using FUZZY finite elements," presented at the 14th International Conference on Thermal, Mechanical and Multi-Physics Simulation and Experiments in Microelectronics and Microsystems (EuroSimE), Wroclaw, 2013.

[111] I. M. Sobol, "Theorems and examples on high dimensional model representation," *Reliability Engineering & System Safety,* vol. 79, pp. 187-193, 2003.

[112] H. Xu and S. Rahman, "Decomposition methods for structural reliability analysis," *Probabilistic Engineering Mechanics,* vol. 20, pp. 239-250, 2005.

[113] D. C. Woods and S. M. Lewis, "Design of experiments for screening," *Handbook of Uncertainty Quantification,* pp. 1143-1185, 2017.

[114] C. J. Wu and M. S. Hamada, *Experiments: planning, analysis, and optimization* vol. 552: John Wiley & Sons, 2011.

[115] X. Lu, L. Chen, Z. Cheng, and J. Liu, "Reliability analysis of embedded chip technique with design of experiment methods," in *Electronics Materials and Packaging, 2005. EMAP 2005. International Symposium on*, 2005, pp. 43-49.

[116] C.-C. Lee, S.-M. Chang, and K.-N. Chiang, "Sensitivity design of DL-WLCSP using DOE with factorial analysis technology," *IEEE transactions on advanced packaging,* vol. 30, pp. 44-55, 2007.

[117] W. Sun, W. Zhu, C. Wang, A. Y. Sun, and H. Tan, "Warpage simulation and DOE analysis with application in package-on-package development," in *Thermal, Mechanical and Multi-Physics Simulation and Experiments in Microelectronics and Micro-Systems, 2008. EuroSimE 2008. International Conference on*, 2008, pp. 1-8.

[118] J. L. Devore, *Probability and Statistics for Engineering and the Sciences*: Cengage learning, 2011.

[119] W. Härdle and L. Simar, *Applied multivariate statistical analysis* vol. 22007: Springer, 2007.
[120] H. Abdi and L. J. Williams, "Principal component analysis," *Wiley interdisciplinary reviews: computational statistics,* vol. 2, pp. 433-459, 2010.
[121] S. C. Cotter, "A screening design for factorial experiments with interactions," *Biometrika,* vol. 66, pp. 317-320, 1979.
[122] D. C. Montgomery, *Design and analysis of experiments*: John Wiley & Sons, 2017.
[123] T. A. Mara and S. Tarantola, "Application of global sensitivity analysis of model output to building thermal simulations," in *Building Simulation*, 2008, pp. 290-302.
[124] S. Shan and G. G. Wang, "Turning black-box functions into white functions," *Journal of Mechanical Design,* vol. 133, p. 031003, 2011.