OPTIMISATION MODELLING FOR MICROELECTRONICS PACKAGING AND PRODUCT DESIGN

Stoyan Kostadinov Stoyanov

A thesis submitted in partial fulfilment of the requirements of the University of Greenwich for the Degree of

Doctor of Philosophy

Centre for Numerical Modelling and Process Analysis
School of Computing and Mathematical Sciences
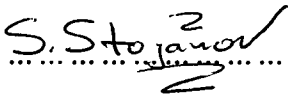The University of Greenwich, London, UK

March 2004

Dedications

For my mother,
Zdravka Mircheva

and my father,
Kostadin Stoyanov Mirchev,

the most extraordinary blessing in my life...

With all my love and heart, always,
Stoyan

DECLARATION

I certify that this work has not been accepted in substance for any degree, and is not concurrently submitted for any degree other than that of Doctor of Philosophy (Ph.D.) of the University of Greenwich. I also declare that this work is the result of my own investigations except where otherwise stated.
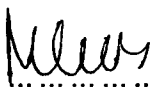
X..

Stoyan Stoyanov

Supervisors:

Prof. Christopher Bailey

Prof. Mark Cross

X..

Prof. Mark Cross

ABSTRACT

OPTIMISATION MODELLING FOR MICROELECTRONICS PACKAGING AND PRODUCT DESIGN

The objective of this research is to develop a design framework for virtual prototyping of electronic packaging. This framework couples computational mechanics and fluid dynamics, based on finite volume method with integrated finite element routines, with numerical optimisation and statistical methods. This integrated approach is intended as a modelling tool for calculating optimal design solutions for electronic packaging and component assembly with a focus on the reliability and the thermal management. The motivation is to introduce numerical optimisation theory as an approach for a fast, systematic and automated design approach for wide range microelectronics applications. The proposed methodology will also benefit from multi-physics numerical analysis to predict complex behaviour of electronic packages, systems and processes subject to different operational or environmental conditions.

This thesis demonstrates multi-physics modelling (i.e. integrated solutions for fluid flow, heat transfer and stress) coupled with gradient/non-gradient based numerical optimisation techniques and associated statistical methods. An explanation and comparison of the two approaches to numerical optimisation – (1) Response Surface Methodology (RSM) based on Design of Experiments (DoE) and (2) direct gradient based and non-gradient methods are given. Both the advantages and limitations of these virtual design strategies, with respect to their integration with multi-physics modelling, are discussed and demonstrated.

This integrated **multiphysics/optimisation** design approach is demonstrated on a variety of problems from the area of microelectronics design and packaging. The thesis demonstrates this for three industrial examples. These are:

1. **Solder Joint Fatigue.**

   This example focuses on the thermo-mechanical reliability of an electronic package assembled using the flip-chip process. A number of important process/design parameters are optimised so as to maximise the lifetime of the solder joints. Finite element simulations are undertaken to predict the stress and strain in the solder due to an imposed temperature cycle.

   This example is a single physics problem and the analysis only stress based. Suitable creep relationships and damage models for solder fatigue are used. The example demonstrates both direct gradient based and response surface optimisation. Important design parameters of a flip-chip package are detected, investigated and efficiently improved to satisfy the component requirements for performance and solder joint reliability. The optimisation of the flip-chip design is performed using direct gradient optimisation and Design of Experiment/Response Surface modelling.

2. **Wave Soldering.**

   This example demonstrates the flow of nitrogen in a wave soldering machine. This is a coupled fluid flow-heat transfer problem. Lead-free solder requires the use of nitrogen in this machine to ensure that no oxides form in the solder bath. In this industrial application, it is important to understand the flow characteristics of the nitrogen and to design the machine and the soldering process such that a minimum amount of nitrogen is used. Finite volume simulations are utilized.

   Both the Incrementally Refined Response Surface (IRRS) modelling and non-gradient optimisation approaches are demonstrated in this example.

3. **Thermal Management.**

   This example demonstrates numerical optimisation integrated with finite element and finite volume coupled solution procedures for fluid-flow, temperature, and stress. A printed circuit board with a number of electronic components on its surface is subjected to air cooling. Each component has a power dissipation that raises its temperature. This rise in temperature will also create stress in the components due to the coefficient of thermal expansion (CTE) miss-match between each of the materials. The aim of passing air over the components is to extract as much heat from the components as possible.

   This example uses the integrated optimisation/multiphysics framework to optimise the placement of a component on the circuit board. This location on the printed circuit board is identified to minimize the silicon die stress, equivalently the junction temperature, of the component.

The software packages used to develop the design tool and to undertake the outlined studies are *PHYSICA* and *VisualDOC*. *PHYSICA* is a multiphysics finite volume based simulation tool with integrated modules for finite element solid mechanics analysis. The software framework is detailed in Chapter 2, Section 2.4 and further in Chapter 4. The *VisualDOC* tool offers a collection of numerical optimisation routines and modules for statistical analysis (Design of Experiments) and approximate Response Surface modelling. *VisualDOC* framework is discussed in Chapter 4, Section 4.8.

The integrated *optimisation* and *multiphysics* modelling approach demonstrated in this research can benefit electronics design engineers in their efforts to reduce design time and the time-to-market for their products. Adopting *optimisation modelling* at the early stages of product development can provide optimal virtual prototypes that can be cheaply and efficiently obtained and will allow competitiveness and profitability of electronics manufacturers and assemblers.

ACKNOWLEDGMENTS

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

The microelectronics industry is one of the world's largest manufacturing sectors today and has an enormous impact on human life and all aspects of society. Electronic components are at the heart of many products and are found almost anywhere, ranging from networks, computers and mobile phones to automotive, aerospace and military devices. The fierce competition in the electronics world market and the ever increasing demands for faster, smaller and cheaper products impose real challenges to manufacturing companies. Some of the main trends that the electronics industry faces at present include:

- Developing new assembly technologies such as flip-chip process due to the demand for further miniaturisation of electronic packages (Figure 1.1);

- Improvements at interconnect level (cost, reliability, etc.);

Figure 1.1: Electronic Circuit

- Environment friendly manufacturing processes and products (e.g. lead free solders);

- Standards for high quality and reliable products (thermal, thermo-mechanical, electrical, etc.);

Throughout the life cycle of a product, the embedded electronic packages and systems are subject to a variety of environmental and operational loads. With the continuous miniaturisation of these products the temperatures and stresses within them become more extreme and impose a real danger for their performance and reliability (Figure 1.2). Knowledge on the failure mechanisms and critical parts of an assembly is essential to find and adopt optimal design solutions leading towards consistent and cost-effective manufacturing process of reliable electronic systems. The present demands in the microelectronic sector and the challenge to resolve effectively existing problems are the major motivation to undertake this research.
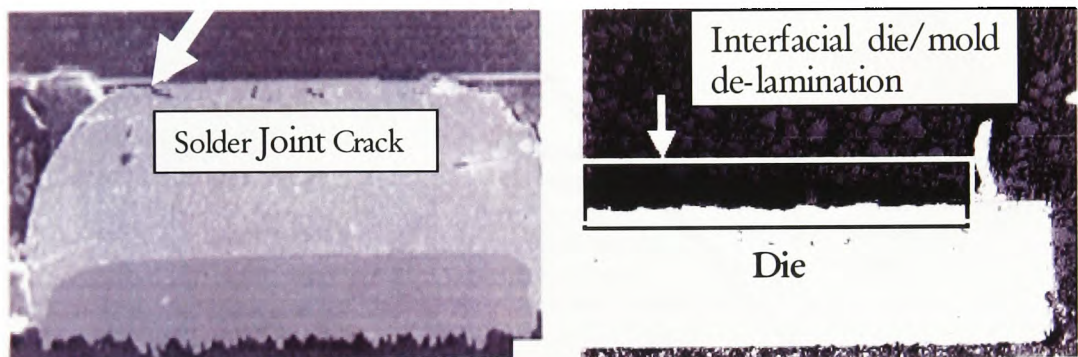
Figure 1.2: Common Failures in Electronic Assemblies

Although optimisation theory has been used in engineering design for a number of years, especially in structural analysis, it has very limited use in electronics packaging design, especially when coupled with multiphysics modelling. **This research focuses on the development of simulation-based optimisation processes as a route towards effective design of modern electronic products.** The methodology and tools that incorporate both computer-aided analysis and optimisation strategies into a robust technology for virtual design are discussed and implemented.

Having time to market of only few months for the modern electronics devices, it is critical for engineers to obtain the best design solutions quickly and efficiently. In the conflicting environment with demands for smaller, faster and cheaper electronic products restrained by increasing reliability and manufacturability issues, the proposed methodology (Figure 1.3) can aid the design process of new products, processes and technologies in microelectronics. In

the core of the design process methodology stays the principle of iterative improvement of the designed system by systematic changes in the system/process parameters realized using fully automated computational procedure. As seen in Figure 1.3 this approach uses extensively both the theory of optimisation through variety of numerically implemented optimisation algorithms and techniques and the power of reliable, physics-based finite element (FE) and finite volume (FV) analyses. It is a virtual design process ultimately leading towards a cost effective and fast way of finding optimal specification for any process or system related to electronics manufacturing.
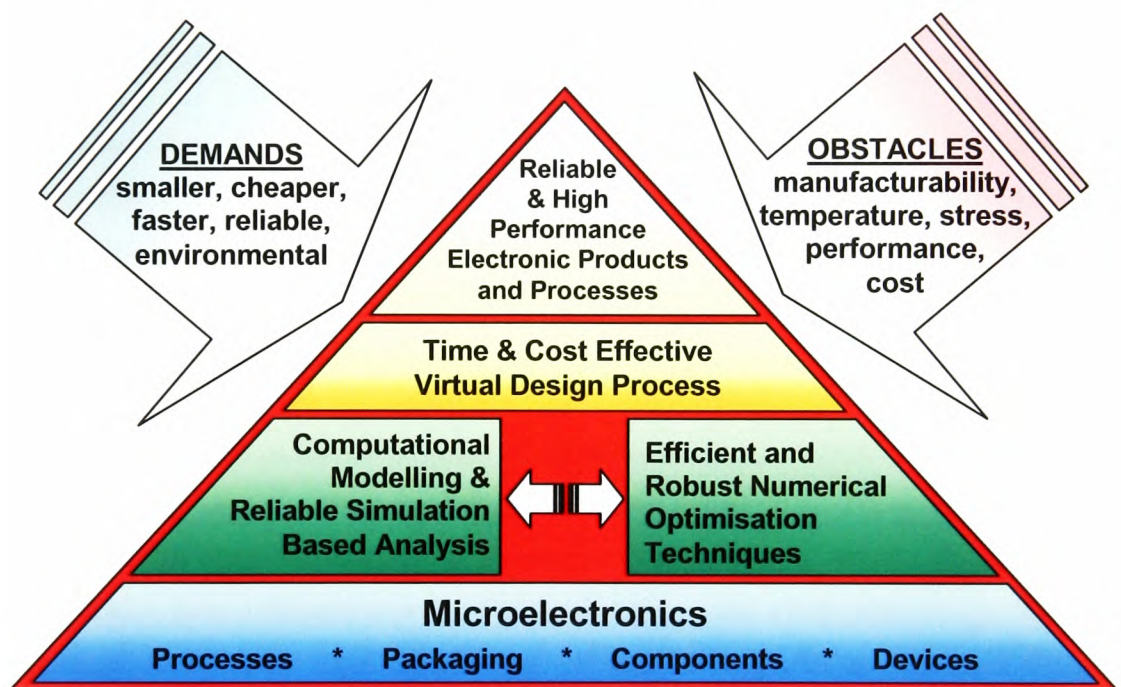
Figure 1.3: Methodology for Optimal Design Solutions in Microelectronics

Chapter 1 provides a brief overview of the following areas:

- **Optimisation theory** : history, achievements and contributions, applications in engineering;

    A brief history and review of the optimisation methods in mathematics involving constrained and unconstrained optimisation, continuous, discrete and integer

programming is presented in Section 1.1 together with some of the major contributors, authors and the corresponding references.

- **Computational modelling of electronic packages, systems and processes:** predictive analysis and optimisation prospective;

    Section 1.2 reviews computational modelling and some of the corresponding packaging issues being investigated in recent years. Research results in the area of improving microelectronics design, via computer simulations, are summarized. The aim here is to emphasize the need of a general design tool such as that adopted in this research.

- **Multiphysics finite element analysis**

    Since one of the important goals of this research is to couple optimisation methods to multiphysics Finite Element (FE) / Finite Volume (FV) analysis, a review of some of the research and published materials is outlined in Section 1.3.

1.1. Review of Optimisation Theory

Optimisation is one of the oldest fields of scientific investigation [202]. The first observed problems with aspects of optimality are recorded before B.C. However, it must be recognized that optimisation started to develop more systematically and as a separate field in mathematics, from a theoretical point of view, about three and a half centuries ago. The first significant contributions were made in the 17th and 18th centuries by some of the most famous mathematicians. While the nature and concepts of optimisation stay the same, the application areas have varied to include the important problems of the day [202].

1.1.1. *Origins of Optimisation*

One of the first recorded cases with observed optimality was considered in the 3-rd century B.C. by Archimedes. He conjectured that the geometric curve with fixed length that will enclose, together with a straight line, the largest possible area is the semicircle. In the first century A.D., Heron of Alexandria - a mathematician, important geometer and worker in mechanics, observed the problem of finding the shortest path between two points. Based on

the physical principle of light reflection of a flat surface, Heron stated the associated optimisation principle that *"light travels between any two points by following the shortest path between these points"*.

Many centuries later, in 1657, the French mathematician Pierre Fermat re-formulated that postulate to state correctly that *"light travels always by the path which incurs least time"* rather than least distance. Fermat introduced the first formal ideas, terminology and fundamental concepts in the field of optimisation, e.g. a method of maxima and minima (the highest and lowest values of a curve) and that at the maximum and minimum the tangent line is horizontal. Apart from the *"optimal"* postulates of natural phenomena, another class of problems that was considered at that time involved the determination of the function that minimises a particular functional.

The famous problem, known as *brachistochrone*, formulated and solved by John Bernoulli in 1696 is another instance of optimal study: what is the shape of the curve joining two given points in space such that a frictionless bead travelling on the curve under influence of gravity will cover the journey from the upper point to the other point in the least time? This problem was posed as a competition and was solved by de l'Hopital, Leibniz and Newton [202].

Problems involving optimality were in the base of the developed by Euler new discipline which in 1766 he called *calculus of variations*. At this time many mechanical laws were formulated for first time, some in terms of principles of optimality. Examples include the Maupertuis's and Euler's principle of *"least action"*, the principle of Gauss for *"least restraint"* and Lagrange's kinetic principle [202]. The example in Figure 1.4 describes the principle of potential energy minimisation at the point of equilibrium.
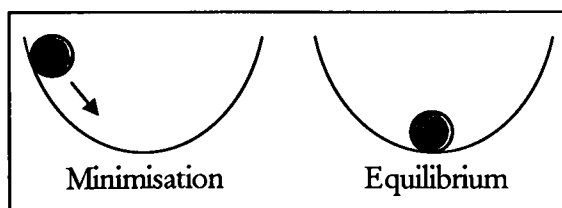
Figure 1.4: Potential Energy and
Mechanical Equilibrium
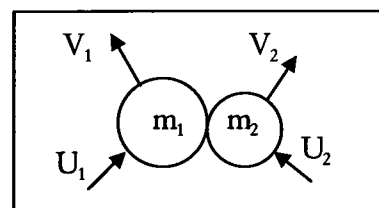
Figure 1.5: Least Action Principle –
Impact of Two Elastic Bodies

Figure 1.5 is an example of collision between two elastic balls with mass and velocity (m_1, U_1) and (m_2, U_2), respectively. From a physical point of view the speeds after impact, V_1 and V_2 will satisfy the momentum conservation law, $m_1U_1 + m_2U_2 = m_1V_1 + m_2V_2$. Considering the example from an optimisation perspective, the speeds V_1 and V_2 are such that to ensure the minimisation of a certain energy function.

The famous method of *Lagrange multipliers* that allows solving optimisation problems with linear constraints was invented by Lagrange in 1760. Another famous name with contributions to many fields, Gauss, developed *the least squares fitting curve* method. Later, in 1834 William R. Hamilton developed a set of functions known as Hamiltonians. These have been used at that time to state a principle of optimality related to the *optics* theory in mechanics. A principle of optimality related to the equilibrium of a thermo-dynamical system was introduced by J.W. Gibbs in 1875. Since then an increasing number of contributions to optimisation theory have been made regularly. More details about above very early contributions to optimisation can be found in reference [202].

1.1.2. *Modern Optimisation*

The mid 1940s and early 1950s were a turning point in the history of optimisation. In 1947 the father of modern optimisation, G. B. Dantzig [1], developed the Simplex method of optimisation, the most popular algorithm in the theory of optimisation. It came out of hs work as a Mathematical Advisor (expert on planning methods) to the U.S. Air Force and was known as "programming" - a military term that, at that time, referred to plans or schedules for training, logistical supply or deployment of men.

One of the outstanding modern contributions to optimisation theory in the area of nonlinear programming was made by Kuhn and Tucker (1951) [2]. They developed what is known as the *Kuhn-Tucker necessary conditions* for optimality.

Outstanding contributions to the rapid development of the modern optimisation theory were made by the mathematician R. E. Bellman (1957) [3]. He developed the *dynamic programming* techniques which are applicable to problems where optimum can be determined using decomposition of the original problem into sub-problems. In 1958 Pontryagin formulated *the maximum principle* [4]   a technique that also deals with

problems of the optimal control of dynamical systems. Both Bellman's and Pontryagin's contributions are closely related to the *calculus of variations*.

Hooke and Jeeves (1961) contributed to optimisation theory with the first non-gradient search method [5] called *pattern search*. The Powell's method (1964) [6] and its variations are among the most popular and representative methods of the class of *zero-order* methods [7].

R. E. Gomory introduced in 1958 [8] the *branch and bound* approach for solving integer programming problems. Important contributions to this method have been made 7 years later by Dakin and Balas [9].

The 1960's/early 70's mark a rapid development of the gradient methods. Steepest descent method, the classical conjugate gradient Fletcher-Reeves algorithm [11] and a variety of its modifications are the early optimisation search techniques that adopt the fundamental concepts of using optimisation function gradients to navigate the optimisation process. Further development of these algorithms is marked by the introduction of the variable metric algorithms such as the Davidon-Fletcher-Powell [12, 13] and Broydon-Fletcher-Goldfarb-Shanno (BFGS) [14-18] methods. The first described method for solving nonlinear optimisation problems with inequality constraints called *Method of Feasible Directions* [10] was developed by Zoutendijk (1960). Since then a huge number of gradient search techniques [24, 25] and second order methods such as the classical Newton's method [26] for constrained optimisation have been proposed and developed.

In recent years different classes of non-gradient based optimisation techniques (called *evolutionary*, or population-based methods) such as Genetic Algorithm (GA) [31], Particle Swarm Optimisation (PSO) [32], Differential Evolution (DE) [33] and many others have become attractive and also used in some design applications.

The list of the above references is just indicative and by no means covers all major developments in optimisation theory. It refers to some of the first important steps in the scientific branch of optimisation followed by years of development and applications in many different areas of optimal control and design.

1.1.3. *Optimisation in Engineering*

Optimisation techniques, in particular numerical optimisation methods, have evolved mainly in the operational research community. Optimisation can be applied in many areas, e.g. economics, finance, transport, ecology, etc. In this research we focus on *engineering*, in particular microelectronics design, as application area of optimisation theory. First contributions to popularization of the numerical optimisation and application of these techniques in the area of engineering design are made by L. A. Schmit [19, 20]. In 1960 he applied numerical optimisation techniques to structural design referring to the approach as *"structural synthesis"*. His work opens a fundamentally new approach in engineering design beyond structural optimisation, in fact applicable to wide range of areas and engineering problems.

At that time the first applications of linear programming theory in the design of structures are made [21-23]. These first efforts to apply optimisation in solving design problems establish the fundamentals of what is now known nowadays as *design optimisation*. The fast development of optimisation numerical techniques and the rapid improvement in their efficiency and robustness open the route for application of the optimisation approach in many analyses.

An important development in design optimisation was made by introducing explicit sensitivity analysis techniques. These methods aid gradient search optimisation through explicit calculations of design sensitivities with respect to some parameters as part of the finite element analysis and are first introduced in static structural design optimisation [27]. Haftka [28] developed a numerical approach for the sensitivity analysis of structures modelled by finite elements. His work was limited to temperature sensitivities for design parameters in a thermal system. Mukherjee *et al.* [29] presented a formulation for accurate determination of design sensitivities for shape optimisation in structural problems subject to small non-elastic (plastic or visco-plastic) deformations. Using this approach, rates of quantities such as displacements, stresses and non-elastic strains with respect to a shape design parameter can be obtained and used in nonlinear gradient based numerical optimisation techniques. Tortorelli and Haber [30] presented an *adjoint* method for design sensitivity analysis of transient heat conduction problems where variations of a general design

functional are expressed in explicit form with respect only to the analysed design parameters (e.g. material properties, geometric dimensions).

Recent engineering applications require a general and robust optimisation framework that can be linked to analysis tools and used for a wide range of problems. One of the recent forums on Optimisation in Industry [197] has focussed on the present advances in design optimisation technology and its implementation into the design process for industrial applications. Design optimisation has been applied to aircraft wing design, automotive radio and navigation systems structural and thermal specification, calibration of combustion engines, models of investment of ethylene and its derivatives, and optimal control of vibration in flexible structures [197].

Recently, a branch of optimisation known as *topology optimisation* (TO) has also been applied to engineering design [198-200]. Topology optimisation finds the optimal structural configuration that meets a specified criterion by changing the topology of the structure starting with a coarse design that represents the outer boundary. Topology optimisation identifies the optimal material distribution in the structure based on finite element stress analysis of the current topology model.

1.2. Computer Modelling and Optimisation for Electronic Packaging

Component package design and assembly process in the electronic manufacturing requires careful observation and pre-planning to achieve the ultimate goal of manufacturing low cost and reliable products [34, 59, 60, 61]. Computer modelling is now being used extensively in microelectronics to complement experimentation, where it can help to identify phenomena that would be too costly and even impossible by experiments alone. Computer modelling analysis, and in particular the finite element method, can predict the behaviour and responses of a package, system or process [34-38]. Different aspects in microelectronics manufacturing and associated problems [54] have been investigated in recent years with special attention to electronic packaging process and corresponding reliability issues.

Assembly technologies such as Ball Grid Array (BGA) [136] and Flip-Chip [34] have been studied extensively through experiments to provide knowledge on the process characteristics [34, 52, 57, 64] and to evaluate reliability performance [34, 58, 137, 138]. Computer modelling

provides knowledge on characteristics such as solder interconnect durability[39, 40], package deformation [41], reflow soldering process [42], potential failure sites [43], solder joint reliability and life predictions[44, 45] and many other packaging issues.

1.2.1. *Design Improvement using Parametric Study Approach*

Predicting the behaviour of an electronic package assembly or process using finite element analysis (FEA) is a widely accepted approach that provides knowledge and better understanding. Normally, a single design parameter is changed and the system is evaluated through finite element analysis to get insight of the physical quantity relationships. This approach, known as *parametric study*, can be observed as a simple form of optimisation practiced widely. Variety of problems and their corresponding investigations are published and available in the science journals, conference proceedings and other literature sources, e.g. references [46-53].

The following paragraphs in this section (Section 1.2.1) outline some published investigations in the area of microelectronics applications where the studies do not exploit any of the forms and advantages of the systematic numerical optimisation. They adopt only the parametric study approach as a strategy to identify the route towards a possible design improvement.

Maximum allowable power dissipation within typical system level environments is investigated in the work by Chambers *et al.* [53] for a number of variables (die size, thermal vias and bumps, heat sink effect, natural/forced convection, PCB head loading and conductivity). Iyengar and Bar-Coohen [79] observed the problem of optimizing the design of multiple fin arrays. Least-material optimisation of different fin arrays is performed using a simple parametric study procedure. The analysis results on heat dissipation from a single fin of the heat sinks for different geometric configurations (fin dimensions and shape) are compared to indicate optimal volumetric efficient fins. An efficient approach of analyzing thermal via pads is discussed in [50] together with the thermal resistance improvement based on a number of via design parameters using analytical models and performing analyses on different geometric scenarios. These simulation results are not incorporated into numerical optimisation routine.

A major area of research relates to solder joint reliability and life time predictions. Lu and Bailey [46, 47] investigated the impact of solder interconnects stand-off height and other geometric package dimensions, for flip-chip packages, to detect key parameters that affect the package lifetime and to determine the "single parameter vs. package response" relationship. Computer modelling analysis predicts the evolution of thermal stresses and strains in the whole package during a thermal cycle. The creep strain values from the analysis are then used in an appropriate lifetime model to estimate the number cycles to failure (lifetime). The same approach is used by Popelar [48] and Michaelides et al. [49] to investigate flip-chip reliability based on solder fatigue modelling. Le Gall et al. [51] investigated the influence of die size on the magnitude of thermo-mechanical stresses in the flip-chip packages.

Typical underfilled flip-chip structures and the FEM-computations on solder shear strain combined with fatigue life evaluations using Coffin-Manson relation are observed by Gektin et al. in a number of publications [55, 56] to investigate underfill properties for improved reliability. Thermal shock (TS) and accelerated thermal cycling (TC) tests in characterization flip-chip reliability performance [62] have been also investigated and compared with finite element modelling results. Subsequently, the authors developed correlation factors for TC and TS tests. Accelerated testing, life prediction methodologies and constitutive modelling for solder joint reliability assessment are discussed and critically reviewed by Qian and Liu [65]. They emphasized the importance of constitutive modelling in terms of correct predictions of inelastic strain ranges, inelastic strain energy density and creep-plasticity interaction under various accelerated test conditions. Possible misleading results in the life-time prediction using energy-based and strain based approach are also outlined.

Using experimental data for reliability of solder joints of power resistors under non-isothermal conditions, Akay et al. [63] developed volume weighted technique for life time predictions using combined heat transfer and thermal stress analyses.

A study carried by Wu et al. also somehow misleadingly contains the words "design optimisation" in the title [80]. The aim of their work is to develop a physics-based, validating modelling methodology for effective simulation of solder joint formation and prediction of possible defects (solder bridging, solder opens, insufficient soldering, etc.). The ultimate goal is to determine the optimal solder joint configuration (bond pad size, stencil aperture size,

solder volume, etc.) in a cost effective manner. Surface Evolver software tool [139] is used to predict solder joint shape formation. Different investigated scenarios are evaluated separately, and subsequently compared to determine the best option. No optimisation algorithms for systematic search of the optimum solution are used.

Although solder joint reliability is the major concern for electronic packaging industry there are concerns in other design aspects. Interfacial de-lamination due to the presence of dissimilar materials and thermal loading is another major area of research where numerical modelling provides knowledge and further understanding of the physical phenomena. Different energy based approaches for failure criteria are developed to predict interfacial de-lamination using energy release rate, stresses and strain energies and the corresponding values for the interfacial strength and fracture toughness [66, 67].

1.2.2. *Design Improvement using Optimisation Approach*

Simulation-based optimisation is the term used to denote virtual improvement of a component design or process using numerical optimisation and statistical techniques that explore the evaluation of the observed design based on computer analysis predictions (FEM in this research). A commonly used approach in optimisation known as Response Surface Methodology (RSM) [142] combines mathematical and statistical techniques. Statistical Response Surface (RS) models are functions in explicit mathematical form (e.g. polynomials, multi-quadratic functions, and etc.) that approximate certain system/process responses based on Design of Experiments (DoE) data. Subsequently, the RS models are used to search for optimal parameters by applying numerical optimisation methods such as gradient-based search algorithms. DoE and RSM approaches are discussed in Chapter 3.

Optimisation techniques where the search for optimal solution relies on direct evaluations from FE analysis is another possible route towards effective design optimisation. This strategy has its own advantages and disadvantages when compared with optimisation process that utilizes the RSM. These aspects are discussed and demonstrated in details in number of papers [177-182] related to this research and throughout the thesis. Numerical optimisation (emphasis on gradient-based search techniques) is outlined in Chapter 4.

Exploration of numerical optimisation techniques in microelectronic design and process improvement has begun and marks a clear step forward in the usage of numerical simulation tools by benefiting not only from their *prediction* capabilities but also the *design* features. The author considers the published material [69-78, 81-88], referred and discussed in the following paragraphs of this section, as the closest one to his work of introducing a general design approach based on integrated FV/FEA (finite volume/finite element analysis) and numerical optimisation. Most of these reviewed references appeared at the same time when outcomes and publications from this research programme took place.

Design of Experiments and RSM are used in many areas of research including recent applications in microelectronics packaging design. First applications of building RS models to describe different aspects of the design behaviour are limited to the point of expressing a certain design response from FEA via approximation which is a function in explicit mathematical form. These RS models are not directly linked to an overall optimisation procedure. They are used to provide knowledge about influence of design parameters on specific aspects of analysed system and to indicate their good (optimal) values. In 1993 Corbin [82] used design of experiments (DoE) techniques and finite element analysis to evaluate the thermal strain sensitivity of solder interconnects in ceramic modules to structural variables (pad diameter and solder volume). Zahn [71] also used DoE and RSM to provide a more accurate thermal characterization of a multi-chip module package. Again, the exploration of the constructed RS models is not linked directly to any optimisation process.

Product Model-Based Analysis Models (PBAMs) are described as highly automated modules for designer usage and utilised for design optimisation in the work presented by Cimtalay *et al.* [69]. Solder joint fatigue life is improved by changing the PBAM inputs parameters in a systematic manner using optimisation methods.

Mahajan [70] explored artificial neural networks techniques to incorporate their capabilities into a modelling framework. The models are tools for simulation, optimisation and process control. The author incorporated in the overall optimisation procedure neural nets as a tool used to approximate behaviour of the system instead of statistical regression polynomial RS models. This tends to be the only difference in the whole classical polynomial-based RSM optimisation procedure since the optimisation still adopts gradient search algorithms.

Response Surface Models are developed also and tested as modular, decomposed analysis procedures by Zhang, Subbarayan and co-authors [72]. They developed and evaluated response surface models based on feed-forward, back propagation neural networks as well as linear regression models for predicting the fatigue life of solder joints in area array packages.

In 1996 Subbarayan [81] developed a procedure for automated finite element shape and life time prediction of solder joints in flip-chips and BGA's. The author discusses some of the issues that have been acting as obstacles for implementing such automated CAD design (e.g. implementation of an automated mesh generation process for general 3D solder geometry) and recognizes the importance of such implementation for automated finite element calculation which can enable subsequent optimisation of design. As a result of this work Deshpande, Subbarayan and Mahajan [73] developed an automated optimisation procedure and demonstrated their software framework by optimizing the solder joint fatigue life. In their work, solder joint shape is parameterized and an approximate function based on the solder joint shape parameters is constructed using design of experiments and artificial neural network models. The work of the authors, published in 1997, demonstrates clearly the ideas of combining different optimisation strategies to finite element predictions on reliability of solder joints in electronic packages. Global approximation techniques as an efficient route towards effective method for optimizing electronic packages were developed and compared with the direct gradient-based optimisation approach.

Li, Mahajan and Nikmanesh applied the statistical-neural network modelling approach to process optimisation of fine pitch stencil printing for solder paste deposition on pads of printed circuit boards [74]. In his study they minimise the variation of the solder paste height by setting optimal design parameters. The approach again is based on design of experiments and neural networks. Numerical techniques for optimisation are used subsequently to find the optimal values for design parameters using approximated functions from the neural network models.

Design-of-simulations methodology is used to study a palletized structure for thin film processing [75]. In their work, Variyam and Sitaraman investigated a low die-to-substrate ratio to reduce the die backside stresses and therefore die-cracking. Simsek and Reichl investigated the electrical characterization of different BGA packages with applications in automotive manufacturing sector [76]. This work focuses on analyzing 2D and 3D models

with respect to signal propagation and electromagnetic field distribution. Variation of through-hole vias size parameter is used to indicate the optimal geometry with respect to signal deflection.

Craig *et al.* [77] used mathematical optimisation techniques to minimize heat sink mass subject to maximum allowable heat sink temperature, constant cooling fan power and heat source. The authors used commercial CFD software to carry out the heat calculations on a parameter defined model of a heat sink. Optimisation is carried out through the transformation of defined constrained design problem into unconstrained by formulating penalty functions [78]. Chang and Lin also addressed in their work thermal optimisation problems [88]. They explored the Taguchi statistical experimental design and analysis of variance to determine the optimal settings of thermally enhanced package materials, geometric dimensions and heat spreader location. The authors minimize the thermal resistance and package warpage subject to a constraint for maximum allowable residual stress in the silicon die.

Significant research and development of a methodology for virtual design in reliability of electronic packages is done in recent years by Zhang *et al.* (Philips Semiconductors) [83-87]. The group proposed so-called "optimized Designing in Reliability based on the method of Physics of failures". The ultimate goal of their virtual thermo-mechanical prototyping strategy is to predict, evaluate, qualify and optimize the thermo-mechanical behaviour of electronic products before their actual physical prototyping and manufacturing. It is a methodology based on "reliable FE thermo-mechanical models" specifically designed for microelectronic applications and capable to predict efficiently and accurately the behaviour of electronic products and processes. Design optimisation is carried out using RSM combined with optimisation methods. Problems addressed by Zhang *et al.* include optimisation of a package geometric design parameters such as the solder and lead-frame thicknesses [84], and die thickness, die length and heat sink thickness [83] against possible failure mode of vertical die cracking.

1.3. Multi-Physics Finite Element / Finite Volume Modelling

In computer modelling, different physical phenomena are often treated as being uncoupled. As a result, separate analyses of the same system are performed to obtain complete predictions of the system response and behaviour. In such an analysis approach the real complex interactions between solved quantities are treated as first-order approximations.

The coupling of various physical processes in one mathematical model can be carried out typically either by considering additional terms in the governing equations (equation coupling) or by allowing the constitutive law to depend on the interacting field (constitutive law coupling). An example of the first category is the effect of the temperature gradient on the stress field, which can be captured by adding a thermal stress term to equilibrium equations. A less familiar example within the same class of problems is the effect of diffusion (clustering) and chemical reactions (reaction products) on the stress field, which can be accounted for by the eigenstrain formulation [89]. An example from the second category is the influence of temperature upon diffusion which can be captured by simply considering the diffusion coefficients to be temperature dependent.

Yu and Fish developed a systematic approach for analyzing multiple physical processes interacting at multiple spatial and temporal scales [90]. They consider interacting physical processes which include mechanical, thermal, diffusion, chemical and electromagnetic fields. The authors propose and apply a computational framework to coupled thermo-viscoelastic composites with microscopically periodic mechanical and thermal properties.

Importance of the multi-physics component in computational modelling of physical processes and phenomena [91, 98] imposes challenges to R&D organizations to develop mathematical methods and software tools for virtual analyses of processes and systems with complex physical behaviour and relations. Numerical algorithms for multiphysics modelling and related software development [99, 100, 101, 118] are a major area of research in recent years at the University of Greenwich. The modelling techniques are based on finite volume methods that are able to represent the physical domain using an unstructured mesh [99, 102-104]. Initially developed for studying fluid dynamics applications, these algorithms have been extended to offer coupled solid mechanics calculations within the same simulation

framework by adopting solution methods for the relevant elastic/visco-plastic constitutive equations [105-109].

Developed multi-physics modelling techniques at the University of Greenwich are adopted by Bailey *at al* [92-96] to predict a range of interacting physics-based phenomena associated with the manufacturing process in microelectronics. Bailey and Boettinger [92] investigated the fillet lift defect and the behaviour of solder materials during the process of solidification and cooling using coupled solidification and stress calculations. To obtain an insight the complexity of solder paste dynamics process, Wheeler and Bailey [93] ran models capable of studying simultaneously heat transfer, melting, fluid convection, surface tension and solidification factors. The authors investigated also the macroscopic processes of solder joint formation [94, 95, 110] by modelling the mechanical response of solder materials as they cool, solidify and then deform. These papers provide a modelling strategy for solder shape, solidification history and resulting stress profile predictions. Physical phenomena of surface tension in joint shape formation, thermal convection in liquid solder, change of phase, deformation across solder-pad-board assembly and rest of the components, and their coupling are addressed in this research.

Bailey *et al.* [96] used the multi-physics approach (coupled heat transfer, solidification, fluid flow, void movement and thermal-stress) with reliable physics of failure and damage models to investigate the reliability of chip components on printed circuit boards. In collaborative work carried out by Flomerics Ltd. [201] and the University of Greenwich, Parry, Bailey and Aldham [97] discussed the benefits for electronics companies of integrated analysis software framework for system-level thermal, stress and electromagnetic predictions if the modelling approach is incorporated in the early stages of a product development. Using the existing knowledge on multiphysics modelling methodology, the team observed and developed [97, 111] numerical strategies for unified modelling framework specifically designed for virtual multi-physics analysis via computer modelling/simulation of electronic systems and components. The technology is using the latest numerical techniques and algorithms to develop software modules for stress and electromagnetic prediction coupled with the Flomerics CFD thermal tool, FLOTHERM.

The importance of reliable multi-physics modelling in the realization of virtual prototyping strategy is discussed by Zhang *et al.* in reference [86]. They demonstrated combined thermal, thermo-mechanical and vapour modelling of a BGA package. The authors carried out also parameter sensitivity analysis based on integrated DoE, solder fatigue prediction models and simulation-based optimisation [87].

In general, multiphysics modelling is subject to continuous development in different aspects [107, 108, 112, 113] and application areas [114-117] that are beyond the scope of the thesis.

1.4. Overview of the Thesis

The introductory **Chapter 1** was intended to outline some of the very important challenges to the microelectronics industry at present. It discussed the major motivation behind this research and why this work can be important for engineers who are working in the area of electronic packaging design. Also, the chapter provides a brief overview on optimisation, outlines some of the major achievements in its history, first applications of these techniques in engineering design and integration with finite element analysis. The role of modelling and optimisation in microelectronic design was reviewed. As the ultimate goal of the research is to develop design framework based on integrated optimisation with multi-phenomena analysis of electronic systems, the multi-physics modelling approach was also observed.

In **Chapter 2**, a background on multiphysics computer analysis that exploits finite element computational solid mechanics and finite volume fluid dynamics techniques is presented. Brief discussion is given on the Partial Differential Equations (physics governing equations), finite element discretization procedures and domain discretization techniques.

Statistical Design of Experiments (DoE) and Response Surface Methodology (RSM) techniques and some related aspect of the theory such as Residual Analysis and Analysis of Variance are observed and detailed in **Chapter 3**. The Design of Experiments and related Response Surface (RS) modelling are key part in the developed virtual design tool and demonstrated for different microelectronics problems in this research.

Chapter 4 discusses first some theoretical aspects of the numerical optimisation techniques. Gradient-based algorithms for unconstrained and constrained optimisation problems are

explained in detail. Non-gradient optimisation is also presented briefly. The outlined numerical techniques follow closely the software implementation of these optimisation algorithms and strategies. The second part of this chapter discusses then the approach of integrating a collection of DoE statistical/numerical optimisation modules (VisualDOC, Vanderplaats R&D) with the multiphysics finite volume analysis tool (PHYSICA, the University of Greenwich). Integrated software framework for optimisation modelling is presented. The underlying interaction mechanism between optimisation calculations and the simulation analysis within the automated software design procedure are specified.

The rest of the thesis focuses on discussing microelectronics applications. The optimal design solutions for different problems with increasing complexity of the analysis, from single to multi-physics, are presented. These applications are analysed using the integrated software framework.

In **Chapter 5** the traditional approach of finite element analysis and design improvement based on pure parametric study (what-if scenarios) is demonstrated. The analysis investigates a recently developed bonding technology known as "no-flow underfill" packaging. The study combines experimental and modelling work for a flip-chip package and provides directions towards successful implementation of the novel process. The problem investigated in Chapter 5 uses none of the integrated optimisation routines. Some of the process optimal trends from this parametric study are compared with the results obtained in Chapter 6, where the same flip-chip package is analysed further but now using the *optimisation modelling* approach.

Thermo-mechanical reliability of the flip-chip package assembled using "no-flow underfill" technology process is discussed in **Chapter 6**. Thermal fatigue life of solder joints is optimised by using the integrated Optimisation-FEA design tool utilizing a single-physics analysis for stress/solder creep strain due to the prescribed thermal loading. Key parameters including underfill properties and package geometry are optimised automatically and systematically. This design process is performed with both DoE-Response Surface Modelling and Direct gradient-based numerical optimisation to allow the demonstration of these different strategies and their comparison.

Chapter 7 details the development of a new wave soldering machine and process for lead free soldering applications. Nitrogen ($N2$) is inert into the machine to ensure oxygen free environment. The study focuses on designing the process in a way to provide the minimum possible consumption of nitrogen. The cost effective soldering is achieved using the proposed *optimisation modelling* methodology and integrated software tools. The virtual computer analysis has increased complexity and couples both the heat and flow calculations. Optimisation is carried out using the Incrementally Refined RS approach combined with novel evolutionary search techniques (Particle Swarm Optimisation).

The last application of the virtual design methodology using simulation-based optimisation demonstrates an example for obtaining optimal thermal management solutions in electronics. The emphasis in **Chapter 8** is on the complex physics and interacting phenomena that take place during the operational life of the electronic systems, requiring multi-physics modelling, and related component reliability issues. The coupled thermal-flow-stress calculations are used to capture realistically the heat transfer and thermally induced stress behaviour of the components on circuit board subject to air cooling. Optimisation methods are used to find the optimal placement of a BGA component on the board that will ensure minimum die stress and junction temperature. This chapter demonstrates a route towards effective and optimal thermal management solutions in electronics using combined multi-physics analysis and design optimisation.

The final **Chapter 9** provides conclusions on the undertaken research. It summarises the proposed methodology for optimal design solutions in electronic packaging, product manufacturing and process control. The chapter outlines the advantages and benefits of adopting such a design approach in electronics, and finally details some aspects of future work.

Most of the results obtained during this research programme are disseminated in a number of national and international forums on electronic packaging. The **List of Publications** at the end of this thesis provides details on all published results (first author and co-author). Two journal publications, in *Soldering & Surface Mount Technology* (2002) and **Springer-Verlag** book *Optimisation in Industry* (2002), summarize the most important results and describe the undertaken *optimisation modelling* approach. A paper presented at the Fourth

International Symposium on Electronic Packaging Technology (co-author) won the **best paper** award of the forum and the extended version of results was published in the *Microelectronics Reliability* (2002) journal. Two other author's publications at the Fifth International IEEE Symposium on High Density Packaging and Component Failure Analysis in Electronics Manufacturing (2002) and at the 4th International Conference on Thermal & Mechanical Simulation and Experiments in Micro-Electronics and Micro-Systems (EuroSIME2003) have been presented as **keynote presentations** in optimisation tracks of both conferences. Other electronic packaging meetings where part of the early research was disseminated include InterPACK2001 (2001, USA), the 2001 TMS Annual Meeting (2001, USA) and APACK2001 (2001, Singapore).

CHAPTER 2

MULTIPHYSICS MODELLING

2.1. Finite Element/Finite Volume Methods for Multiphysics Modelling

The finite element method (FEM) [128] is a very popular numerical method for solving engineering problems and applicable to areas such as structural analysis, heat transfer, fluid flow and electromagnetic potential. The finite element method has developed rapidly over the past 40 years and its growth is linked to the growth of modern high-speed digital computers.

In 1940's, Hrennikoff [120] and McHenry [121] made the first contributions to the development of finite element method by solving stresses in solids using one-dimensional elements. In 1956, Turner *et al* [122] proposed, for the first time, the numerical procedures of the direct stiffness method and two-dimensional finite elements. Clough introduced the *finite element* term in his work on solving plane stress problems using both triangular and rectangular elements [123].

In the early 60's Martin [124], Gallagher *et al* [125] and Melosh [126] extended the finite element method to three-dimensional problems. Applications of FEM in the early 60's were related to the modelling of small strains, elastic material behaviour and static loading. The first applications of material non-linear behaviour were presented by Gallagher *et al* [125] and later, in 1968, were extended by Zienkiewicz *et al* [127] to visco-elastic problems.

The finite element method – originally developed to solve solid mechanics problems - was extended and modified to general solutions of continuum phenomena in fields such as acoustics, heat transfer, fluid flow and electromagnetics [128, 129].

In the area of Computational Fluid Dynamics (CFD) the finite volume (FV) method is used to analyze problems ranging from modelling combustion and free surface flows to multiphase flows [130, 131]. The ability of CFD techniques to conserve physical quantities at a local and global level allows the FV-CFD codes to model non-linear physical phenomena [130,131]. Finite volume methods on structured [132] and unstructured [109] mesh domains for solid mechanics problems were developed in recent years. Details of the finite volume (FV) methods are given in Section 2.3.

The partial differential equations of equilibrium, compatibility and stress-strain for problems in linear static elasticity and plasticity are used in computational mechanics to predict the deformation and stress/strain of the modelled system. In computational fluid dynamics, the governing partial differential equations have terms representing the physics of the process (e.g. conduction, convection, etc.). In the finite element method (FEM) the geometric domain is split into a number of simpler sub-regions. Each sub-region is called an *element*. The assembly of elements is the *mesh*. The fixed points in the elements defining their vertices are called mesh *nodes* (Figure 2.1)
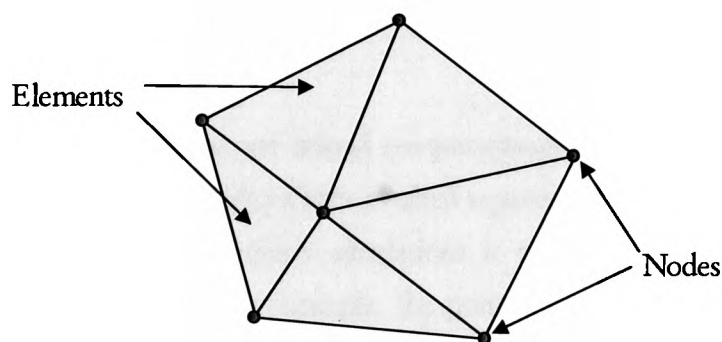
Figure 2.1: Finite Elements and Mesh Nodes

The relevant physical quantity or property of the element is predicted within each element, for example, the relationship between forces and displacements for a structural element. In the Computational Solid Mechanics (CSM), the element equations that represent the values

of the solved property are formed at the mesh nodes. In the Computational Fluid Dynamics (CFD) and related finite volume discretization techniques, the element equations are for the solved property at the mesh element centres. The process of assembling the element equations takes into account the properties of all elements and the interactions between them.

The FEM procedure for a CSM, or the FV procedure for a CFD analysis has the following steps:

1. Divide domain of the structure into elements;

2. Discretize the Partial Differential Equations (PDE's) to form Algebraic Equations solvable by computers;

3. Assemble the elements to obtain an approximate system of equations for the solved quantity at the mesh nodes (FEM) and/or centres (FV) for the domain as a whole;

4. Solve the system of equations and obtain the solution of the observed quantity at the nodes (FEM, e.g.. displacement) or element centres (FV, e.g. temperature);

5. Derive the secondary quantities associated with the elements (e.g.. stress, strain in a solid mechanics problem);

6. Interpret the results. For example, a structural stress analysis provides important information on the structure locations that are characterized with high levels of stress/deformation.

Historically, numerical techniques and related computational technologies for solving solid mechanics and fluid dynamics problems have evolved separately. The design of an electronic package or component using computer simulations is an area where the multi-physics modelling approach is essential. For example, the printed circuit board and its enclosure design, and components and fan locations affect the profile of the cooling airflow. This in turn influences the temperature field across electronic components and will have also an impact on the thermal stresses. Thermal stresses are relevant to the thermo-mechanical reliability of the entire system and its performance.

A possible option to perform a multi-physics analysis is by using the multiple-codes approach. This approach uses integration of different analysis codes capable of covering the

multi-physics aspects of the process. In this case, the user needs to ensure the required interaction between the software packages, i.e. data transfer and data translation in the process.

This research utilizes the other possible, so-called *single-code*, approach as a route towards executing multi-physics simulations. In this approach, the existing libraries of solvers for computational solid mechanics (CSM) and computational fluid dynamics (CFD) (i.e. numerical procedures to solve simultaneously for flow, solid mechanics, heat transfer, electromagnetics, etc.) are integrated into a single software package. The simulation results presented in this thesis were carried out using *PHYSICA* [118] - a single-code multi-physics software package, detailed further in Section 2.4.

2.2. Computational Solid Mechanics (CSM)

2.2.1. *Basic Concepts*

The main topics in the computational solid mechanics and in particular the linear elastic and visco-plastic analysis are:

- Transmission of the loads when they are to solid mechanical systems;

- Theoretical basis to the fundamental equations that represent relations between strain, stress and displacements;

- Finite element method;

The most commonly used finite element method for computational solid mechanics is based on the displacement shape functions (stiffness method) [128]. The behaviour of an element is defined by the matrix equation $\mathbf{F}^e = [k]\mathbf{d}^e$ where $\mathbf{F}^e$ is the element nodal force vector, $[k]$ is element stiffness matrix and $\mathbf{d}^e$ is the element nodal displacement vector. The vector notations in this chapter are for vector columns. In 3-D Cartesian coordinate system, three independent displacements in x, y and z directions define each nodal displacement (similarly for the nodal forces). Components of the displacement vector $\mathbf{d}^e$ represent the degree of freedom for the element. At the first step of the finite element procedure, the individual

element stiffness matrices and the nodal force vectors are obtained for all mesh elements. The individual element equations are assembled after that into a global system related to the whole mesh domain.

The matrix equation for the whole system is expressed as

$$\mathbf{F} = [K]\,\mathbf{d}$$

where $\mathbf{F}$ is the vector of nodal force, $\mathbf{d}$ is vector of nodal displacements, and $[K]$ is the global stiffness matrix. The numerical solution of the global system of equations provides the approximate solution for the solved variable (displacement). The solution for displacements at the nodes of the mesh domain is used subsequently to calculate the secondary element quantities (i.e. the stress or strain).

In linear static elasticity, the *total stress* at central point C across a plane through a solid system with small flat area, for example, δA and normal direction given by the *x-axis* of Cartesian coordinate system (Figure 2.2) is

$$\sigma = \lim_{\delta A \to 0} \frac{\delta F}{\delta A}$$

where δF is the resultant of the forces exerted by the material on the positive side of δA (positive side is defined by the normal vector to the plane, i.e. the *x-axis* in this example).
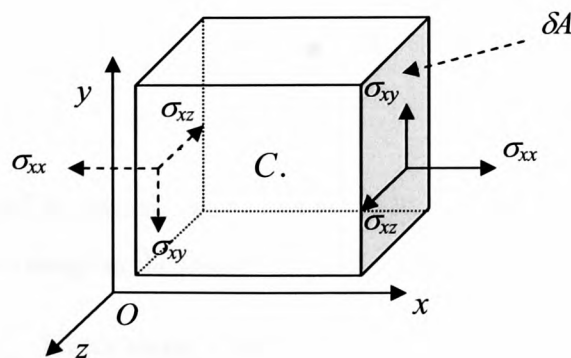
Figure 2.2: Stress Components

The δA lies in the y-z plane (the cutting plane) for the Cartesian coordinate system shown in Figure 2.2. The total stress σ can be resolved into three components, σ_{xx}, σ_{xy}, σ_{xz}, in the Ox, Oy and Oz directions. The *normal stress* is the component σ_{xx}, and the *shear stresses* are σ_{xy} and σ_{xz}. If σ_{xx} is positive then the direct stress is **tensile**, otherwise σ_{xx} is **compressive**. Similar expressions can apply to cutting planes perpendicular to the y and z directions. Figure 2.2 represents the 3-D stresses at the point C - the centroid of the very small cuboid with sides parallel to the Cartesian coordinate system and surface area δA. Stresses at the two of the sides are shown only, but similar stresses are defined on the rest of the sides. For an isotropic system in the equilibrium state, the shear stresses satisfy the relations $\sigma_{xy} = \sigma_{yx}, \sigma_{yz} = \sigma_{zy}, \sigma_{zx} = \sigma_{xz}$. Therefore, the six quantities of stress $\sigma_{xx}, \sigma_{yy}, \sigma_{zz}$, $\sigma_{xy}, \sigma_{yz}, \sigma_{zx}$ define the stresses acting on the coordinate planes associated with this small domain and the central point C.

2.2.2. Differential Equations of Equilibrium

Let f_x, f_y and f_z be the components of the body forces per unit volume when they are resolved in the Cartesian coordinate system. The general equilibrium equations governing the conservation of force are:

$$f_x + \frac{\partial \sigma_{xx}}{\partial x} + \frac{\partial \tau_{xy}}{\partial y} + \frac{\partial \tau_{zx}}{\partial z} = 0$$

$$f_y + \frac{\partial \tau_{xy}}{\partial x} + \frac{\partial \sigma_{yy}}{\partial y} + \frac{\partial \tau_{yz}}{\partial z} = 0 \tag{2.1}$$

$$f_z + \frac{\partial \tau_{zx}}{\partial x} + \frac{\partial \tau_{yz}}{\partial y} + \frac{\partial \sigma_{zz}}{\partial z} = 0$$

The stresses cause small linear and angular displacement in the deformable structure. *Direct strain* ε_{xx} defines the change in the length of a line, oriented in x-direction, and produced by acting direct stress σ_{xx}. Direct strain is defined as $\varepsilon_{xx} = \dfrac{\Delta x}{L}$, where L is the length of the same line (on the normal n) that undergoes a change in the length Δx. *Shear strain* at a point is defined as the change in the angle between two mutually perpendicular lines

associated with the planes intersecting at the point. The shear strains ε_{xy} and ε_{xz} define the change in the angle produced by the shear stresses acting in x-y and x-z planes respectively. Similar strain expressions exist in the y- and z- directions in the 3-D Cartesian co-ordinate system.

Based on geometric observations, the relations between strain and displacement at a point are specified as $\varepsilon_{xx} = \partial u / \partial x$ (direct strain in x direction) and $\varepsilon_{yz} = \dfrac{\partial v}{\partial z} + \dfrac{\partial w}{\partial y}$ (shear strain) where $\overline{\mathbf{d}} = (u, v, w)^T$ stands for the displacement vector at the point. Similar expressions of the strain-displacement relationship associated with the other directions of co-ordinate system exist. The total strain ($\varepsilon^{(tot)}$)-displacement relationship is given in matrix equation form as $\varepsilon^{(tot)} = \mathbf{L}\overline{\mathbf{d}}$, i.e.

$$
\begin{Bmatrix} \varepsilon_{xx} \\ \varepsilon_{yy} \\ \varepsilon_{zz} \\ \varepsilon_{xy} \\ \varepsilon_{yz} \\ \varepsilon_{zx} \end{Bmatrix} = \begin{bmatrix} \dfrac{\partial}{\partial x} & 0 & 0 \\ 0 & \dfrac{\partial}{\partial y} & 0 \\ 0 & 0 & \dfrac{\partial}{\partial z} \\ \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial x} & 0 \\ 0 & \dfrac{\partial}{\partial z} & \dfrac{\partial}{\partial y} \\ \dfrac{\partial}{\partial z} & 0 & \dfrac{\partial}{\partial x} \end{bmatrix} \begin{Bmatrix} u \\ v \\ w \end{Bmatrix} \tag{2.2}
$$

The linear relationship (Hooke's low) between the stress and the elastic strain for a piece of material in tension or compression gives the *modulus of elasticity* (Young's modulus) E. When the material undergoes stretching, there is also a lateral contraction. The direct elastic strains, denoted via superscripts "*(el)*", (say in y- and z- direction) are related to the direct strain in the x direction by a constant v, $\varepsilon_{yy}^{(el)} = \varepsilon_{zz}^{(el)} = -v\varepsilon_{xx}^{(el)}$, where v is called *Poison's ratio*. Shear stress (e.g. in x-y plane) σ_{xy} and shear elastic strain $\varepsilon_{xy}^{(el)}$ are also linearly dependent via *shear modulus of elasticity* $G = \dfrac{E}{2(1+v)}$. The generalization of Hooke's law

in 3-D gives the relationship between six stresses and six elastic strains for an isotropic homogeneous linear elastic material:

$$\begin{Bmatrix} \sigma_{xx} \\ \sigma_{yy} \\ \sigma_{zz} \\ \sigma_{xy} \\ \sigma_{yz} \\ \sigma_{zx} \end{Bmatrix} = \frac{E}{(1+v)(1-2v)} \begin{bmatrix} 1-v & v & v & 0 & 0 & 0 \\ v & 1-v & v & 0 & 0 & 0 \\ v & v & 1-v & 0 & 0 & 0 \\ 0 & 0 & 0 & \dfrac{1-2v}{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & \dfrac{1-2v}{2} & 0 \\ 0 & 0 & 0 & 0 & 0 & \dfrac{1-2v}{2} \end{bmatrix} \begin{Bmatrix} \varepsilon_{xx}^{(el)} \\ \varepsilon_{yy}^{(el)} \\ \varepsilon_{zz}^{(el)} \\ \varepsilon_{xy}^{(el)} \\ \varepsilon_{yz}^{(el)} \\ \varepsilon_{zx}^{(el)} \end{Bmatrix} \qquad (2.3)$$

In the matrix form, Equations 2.3 are written in the form $\boldsymbol{\sigma} = \mathbf{D}\boldsymbol{\varepsilon}^{(el)}$. The symmetric matrix in the above relation, $\mathbf{D}$, is called elastic stiffness matrix (also material property matrix).

The constitutive Equations 2.3 state the stress dependency on the elastic strains. The elastic strains $\boldsymbol{\varepsilon}^{(el)}$ depend on the total, $\boldsymbol{\varepsilon}^{(tot)}$, the visco-plastic (or pure plastic), $\boldsymbol{\varepsilon}^{(VP)}$, and the thermal, $\boldsymbol{\varepsilon}^{(TH)}$, strains and are related as follows:

$$\boldsymbol{\varepsilon}^{(el)} = \boldsymbol{\varepsilon}^{(tot)} - \boldsymbol{\varepsilon}^{(VP)} - \boldsymbol{\varepsilon}^{(TH)} \qquad (2.4)$$

Generally, the plastic strain is a function of the stress and the temperature. There are analytical models developed based on experimental data and capable to represent the plastic or visco-plastic behaviour of some materials. For example, a constitutive law for time dependent visco-plasticity of solder alloy, discussed in details in Chapter 5, is adopted in the research to aid the failure modelling of the solder interconnects in electronic packages.

The thermal strains are given by

$$\boldsymbol{\varepsilon}^{(TH)} = \alpha \, \Delta T \qquad (2.5)$$

where $\alpha = (\alpha_1, \alpha_2, \alpha_3, 0, 0, 0)^T$ is the vector of coefficients of thermal expansion (CTE) and ΔT is the thermal load (the temperature change). The first three components of the vector α^T represent the thermal expansion coefficients in x, y and z direction respectively.

For an isotropic material $\alpha_1 = \alpha_2 = \alpha_3$ and for an orthotropic material these values are different, i.e. the material expands and contracts at different level in the different directions under thermal loading. A typical example for an orthotropic material in electronic packaging is the FR4 board with CTE in in-plane directions of order 14-18 ppm/°C and CTE out-of-plane of approximately 60-80 ppm/°C.

2.2.3. *Discretization of Solution Domain*

In the finite element method, the quantities solved in a solid mechanics problem (displacement, stress, strain) are calculated at the nodes (vertices) of the mesh elements, which are placed across the geometric domain. Each of the mesh elements are associated with a material and the relevant materials properties. Figure 2.3 illustrates quadrilateral and triangle mesh elements for a 2D domain.
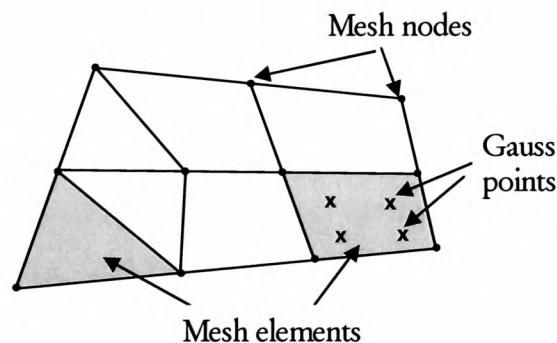
Figure 2.3: Mesh Elements and Gauss Points for Finite Element Method

In the FEM, the governing equations of equilibrium are integrated over the mesh elements. The volume integrals in the three dimensional, or the area integrals in the two dimensional space are approximated by using numerical integration techniques such as the Gauss quadrature. The discretized equations are delivered by using the so-called *reference* element, i.e. each mesh element in terms of global coordinates of its nodes is mapped into an element in a local coordinate system *(s,t,u)* (Figure 2.4).
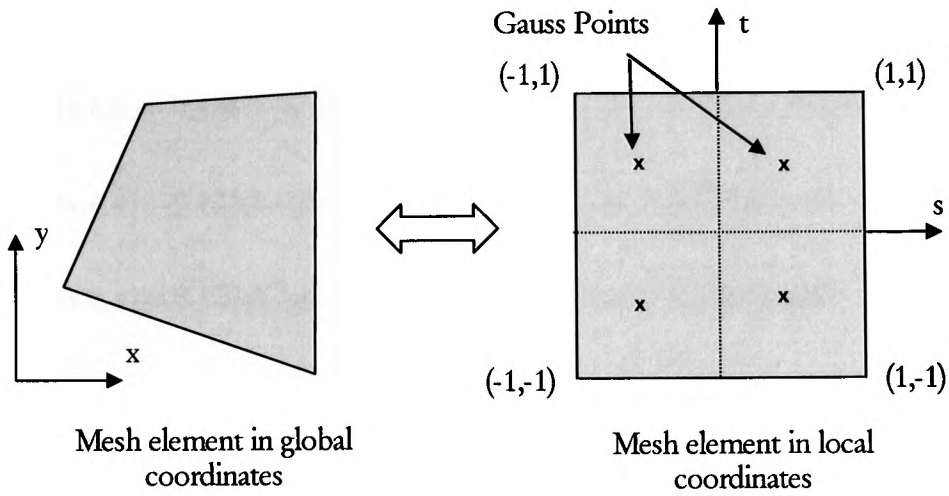
Figure 2.4: Global and Reference Element

The discretization techniques use the reference elements to obtain approximations to the governing equations in local coordinates. The mapping of derivatives of the solved equations back into global coordinates uses the Jacobian matrix.

Calculating a variable ϕ (e.g. displacement) and its derivatives anywhere in a mesh element uses the so-called *shape* functions $N_i(s,t,u)$ and the variable values ϕ_i at the mesh element nodes ($i = 1, m$). These approximations are in local coordinates and are given as

$$\phi(s,t,u) = \sum_{i=1}^{m} N_i(s,t,u)\phi_i = \mathbf{N}\,\boldsymbol{\varphi}$$

$$\frac{\partial \phi(s,t,u)}{\partial k} = \sum_{i=1}^{m} \frac{\partial N_i(s,t,u)}{\partial k}\phi_i \quad , \quad k = s,t,u$$

(2.6)

where m is the number of mesh nodes that define the mesh element. As an example, the shape functions for a four-noded quadrilateral are

$$N_1(s,t) = 0.25(1+s)(1+t) \quad N_2(s,t) = 0.25(1-s)(1+t)$$

$$N_3(s,t) = 0.25(1-s)(1-t) \quad N_4(s,t) = 0.25(1+s)(1-t)$$

In addition, for an eight-noded brick the shape functions are

$$N_1(s,t,u) = 0.125(1+s)(1+t)(1+u) \qquad N_2(s,t,u) = 0.125(1-s)(1+t)(1+u)$$

$$N_3(s,t,u) = 0.125(1-s)(1+t)(1-u) \qquad N_4(s,t,u) = 0.125(1+s)(1+t)(1-u)$$

$$N_5(s,t,u) = 0.125(1+s)(1-t)(1+u) \qquad N_6(s,t,u) = 0.125(1-s)(1-t)(1+u)$$

$$N_7(s,t,u) = 0.125(1-s)(1-t)(1-u) \qquad N_8(s,t,u) = 0.125(1+s)(1-t)(1-u)$$

If the displacement is the solved variable and the mesh element is a four-noded quadrilateral (2-D), then the vector of displacements at the mesh element nodes is $\mathbf{d} = (u_1, v_1, ..., u_4, v_4)^T$ and the shape functions matrix is

$$\mathbf{N} = \begin{pmatrix} N_1 & 0 & N_2 & 0 & N_3 & 0 & N_4 & 0 \\ 0 & N_1 & 0 & N_2 & 0 & N_3 & 0 & N_4 \end{pmatrix}$$

A mesh across the solution domain may consist of different shape elements including also triangles, wedges and tetrahedrals. Specific shape functions are formulated for any of the shapes that can be associated with a mesh element. They will apply in the global-local coordinate mapping calculations [105].

In Equations 2.6, the local co-ordinates are used to approximate the first-order derivatives of displacements in the integrated governing equations. To obtain these derivatives back in global co-ordinates, the Jacobian matrix for the mesh element is used. The Jacobian matrix is defined as

$$\mathbf{J} = \begin{pmatrix} \dfrac{\partial x}{\partial s} & \dfrac{\partial y}{\partial s} & \dfrac{\partial z}{\partial s} \\ \dfrac{\partial x}{\partial t} & \dfrac{\partial y}{\partial t} & \dfrac{\partial z}{\partial t} \\ \dfrac{\partial x}{\partial u} & \dfrac{\partial y}{\partial u} & \dfrac{\partial z}{\partial u} \end{pmatrix}^{-1} \tag{2.7}$$

where each of the components in the matrix is calculated using Equation. 2.6. If X_i are the global x-coordinates of the nodes of a mesh element and x is the local x-coordinate then

$$x = \sum_{i=1}^{m} N_i(s,t,u) X_i$$

$$\frac{\partial x}{\partial k} = \sum_{i=1}^{m} \frac{\partial N_i(s,t,u)}{\partial k} X_i \quad , \quad k = s,t,u \tag{2.8}$$

Similar expressions exist for the y- and z- local-global coordinate transformations of the nodes associated with a mesh element.

The displacement vector at a mesh element point, $\overline{\mathbf{d}} = (u, v, w)^T$, is approximated using the mesh element shape functions and mesh element nodal displacements as $\overline{\mathbf{d}} = \mathbf{N}\mathbf{d}$.

The Jacobian matrix is used in the transformation

$$
\begin{pmatrix}
\dfrac{\partial N_i}{\partial x} \\[2mm]
\dfrac{\partial N_i}{\partial y} \\[2mm]
\dfrac{\partial N_i}{\partial z}
\end{pmatrix}
=
\begin{pmatrix}
\dfrac{\partial x}{\partial s} & \dfrac{\partial y}{\partial s} & \dfrac{\partial z}{\partial s} \\[2mm]
\dfrac{\partial x}{\partial t} & \dfrac{\partial y}{\partial t} & \dfrac{\partial z}{\partial t} \\[2mm]
\dfrac{\partial x}{\partial u} & \dfrac{\partial y}{\partial u} & \dfrac{\partial z}{\partial u}
\end{pmatrix}^{-1}
\begin{pmatrix}
\dfrac{\partial N_i}{\partial s} \\[2mm]
\dfrac{\partial N_i}{\partial t} \\[2mm]
\dfrac{\partial N_i}{\partial u}
\end{pmatrix}
\tag{2.9}
$$

This transformation is used in the process of mapping the solved variable derivatives from local to global co-ordinates.

2.2.4. *Finite Element Discretization of the Equilibrium Equations*

The equilibrium equations (Equation 2.1) can be expressed in tensor form as

$$\frac{\partial \sigma_{ij}}{\partial x_j} = f_i \quad i, j \in \{x, y, z\} \tag{2.10}$$

The finite element procedure requires the integration of these equations over the domain control volume. The first step using the weighted residual approach is to define the overall residual for the whole domain and to require this to be zero:

$$R_\Omega = \int_\Omega W \left(\frac{\partial \sigma_{ij}}{\partial x_j} - f_i \right) d\Omega = 0 \quad i, j \in \{x, y, z\} \tag{2.11}$$

where W are the weighting functions and Ω is the volume of the domain. The Equation 2.11, specified in a tensor form, is equivalent to three integral equilibrium equations for the general three-dimensional discretization procedure. For example, the x-direction equation has the following form:

$$\int_\Omega W\left(\frac{\partial \sigma_{xx}}{\partial x}+\frac{\partial \sigma_{xy}}{\partial y}+\frac{\partial \sigma_{xz}}{\partial z}-f_x\right)d\Omega = 0 \tag{2.12}$$

The integral for each of the stress terms in Equation 2.12 can be split in two integrals. This procedure relies on the product rule of differentiation

$$\int_\Omega f\frac{\partial g}{\partial x}d\Omega + \int_\Omega \frac{\partial f}{\partial x}g d\Omega = \int_\Omega \frac{\partial(fg)}{\partial x}d\Omega$$

If the rule is applied to the σ_{xx} stress term, it will result in the following relationship:

$$\int_\Omega \left(W\frac{\partial \sigma_{xx}}{\partial x}\right)d\Omega = \int_\Omega \frac{\partial(W\sigma_{xx})}{\partial x}d\Omega - \int_\Omega \frac{\partial W}{\partial x}\sigma_{xx}d\Omega \tag{2.13}$$

In Equation 2.13, the divergence theorem is used to transform the integral equation by replacing the volume integral with the corresponding integral over the boundary of the volume domain. If Γ denotes the boundary of the control volume Ω and the $\mathbf{n} = (n_x, n_y, n_z)$ is the outward normal vector at the boundary surface then

$$\int_\Omega \frac{\partial(W\sigma_{xx})}{\partial x}d\Omega = \int_\Gamma W\sigma_{xx}\, n_x\, d\Gamma \tag{2.14}$$

By applying the rules of substitution expressed in Equations 2.13 and 2.14 to the residual form of the other stress terms of the equilibrium equation (Equation 2.11), this equation can be transformed in the form

$$-\left(\int_\Omega \frac{\partial W}{\partial x}\sigma_{ix}d\Omega + \int_\Omega \frac{\partial W}{\partial y}\sigma_{iy}d\Omega + \int_\Omega \frac{\partial W}{\partial z}\sigma_{iz}d\Omega\right) + \int_\Gamma W\sigma_{ij}\, n_j\, d\Gamma = \int_\Omega W\, f_i d\Omega$$

where $i, j \in \{x, y, z\}$. In the matrix form the above equation states

$$-\left(\int_{\Omega} \mathbf{L}^{T} \mathbf{W} \boldsymbol{\sigma} \, d\Omega\right) + \int_{\Gamma} \mathbf{W} \boldsymbol{\sigma} . \mathbf{n} \, d\Gamma = \int_{\Omega} \mathbf{W} \mathbf{f} \, d\Omega \tag{2.15}$$

Let observe the Equation (2.15) with respect to a mesh element control volume. From Equation 2.3 ($\boldsymbol{\sigma} = \mathbf{D}\boldsymbol{\varepsilon}^{(el)}$) and Equation 2.4 ($\boldsymbol{\varepsilon}^{(el)} = \boldsymbol{\varepsilon}^{(tot)} - \boldsymbol{\varepsilon}^{(VP)} - \boldsymbol{\varepsilon}^{(TH)}$) the stress can be expressed as function of the total, inelastic and thermal strain:

$$\boldsymbol{\sigma} = \mathbf{D}(\boldsymbol{\varepsilon}^{(tot)} - \boldsymbol{\varepsilon}^{(VP)} - \boldsymbol{\varepsilon}^{(TH)}) \tag{2.16}$$

If Equation 2.2 ($\boldsymbol{\varepsilon}^{(tot)} = \mathbf{L}\overline{\mathbf{d}}$) is used to substitute the total strain in Equation (2.16) and taking into account the relationship $\overline{\mathbf{d}} = \mathbf{N}\mathbf{d}$ between the displacement vector at a point ($\overline{\mathbf{d}}$) and the nodal displacement vector ($\mathbf{d}$), the following stress-displacement relationship in matrix form is derived:

$$\boldsymbol{\sigma} = \mathbf{D}\,\mathbf{L}\mathbf{N}\mathbf{d} - \mathbf{D}\boldsymbol{\varepsilon}^{(VP)} - \mathbf{D}\boldsymbol{\varepsilon}^{(TH)}$$

If $\mathbf{B} = \mathbf{L}\mathbf{N}$ and by substituting the stress in Equation 2.15 using above relationship, Equation 2.15 can be expressed in terms of the nodal displacement $\mathbf{d}$ for each mesh element control volume CV:

$$-\left(\int_{\Omega} \mathbf{L}^{T} \mathbf{W} (\mathbf{D}\mathbf{B}\mathbf{d} - \mathbf{D}(\boldsymbol{\varepsilon}^{(VP)} + \boldsymbol{\varepsilon}^{(TH)})) \, d\Omega\right) + \int_{\Gamma} \mathbf{W} (\mathbf{D}\mathbf{B}\mathbf{d} - \mathbf{D}(\boldsymbol{\varepsilon}^{(VP)} + \boldsymbol{\varepsilon}^{(TH)})) . \mathbf{n} \, d\Gamma = \int_{\Omega} \mathbf{W} \mathbf{f} \, d\Omega \tag{2.17}$$

The Galerkin finite element approach takes the weighting functions $\mathbf{W}$ are equal to the shape functions for the element control volume ($\mathbf{W} = \mathbf{N}^{T}$) and zero elsewhere. For the internal adjacent control volumes the surface integral balances and does not contribute explicitly in the discretized equation for displacement. The surface integral contributes to the mesh element control volumes at the boundary of the domain. It is approximated by using the specified boundary conditions. The Galerkin finite element method expresses the discretized equation for an internal mesh element with volume V as

$$-\int_{V} \mathbf{B}^{T} \mathbf{D}\mathbf{B}\mathbf{d} \, dV = \int_{V} \mathbf{N}^{T} \mathbf{f} \, dV - \int_{V} \mathbf{B}^{T} \mathbf{D}\boldsymbol{\varepsilon}^{(VP)} \, dV - \int_{V} \mathbf{B}^{T} \mathbf{D}\boldsymbol{\varepsilon}^{(TH)} \, dV \tag{2.18}$$

The mesh element control volumes with faces on the boundary of the domain are still discretized using Equation 2.18 but now this equation contains also the surface integral term from Equation 2.17 in the form

$$- \int_\Gamma \mathbf{N^T\, D\, B\, d\,.n}\, d\Gamma$$

whish is added to the right side of Equation 2.18. The approximation of the above integral uses the boundary conditions being imposed. The boundary conditions can be in the form of prescribed displacement, force or pressure.

Let us denote the system matrix $\mathbf{A}$,

$$\mathbf{A} = \sum_{ele=1}^{totele} \left[\int_V - \mathbf{B^T D B}\, dV \right]_{ele}$$

In the above notations, *totele* is the total number of mesh elements (domain control volumes) and the sum indicates the assembling contributions from all control volumes.

The vector of all degrees of freedom is

$$\mathbf{x} = (u_1, v_1, w_1, \ldots, u_n, v_n, w_n)^T,$$

where n is the number of nodes in the mesh domain, and the vector of the source terms is

$$\mathbf{b} = \sum_{ele=1}^{totele} \left[\int_V \mathbf{N^T f}\, dV - \int_V \mathbf{B^T D} (\varepsilon^{(VP)} + \varepsilon^{(TH)})\, dV \right]_{ele} - \int_\Gamma \mathbf{N^T D B d\,.n}\, d\Gamma$$

The integrals in Equation 2.18 are over the volume of a mesh element control volume and they are calculated in local coordinates using Gauss quadrature. As discussed previously, the transformation to global coordinates uses the Jacobian matrix, e.g. the transformation of the left side integral in Equation 2.18 is

$$- \int_V \mathbf{B^T\, D B}\, dV = - \int_{-1}^{1}\int_{-1}^{1}\int_{-1}^{1} \mathbf{B^T\, D B} |\mathbf{J}|\, ds\, dt\, du$$

where $|\mathbf{J}|$ is the determinant of the Jacobian matrix associated with the mesh element control volume.

The matrix equation $\mathbf{Ax} = \mathbf{b}$ is the final system of discretized equilibrium equations using the finite element procedure. The solution of the system of equations is for the unknown nodal displacements, i.e. the degrees of freedom of each node in the domain.

2.2.5. *Methods for Solving Linear System of Equations*

Two major categories of numerical methods to solving a system of equations $\mathbf{Ax} = \mathbf{b}$ exist:

- Direct methods;

- Iterative methods;

In direct methods, the solution of system of equations is obtained as a result of finite number of mathematical operations. These methods terminate with obtaining the exact solution. The most popular method of this type is the Gaussian elimination.

In iterative methods, an initial guess at the solution is required. After that a sequence of matrix-vector multiplications is performed at each of the iterations of the solution procedure, so that at any iteration a new estimation of the solution is calculated. This estimates become closer and closer to the exact solution of the system of equations, and after certain number of iterations a satisfactory solution with respect imposed convergence test is obtained. Popular iterative methods are the Jacobi's method and the method of Gauss-Seidel.

For large systems such those obtain in finite element discretization procedure the iterative methods are more efficient than direct methods. Two major advantages of iterative methods are:

1. They require significantly less amount of storage on a computer than the one required for direct methods if the coefficient matrix of the system is sparse.

2. Can be implemented easier on high-performance computers than direct methods.

Section 2.4.2 gives details about numerical techniques used to carry out the simulations associated with the research programme.

2.3. Computational Fluid Dynamics (CFD)

This section outlines briefly the partial differential equations governing the transport physical phenomena (e.g. heat, flow, etc) and the discretization procedures using finite volume techniques.

2.3.1. *Continuity Equation*

The equation that needs to be satisfied for mass conservation is known as the *continuity* equation. The form of the continuity equations is

$$\frac{\partial(\rho)}{\partial t} + \text{div}(\rho \underline{V}) = 0 \qquad (2.19)$$

where ρ stands for the density and $\underline{V}$ stands for the velocity vector. The continuity equation is solved in addition to the equation governing the transportation of the solved quantity.

2.3.2. *Momentum Equations*

In the three-dimensional Cartesian co-ordinate system the partial differential equations that are governing the law of the momentum conservation for a fluid flow are:

$$\frac{\partial(\rho u_i)}{\partial t} + \text{div}(\rho \underline{V} u_i) = \text{div}(\mu \, \text{grad}(u_i)) - \frac{\partial p}{\partial x_i} + S_{u_i} \quad , \quad i \in \{x, y, z\} \qquad (2.20)$$

The nomenclature being used in these equations include: μ stands for the dynamic viscosity, ρ is the density, p is the pressure, $\underline{V}$ is the resultant velocity and u_i, $i \in \{x, y, z\}$ are respectively the Cartesian velocity components in x, y and z direction. The source term, S, with the corresponding subscript to denote the co-ordinate direction contributes to each of the three equations. The source term is dependent on the flow features and can vary from application to application.

- The buoyancy source. The source is added to the momentum equation if the gravity vector is non-zero.

In the case of an incompressible flow, the Boussinesq approximation can be used to represent the buoyancy effect. The source equation in this case is given by

$$S_i = \rho_{ref}\beta(T_{ref} - T)g_i \qquad (2.21)$$

where $i \in \{x, y, z\}$, ρ_{ref} is the reference density, β is the coefficient of the thermal expansion, T_{ref} is the reference temperature and T is the temperature. The components of the gravity vector are represented by g_i ($i \in \{x, y, z\}$). In the case of a compressible fluid, the source term equation becomes

$$S_i = (\rho_{ref} - \rho)g_i \qquad (2.22)$$

where in addition to the above listed nomenclature we have ρ standing for the density material property.

- Darcy Source. The source is added if the solidification phenomena are considered. The source term is equal to

$$-u_i \frac{(1-f)^2}{f^3}K \qquad (2.23)$$

where f is the liquid fraction and K is the permeability coefficient.

2.3.3. Heat Transfer Conservation Equation

The equation governing the heat transfer phenomenon can be expressed as

$$\frac{\partial(\rho c_p T)}{\partial t} + \mathrm{div}(\rho c_p \underline{V}\, T) = \mathrm{div}(k\,\mathrm{grad}(T)) \qquad (2.24)$$

where ρ is the density, c_p is the specific heat, k is the thermal conductivity and the conserved variable is the temperature T. If the temperature is the solved variable and in the

case of low speed flow and no phase change process there is no source term in Equation 2.24. If solidification effects are analyzed then source terms that account for the heat energy release and convective transfer of the liquid must be included in the right side of Equation 2.24 [118]. If the heat equation is expressed in terms of the enthalpy rather than temperature then a source term is also added [118].

2.3.4. *General Conservation Equation*

The generalized governing equation for a conserved quantity ϕ can be expressed mathematically by the following differential equation

$$\frac{\partial(C_T\phi)}{\partial t} + \text{div}(C_C\underline{V}\,\phi) = \text{div}(\Gamma_\phi\,\text{grad}(\phi)) + S_\phi \qquad (2.25)$$

In the conservation equation C_T and C_C denote the transient coefficient and the convection coefficient respectively, Γ_ϕ is the diffusion coefficient and S_ϕ stands for the source term.

Each of the terms in the general conservation equation represents a different phenomenon or aspect of quantity transportation. The first term on the left side of Equation 2.25 is the transient term and the second is the convection term. On the right side of the equation, we have the diffusion term and the source term respectively. Depending on the nature of the analyzed process, all or just some of these terms will exist in the equation to govern the corresponding physical phenomenon.

Apart from the mentioned dependent variables such as the velocity components in the momentum equation or the temperature in the heat transfer equation, ϕ could stand for any other physical quantity that obeys the transportation low. Examples may vary from mixture fraction to turbulence kinetic energy to concentration.

2.3.5. *Finite Volume Discretization Techniques*

The partial differential equation of the general low of governing the time dependent three-dimensional flow of a physical quantity is used in the general finite volume computational

procedures. In the finite volume approach, the transport equation is integrated over a three-dimensional control volume V :

$$\int_V \frac{\partial(C_T \phi)}{\partial t} dV + \int_V \text{div}(C_C \underline{V} \phi) dV = \int_V \text{div}(\Gamma_\phi \text{ grad}(\phi)) dV + \int_V S_\phi dV \qquad (2.26)$$

Using the Gauss' divergence theorem, the integral form of equation can be re-written as

$$\frac{\partial}{\partial t}\left(\int_V C_T \phi \, dV \right) + \int_A \mathbf{n} . (C_C \underline{V} \phi) dA = \int_A \mathbf{n} . (\Gamma_\phi \text{ grad}(\phi)) dA + \int_V S_\phi dV \qquad (2.27)$$

where A is the entire bounding surface of the integrated volume V .

In the general case of a time-dependent problem, it is also necessary to integrate with respect the time t over a small time interval Δt . The final form of the integrated transport equation is:

$$\int_{\Delta t} \frac{\partial}{\partial t}\left(\int_V C_T \phi \, dV \right) dt + \int_{\Delta t} \int_A \mathbf{n} . (C_C \underline{V} \phi) dA \, dt = $$
$$\int_{\Delta t} \int_A \mathbf{n} . (\Gamma_\phi \text{ grad}(\phi)) dA \, dt + \int_{\Delta t} \int_V S_\phi \, dV dt \qquad (2.28)$$

The discretized form of each of the terms using finite volume techniques follows:

- **The transient term** : In the case of stationary mesh the transient term is discretized in the following manner

$$\int_{t-\Delta t}^{t} \frac{\partial}{\partial t}\left(\int_V C_T \phi \, dV \right) dt = \left((C_T)_P \phi_P - (C_T^0)_P \phi_P^0 \right) V_p \qquad (2.29)$$

In this discretization, the subscript " P " stands for the centre of the control volume and denotes a value at that centre which is taken as an approximation to the average value for the control volume V . The superscript " ^{0} " indicates the values used or obtained through that equation at the previous time step. Since the time integration of the other terms in the transport equation leads only to a multiplication of factor Δt to their discretized forms,

the whole equation can be divided by Δt. Thus, the time step contributes only to the transient term. After this transformation the transient term becomes

$$\frac{(C_T)_P \phi_P - (C_T^0)_P \phi_P^0}{\Delta t} V_P$$

Above discretization assumes no changes in the volume of the control volume V with time. The rest of the terms are staying independent by the time step. The discretization of the other terms given below follows the above made assumption, i.e. the integration with respect time will be ignored.

- **The diffusion term** : The diffusion term from originally a volume integral is replaced, as shown above, into a surface integral

$$\int_A n \cdot (\Gamma_\phi \; \text{grad}(\phi)) \, dA \qquad (2.30)$$

Since the surface of a polyhedral control volume consists of a set of surfaces, the diffusion term can be expressed in the form of a sum of integrals over each face that bounds the control volume:

$$\sum_f \int_f \Gamma_\phi \frac{\partial \phi}{\partial n} \, dA = \sum_f (\Gamma_\phi)_f \left(\frac{\phi_A - \phi_P}{d_{AP}} \right) A_f \qquad (2.31)$$

The discretization technique assumes orthogonal mesh and the normal gradient at the face f connecting two adjacent elements (centres denoted respectively with subscripts A and P) is approximated using the distance d_{AP} between their centroids. The area of the surface f is denoted as A_f and $(\Gamma_\phi)_f$ is the diffusion coefficient at the surface boundary between the two adjacent elements. The value of the diffusion coefficient at the boundary face can be evaluated either by arithmetic mean

$$(\Gamma_\phi)_f = \alpha_f (\Gamma_\phi)_P + (1 - \alpha_f)(\Gamma_\phi)_A \qquad (2.32)$$

or by using the harmonic mean expression

$$(\Gamma_\phi)_f = \frac{(\Gamma_\phi)_P (\Gamma_\phi)_A}{\alpha_f (\Gamma_\phi)_P + (1-\alpha_f)(\Gamma_\phi)_A} \tag{2.33}$$

where $\alpha_f = \dfrac{d_{Af}}{d_{Af} + d_{fP}}$, and d_{Af} and d_{fP} are the distances respectively from the element centre A to the element face f and from the adjacent element centre B to the same face f connecting both elements.

- **The convection term:** Similarly, to the diffusion term, the convection term in its integral form is initially transformed using the divergence theorem from a volume to a surface integral. The subscript " f " denotes again the value at the face bounding the control volume.

$$\int_A n \cdot (C_C \underline{V} \phi) dA = \sum_f (C_C)_f (\underline{V} \cdot n) A_f \, \phi_f \tag{2.34}$$

The upwind element gives the value of the convection coefficient. The mathematical expressions state:

$$\begin{aligned}
(C_C)_f &= (C_C)_P & \text{if} \quad (\underline{V} \cdot n) &\geq 0.0 \\
(C_C)_f &= (C_C)_A & \text{if} \quad (\underline{V} \cdot n) &< 0.0
\end{aligned} \tag{2.35}$$

In the discretization procedure, the calculations of the normal velocity components at the face $(\underline{V} \cdot n)_f$ can be based on the Rhie-Chow interpolation technique [133]. If the arithmetic mean is used for estimation of the value of the solved variable at the face, the final form of the discretized convection term becomes

$$\sum_f (C_C)_f (\underline{V} \cdot n)_f A_f \left\{ \alpha_f \phi_P + (1-\alpha_f)\phi_A \right\} \tag{2.36}$$

- **The source term:** The source term S_ϕ can be represented initially in a linear form as it is discussed by Patankar [130]:

$$S_\phi = S_C + S_P \phi \tag{2.37}$$

The equations that specify S_C and S_P have significant impact on the rate (speed) of convergence and the stability of solution. The discretized form of the linearised source term is

$$\int_V S_\phi \, dV = (S_C + S_P \phi_P) V_P \tag{2.38}$$

After the integration over the control volume, all terms in the discretized form are calculated at centre of the control volume. A detailed description of the overall finite volume procedure (the physical transport phenomena, governing equations, discretization and the calculation procedure) can be found in references [130, 131, 134, 103, and 135] or other CFD related literature.

2.4. PHYSICA Software Framework

2.4.1. *PHYSICA – a Multiphysics Analysis Tool*

In this research and related applications from the area of electronic packaging, the software package used as a simulation tool is *PHYSICA* [99, 118]. *PHYSICA* is a software environment for finite volume modelling of multi-physics phenomena. The analysis tool is a collection of interacting solution procedures, including CFD finite volume based techniques [101, 103, 130, 131, 135] for heat transfer and fluid flow and FE based elasto-visco-plastic solid mechanics [106, 109]. Brief discussion on these numerical techniques, mathematical methods and the FE/FV discretization algorithms was provided in the previous sections of the chapter.

Analysis tool *PHYSICA* consists of three major modules:

- **The pre-processor.** *PHYSICA* is linked to a software package, FEMGV [119] , used as a tool for geometry and mesh generation. The major steps performed using the pre-processor are:

    1. Define the geometry in a computational form;

    2. Generate the mesh (nodes and elements) across the model;

3. Apply the boundary patches;

4. Define elements with same material and physical properties.

- **The analysis modules (the solver):** *PHYSICA* has a variety of modules and interacting routines that allows the user to solve simultaneously different physical phenomena (multi-physics analysis). The relevant mathematical equations are discretized and solved using finite volume techniques and the analysis input specifications (i.e. the material properties, type of analysis, etc.). Finite element based module for stress analysis is also available. *PHYSICA* is an open framework that allows the user to write and implement his own code as well as to obtain internally any type of the analysis data and/or results. The modules have the capability for an analysis restart from the previously recorded in the database results.

- **The post-processor:** PHYSICA is integrated with the post-processing software package FEMGV for post-processing and visualization of the analysis results. Some of the options for presenting the result data include vector plots of the flow, plots of displaced shape of the design, contour plots of the solved quantity, etc.

2.4.2. *Analysis Procedure using PHYSICA*

As described earlier, the solid mechanics or the fluid dynamics mathematical analysis relies on the set of partial differential equations that govern the physics of the phenomenon and describe the system behaviour. The process of a numerical analysis includes the following major steps:

1. The **initial stage** includes a detailed observation and investigation of the problem. Aspects of the problem such as the geometrical shape of the design, the system or process characteristics subject to specified conditions and the material selection are all important from modelling point of view;

2. The **geometry and mesh generation** is the most time consuming step. At this step, the geometry of the investigated system is build. Then the geometry is discretized to the form of mesh elements and nodes (the model is split into small sub-regions);

3. The **boundary conditions** and **material properties** specification, used to define the analysis inputs, is the next step in this procedure. The boundary conditions may specify, for example, mechanical forces or thermal loads, known displacements at given nodes, velocity and pressure at inlets/outlets for a flow domain, and etc.;

4. The **solution of the global system** of the discretized governing equations provides the approximate numerical solution for the solved quantity. Different numerical solvers based on iterative methods are available in *PHYSICA*: the over relaxed forms of the Jacobi method (JOR), Gauss Seidel (SOR), the Jacobi preconditioned conjugate gradient method (JCG), the bi-conjugate gradient (BICG) and the bi-conjugate gradient stabilized method (BICGSTAB). Reference [118] provides details on the mathematical algorithms associated with the above solvers. The bi-conjugate gradient method was used to carry out the calculations in the undertaken studies outlined in the thesis.

5. **The analysis of the results** is the last step in this process.

CHAPTER 3

DOE STATISTICAL ANALYSIS AND RSM

Modern semiconductor devices are characterised with dependencies which in many cases are very difficult to predict and to observe explicitly. A way to analyze the characteristics of an electronic package, device or process is to evaluate the statistical variations of a set of selected parameters called *factors* (statistical terminology) or *design variables* (optimisation terminology) and to examine their results. In general, the experiments are designed to show the dependencies between the input and output parameters based on a small number of observations. Computational modelling can benefit the evaluation of each set of design variables in the same way as the real experiments can.

Statistical methods for the generation of the parameter values for several designs are known as *Design of Experiments* (DoE) methods [139,149]. The real experiment or computer analysis results of interest are called *response variables* (or simply *responses*) and can be used to fit to them an approximated analytical model. Such an analytical model is known as a *Response Surface* (RS) approximation, and the related concepts are known as the *Response Surface Methodology* (RSM) [142,150]. Historically, the theory of response surfaces and experimental designs began in the chemical industry in the 50's, when statisticians and chemical engineers started to collaborate in their work.

The DoE and RS methods can be adapted and applied to many sectors of industry and science. This research programme uses the DoE and RS modelling techniques coupled with computational mechanics tools. The methodology is part of the developed strategy for identifying the optimal design and process conditions in the microelectronics application area. This chapter focuses on the main concepts in the DoE and RSM theory.

3.1. Introduction to Design of Experiments (DoE) & Response Surface (RS) Modelling

The purpose of running experiments is to characterise unknown relations and dependencies that exist in the observed design or process, i.e. how this design or process is influenced by a set of chosen design variables and how it will respond to variations in the design variable values. For example, in electronic packaging it is important to know how the life-time and the reliability of the package are influenced when materials with different properties are used. Since the exact and explicit relationship between the design variables and the design responses is impossible to be obtained, a possible strategy to resolve the problem is to approximate the underlying relationship with an empirical model (Response Surface, or RS) of the form:

$$Y_i = f(X_1, X_2, ..., X_n) \qquad (3.1)$$

where X_1, X_2, ..., X_n are the design variables and Y_i is the system response of interest. Based on the mathematical formulation, a set of design variables can be considered as a design point in the n-dimensional design space, $X = (X_1, X_2, ..., X_n)$, where n is the total number of the design variables. The $f(X)$ analytical model is the curve fit of interest, i.e. the Response Surface (RS) that approximates the Y_i response.

Response Surface Methodology (RSM) is an approach for creating Equation 3.1, an explicit approximate function of a system response using experimental designs and numerical approximation techniques. These approximations (the RS models) can then be used to investigate and evaluate inexpensively different designs and to perform optimisation with no further experiments. When the investigated system is specified with a relatively small number of design variables and the cost of analyzing a single design point is high, the RSM can be

used to evaluate efficiently different design configurations and to identify the design parameter relationships. The RS models are typically low-order polynomials (linear or quadratic) constructed by fitting the design of experiments data using the least squares regression techniques.

The automated design optimisation of a system or process can also benefit from RS models as it will be discussed in Chapter 4.

The experimental design data, in terms of sets of pairs (*design point, response value*), used to construct a Response Surface model, is important for the accuracy of approximation and has influence on the efficiency of the entire response surface modelling. A random selection of design points used to obtain values for the analysed response may result in a set of data, which – if used to build a RS model - may produce response approximation with bad quality. The random selection of design points will require evaluation of large number of design points to ensure the accuracy of RS model. This is not a practical approach as it will require a significant computational and other expense.

The strategy of changing only one design variable at a time is also not desirable because important interactions between the design variables can be missed. Thus, a proper selection of points in the design space that will ensure representative and high quality RS approximation is essential.

In the Design of Experiments methods, the system behaviour is observed using only a small number of experimental points [139] distributed in a certain way across the design space.

Response Surface Modelling using Design of Experiments proceeds in a number of steps as illustrated in Figure 3.1. The general purpose of DoE combined with Response Surface techniques is to:

1. **Create a set of experimental design points.**

   During the planning stage of a set of experiments, the ranges in which the input design variables (the factors) can change are defined. The *layout* of the experimental design specifies how the sample points are chosen in the selected design subspace of the input design variables. Different algorithms - the experimental design methods -

can be used to identify the suitable sets of design variable values. These methods are outlined in Sections 3.2, 3.3 and 3.4.
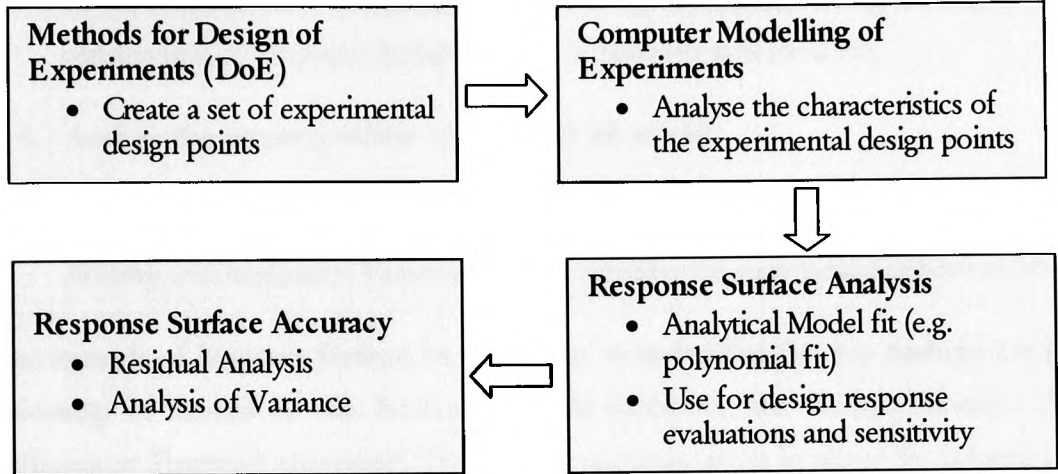
Figure 3.1: Response Surface Modelling Procedure

2. **Analyze the characteristics of these design points.**

Analysis of design points requires the evaluation of the actual design response at each of these points via experiments or computer analysis. When the DoE is used within the computational modelling, each of the experimental points is evaluated by using the simulation (e.g. FEA, see Chapter 2) outputs. The simulations instead of real experiments provide reproducible results, so one observation does not have to be repeated.

3. **Fitting a RS model to the experimental design response data (e.g. linear or polynomial fit).**

The data for experimental points and the corresponding design response values is used to build an approximation to design response of interest (Y), e.g. a second order polynomial:

$$Y(X) = Y(X_1, X_2, ..., X_n) = a_0 + \sum_{i=1}^{n} a_i X_i + \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} X_i X_j$$

In the above polynomial form, $X = (X_1, X_2, ..., X_n)$ is the vector of design variables, n is the number of design variables and a_0, a_i, a_{ij} $(i = 1, n$ and $j = 1, n)$ are the coefficients of polynomial obtained using least squares fitting techniques. The construction of Response Surface Models is discussed in Section 3.5.

4. **Analyse the accuracy of the constructed RS model.**

The accuracy of Response Surface model can be observed by performing Residual Analysis and Analysis of Variance. These techniques are summarized in Section 3.6.

As an example of Response Surface, let us observe an underfilled flip-chip package. Let the relationship of interest be that between underfill properties, the Young's Modulus (E), Coefficient of Thermal Expansion (CTE), and the maximum stress in silicon die induced as a result of an applied thermal load. In this case we have two design variables, X_1 (the underfill Young's Modulus) and X_2 (the underfill CTE), and the pair of these two design variables (X_1, X_2) is a design point in the 2-dimensional underfill E-CTE design space. The flip-chip design response Y_1 is the maximum stress in silicon die. For this example, a second order polynomial that can serve as a Response Surface model will have the form

$$Y_1 = f(X_1, X_2) = a_0 + a_1 X_1 + a_2 X_2 + b_1 X_1^2 + b_2 X_2^2 + c_1 X_1 X_2$$

where a_0, a_1, a_2, b_1, b_2 and c_1 are real value constants which are determined in a way to fit the available data for maximum stress in the die for different sets of design variables (i.e. different design points).

3.2. Design Variables and Responses in DoE

3.2.1. *Design Variables Normalization (scaling)*

In the DoE, it is convenient to scale the design variables with respect their upper and lower limits. The usual way of performing this scaling is to normalize the design variables to the range [-1, 1]:

$$X_i = \frac{z_i - (\max[z_i] + \min[z_i])/2}{(\max[z_i] - \min[z_i])/2} \quad , \quad i = 1, 2, ..., n \tag{3.2}$$

In Equation 3.2, z_i is the un-scaled design variable, X_i is the scaled design variable and n is the number of design variables. Based on this normalization, each of the scaled design variables has lower limit equal to -1 and upper limit equal to 1.

For example, if for a flip-chip package the underfill Young's Modulus (E) is a design variable and the range of its variation is between 1GPa and 6GPa, then the scaled value of Young's Modulus X_1 over [-1, 1] can be calculated using the un-scaled value z_1 and the relation

$$X_1 = \frac{z_1 - (6+1)/2}{(6-1)/2} = \frac{z_1 - 3.5}{2.5} \quad [\text{GPa}]$$

Based on the above transformation, for example, un-scaled value of E = 2 GPa is equivalent to scaled value of -0.6, and un-scaled value of 4.5 GPa gives scaled value of 0.4.

The experimental designs discussed in this chapter assume normalised design variables using the transformation of Equation 3.2.

3.2.2. Design Matrix

An experimental design is represented in a matrix form, called *design* matrix D. The number of the columns in D is equal to the number of the design variables (n) and the number of the rows is equal to the number of the experimental points (m). Thus, each of the rows in the design matrix D represents a particular configuration of the design variables that relates to this experimental point in the design space. The design matrix is discussed further in Section 3.5.1.

3.2.3. Model Matrix

The *model* matrix M relates to the different efficiency measures of the experimental design if a polynomial RS model is used. The model matrix M, similarly to the design matrix D, has as many rows as the number of the experimental points (m). The number of columns is equal to the number of the terms in the polynomial RS model (p). Each of the rows of the model

matrix M represents a particular combination of the polynomial RS terms (e.g. linear, mixed, quadratic, etc.) for the corresponding design point. More details follow in Section 3.5.1.

3.3. Methods for Obtaining Experimental Design Points

The experimental designs primarily considered and used as a part of the developed optimisation modelling framework are summarized in Table 3.1 [139-143, 149].

Table 3.1: Experimental Designs

Experimental Design			
1.	Screening Analysis	7.	Three-level Factorial
2.	Full Factorial	8.	Box-Behnken
3.	Central Composite Face-centred	9.	Notz
4.	Latin Hypercube	10.	Koshal
5.	Plackett-Burman	11.	Random
6.	Orthogonal Array	12.	D-optimal

In the following section, n denotes the number of design variables given to an experimental design. For each of the design variables, the minimum and maximum values are specified. These limits are used to obtain the n-dimensional design subspace (n-dimensional hypercube) for the constructed experimental design. The review of the following DoE methods utilizes the term *factor* instead of *design variable* as being the most accepted in statistics. However, from the design optimisation point of view it will be more convenient to refer to factors as design variables throughout the rest of the thesis.

3.3.1. *Screening Analysis*

In the DoE, the *nominal design* is an experimental point at the centre of the n-dimensional design subspace, i.e. all factors have value zero on the normalized [-1, 1] interval design subspace. The *Screening Analysis* is based on $2n$ experimental points in total. A design point created using this method is specified by giving one of the factors its minimum or maximum value and all other factors are fixed to their nominal values. The design points

constructed using the above rule are called *axial* points. The DoE *Screening Analysis* in the three dimensional design space is illustrated in Figure 3.2.
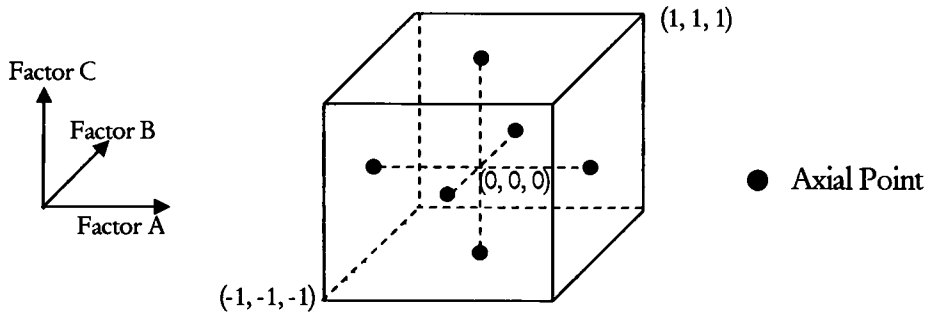
Figure 3.2: DoE Screening Analysis

3.3.2. *Full Factorial Design*

The experimental design points in a full factorial design are the vertices of a hyper cube in the n-dimensional design space defined by the minimum and the maximum values of each of the factors. These experimental points are also called *factorial* points. The *Full Factorial* method creates 2^n experimental points and these include all possible combinations of the minimum and maximum values of the factors. The experimental points in a factorial design can be determined by perturbing all factors simultaneously by 1 (if factors scaled in [-1, 1]) from the nominal values (the centre of the multi-dimensional box, 0 if scaled factors). Because each of the factors can take only two values (minimum or maximum in the range), the full factorial design belongs to the *two-level* class of experimental designs.

The *Full Factorial* design in the three-dimensional design space is illustrated in Figure 3.3.
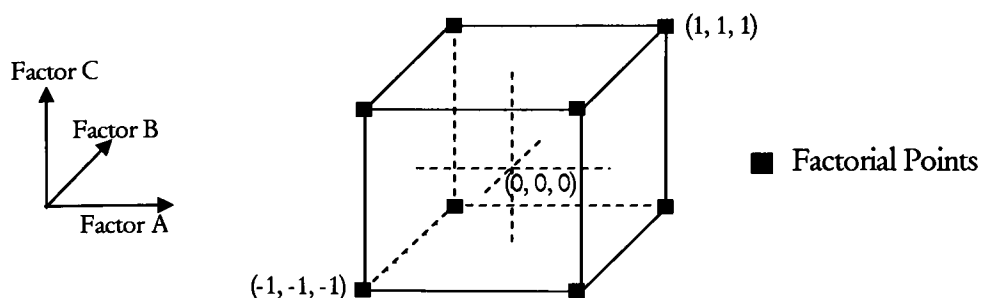
Figure 3.3: *Full Factorial* Design with Three Factors

In *Fractional Factorial* design, only a subset of all vertices of the multidimensional box representing the design space is used. Such an experimental design is denoted with the number of the points in the design, $2^{(n-k)}$, where k is an integer number smaller than the total number of factors n.

3.3.3. *Central Composite Design (CCD)*

This experimental design is frequently used to fit the available response data to a second order polynomial RS models. To establish the coefficients in a polynomial with quadratic terms, the experimental design must have at least three levels for each factor. If the n-dimensional hyper-cube defined by the factor's limits is placed at the origin of the coordinate system then the CCD consists of three different types of points:

1. **Factorial points.**

   These are the vertices of the n-dimensional cube (factorial portion). Either 2^n points coming from the *Full Factorial* design (Section 3.3.2) can be used or only a fraction of the factorial design;

2. **Central point.**

   A single point at the centre of the design space (the nominal design);

3. **Axial points.**

   The number of axial points is $2n$. These points are located on the axes of the coordinate system and symmetrically with respect the central point (created by a *Screening Analysis*, see Section 3.3.1). If normalised factors are observed, i.e. each factor varies in [-1, 1], then the axial points are typically located at distance from 1 to $\sqrt{n}$ from the central point (n is the total number of factors). Axial points at distance >1 are not within the specified design subspace, and should be used only if the corresponding un-scaled design points are still feasible and can be analysed.

Example of a three dimensional *Central Composite* design is shown in Figure 3.4.
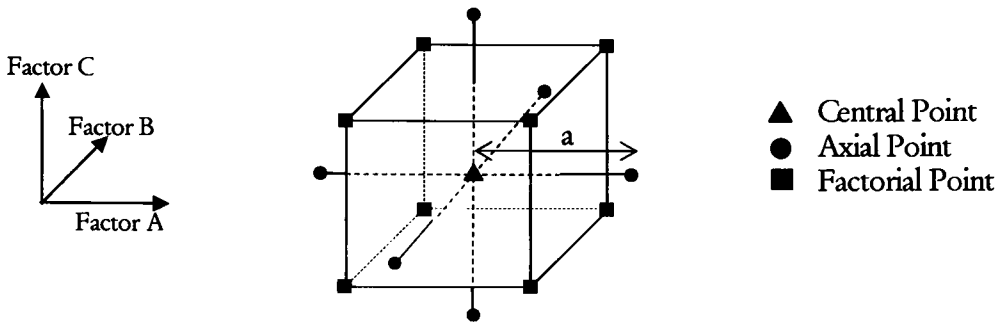
Figure 3.4: *Central Composite* $(a = \sqrt{n})$ DoE with Three Factors

The CCD is an effective experimental design if the number of factors is very small. Otherwise, the CCD can be replaced by the so-called *Small Composite* Design (SCD). The SCD can be obtained from the CCD by replacing the full factorial portion of the design with a fractional factorial design.

If quadratic polynomial is used as RS model, then the design points in the central composite design contribute in different ways to the terms in this RS quadratic approximation. The factorial points have effect mainly on estimating the linear terms (coefficients) and the interactions between two factors. The axial points and the central point contribute in a major way to determine the quadratic terms in the RS model.

3.3.4. *Latin Hypercube* Design

The *Latin Hypercube* Design defines the experimental points according to the specified number of points in the design space m.

The algorithm consists of two loops:

1. For each of the n factors the range of the factor is divided into m non-overlapping intervals on the basis of equal probability;

2. From each interval one value is selected randomly with respect to the probability density in the interval.

The *m* values obtained for the first factor are combined in a random manner (equally likely combinations) with the *m* values for the second factor. These *m* pairs are combined again in a random manner with the *m* values for the third factor, thus obtaining *m* triplets, and so on, until *m* *n*-tuplets are specified. These *m* *n*-tuplets are the experimental points of the *Latin Hypercube* design.

A common rule is to use the uniform probability distribution rule to form the *Latin Hypercube* designs. In Table 3.2, an example of how four points (Points 1 to 4) can be arranged in the *Latin Hypercube* design in four dimensions (Factor F1 to F4) with ranges [-1, 1] is shown.

Table 3.2: *Latin Hypercube* Design

	Factor F1	Factor F2	Factor F3	Factor F4
Point 1	0.5669910	-0.2465637	-0.3702434	-0.7918264
Point 2	-0.8699596	-0.7146524	0.1408818	0.8313969
Point 3	-0.2900985	0.4635731	0.5991107	0.09391561
Point 4	0.09718282	0.6744157	-0.7708488	-0.09295038

The values of factors F1 and F2 are shown in Figure 3.5. The small boxes represent the subsections of the factor ranges for F1 and F2, and the numbers represent the experimental points for evaluation located in the sub-region and corresponding to the enumeration used in Table 3.2.
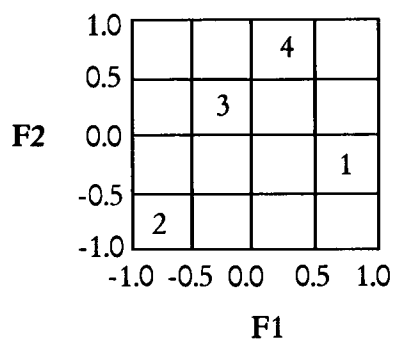
Figure 3.5: Positions of the Four Design Points in the Subspace Spanned by F1 and F2

3.3.5. *Plackett-Burman* Design

The *Plackett-Burman* design is a two-level design for studying n factors using $k = n+1$ experimental points. The *Plackett-Burman* design is limited and can be used only if the k is a multiple of 4. Plackett and Burman [140] supply signs for the first row of the design matrix as shown in Table 3.3. The "-" sign corresponds to the minimum value and the "+" sign corresponds to the maximum value specified by the range of the factor. All other rows of the design are built by shifting the previous one. An extra row is added to the design consisting of the minimum values of all parameters.

If k is a power of 2 the design is equivalent to a fractional factorial design, but for the other numbers (e.g. 12, 20, 24, 28, etc) the design structure provided by Plackett and Burman must be followed.

Table 3.3: The First Row for *Plackett-Burman* Designs

n	k	String
11	12	+ + - + + + - - - + -
15	16	+ + + + - + - + + - - + - - -
19	20	+ + - - + + + + - + - + - - - - + + -
23	24	+ + + + + - + - + + - - + + - - + - + - - - -

In Table 3.4 the *Plackett-Burman* design for 11 factors (A to K) and 12 experimental points is shown.

Table 3.4: The *Plackett-Burman* Design for 11 Factors (named *A* to *K*).

	A	B	C	D	E	F	G	H	I	J	K
1	+	+		+	+	+	-	-	-	+	
2		+	+		+	+	+				+
3	+		+	+		+	+	+	-	-	
4		+	-	+	+	-	+	+	+	-	-
5	-	-	+		+	+		+	+	+	
6				+	-	+	+		+	+	+
7	+			-	+	-	+	+		+	+
8	+	+		-		+	-	+	+		+
9	+	+	+			-	+		+	+	
10		+	+	+	-	-	-	+	-	+	+
11	+		+	+	+	-			+		+
12	-	-	-		-						

The *Plackett-Burman* design can help to investigate only the main effects of the factors by using almost the minimum possible number of experimental points. The design can be used also as a factorial portion inside the small composite design.

3.3.6. *Orthogonal Array* Design

Orthogonal Arrays are special matrices used as the design matrices in the fractional factorial design. These arrays can provide estimation of the effect of several factors in a highly efficient way[141].

The distribution of the experimental points is tabulated in tables and can be found in reference [139].

In Table 3.5, the design matrix of a two-level *Orthogonal array* design with three factors is shown assuming the normalised range [-1, 1] for the factors.

Table 3.5: 2^3 *Orthogonal Array* Design

	F1	F2	F3
1	1	1	-1
2	1	-1	1
3	-1	1	1
4	-1	-1	-1

The three-level orthogonal arrays have matrices where each of the factors has exactly three levels represented in the matrix columns. For any pair of columns in the design matrix, all combinations of the factor levels appear equal number of times. The three-level Taguchi orthogonal arrays can be based on 9 (L9), 27 (L27) or 81 (L81) experimental points. The experimental design L9 for three factors is given in Table 3.6. This design aids the investigation of up to 4 factors and their interactions.

Table 3.6: L9 *Orthogonal Array* Design Matrix

	F1	*F2*	*F3*
1	-1	-1	-1
2	-1	0	0
3	-1	1	1
4	0	-1	0
5	0	0	1
6	0	1	-1
7	1	-1	1
8	1	0	-1
9	1	1	0

3.3.7. *Three-Level Factorial* Design

In the *Three-Level Factorial* design, all possible combinations of three discrete values of the factor are used. The number of the design points in the higher order *Factorial* design grows rapidly and makes them impractical in the case where more than a few factors are investigated.

Table 3.7 provides a comparison between some of the experimental designs outlined in the previous sections. The number of the required design points can be reduced by skipping some of the higher order interactions between the factors.

Table 3.7: Comparison of the Number of Experimental Points for Different DoE

Number of factors	Full Factorial design	Screening Analysis design	Central Composite design	Three-level Full Factorial design
2	4	4	9	9
3	8	6	15	27
4	16	8	25	81
5	32	10	43	243
6	64	12	77	729
7	128	14	143	2187
8	256	16	273	6561
n	2^n	$2n$	$1+2n+2^n$	3^n

3.3.8. *Box-Behnken* Design

As discussed in Section 3.3.2, the experimental points in a factorial design can be determined by perturbing all the factors simultaneously by 1 (factors scaled in [-1,1]) from their nominal values. If the experimental points are obtained using the same rule but instead of perturbing simultaneously all factors, now only one, two, or three factors are perturbed at a time and the rest of the factors are kept to their nominal values, so-called *Box-Behnken* design will be constructed.

Table 3.8: Box-Behnken Design for Three Factors

	F1	*F2*	*F3*
1	1	1	0
2	1	-1	0
3	-1	1	0
4	-1	-1	0
5	1	0	1
6	1	0	-1
7	-1	0	1
8	-1	0	-1
9	0	1	1
10	0	1	-1
11	0	-1	1
12	0	-1	-1

An example of the scaled Box-Behnken design for three factors in total is given in Table 3.8. Two of the factors are perturbed at a time and the third stays at the nominal (zero) value.

3.3.9. *Notz* Design

The experimental points in the *Notz* design, similarly to the composite design, contain both the axial points and the factorial portion. The factorial portion consists from the full factorial design minus the point obtained by perturbing all factors by +1 in terms of scaled factors. The axial points are one factor at a time points perturbed by +1 in terms of scaled variables.

The *Notz* design is suitable to fit a second order RS model to the responses at the experimental points. An example of *Notz* design with three factors is given in Table 3.9. It is

clear that the factorial portion consists of 7 points from a 2^3 full factorial design, and the other three experimental points contributing to the design are the axial points.

Table 3.9: Notz design with three factors

	F1	F2	F3
1	1	-1	1
2	1	1	-1
3	1	-1	-1
4	-1	1	1
5	-1	1	-1
6	-1	-1	1
7	-1	-1	-1
8	1	0	0
9	0	1	0
10	0	0	1

3.3.10. *Koshal* Design

The *Koshal* design contains relatively small number of experimental points and can be used to construct different order RS models. If first order RS model is required then the Koshal design is simply one-factor-at-a-time design plus the central point. For example, for three factors and first order polynomial the Koshal design matrix is

	F1	F2	F3
1	0	0	0
2	1	0	0
3	0	1	0
4	0	0	1

To build a first order RS model with interactions, it is necessary to add to the first order design matrix the "interaction rows". The matrix in this case - for three factors only - is

	F1	F2	F3
1	0	0	0
2	1	0	0
3	0	1	0
4	0	0	1
5	1	1	0
6	1	0	1
7	0	1	1

To fit a second-order model three levels of points are required. The "one-factor-at-a-time" rule is used again to add the required experimental points.

The Koshal design matrix for three factors and second order RS polynomial model is

	F1	F2	F3
1	0	0	0
2	1	0	0
3	0	1	0
4	0	0	1
5	1	1	0
6	1	0	1
7	0	1	1
8	-1	0	0
9	0	-1	0
10	0	0	-1

3.3.11. *Random* Design

A *random* design contains simply a specified number of design points generated randomly throughout the design space. The values of the factors at these points are calculated by mapping the result of a random number generator in the range defined by the minimum and maximum values of the factors. The *Random* designs can be used only if a large number of experimental points are acceptable in terms of their evaluation. In general, this experimental design should be avoided since no statistical criterion is used. The random design can be used in conjunction with the *D-Optimal* design.

3.3.12. *D-Optimal* Design

The *D-optimal* designs are selected using the D-efficiency criterion (see Section 3.5.1.1). Different points are ranked quantitatively to identify the "best" one. Thus, a set of points is specified to ensure certain properties in the information matrix $M^T M$ (M is the model matrix) used in the least squares regression techniques. The D-optimal design is the one that maximizes the determinant of the matrix $M^T M$.

3.4. Comparison of Experimental Designs

The experimental designs have different properties and the optimal choice of the DoE method is significantly dependent on the features and the characteristics of the performed real experiment or computer analysis. The choice of the experimental design affects the degree at which the response values at the experimental points are fitted by the RS model.

The *Central Composite* design is used normally to determine the coefficients of a second order polynomial RS model. The advantage is in the small number of required experiments. Since for small and moderate number of factors this design requires acceptable number of experimental points, it is suitable for DoE applications that involve expensive design evaluations.

The three-level Taguchi orthogonal arrays are also very efficient DoE technique and can be applied to construct second order RS models, still evaluating small number of experimental points.

The *Plackett-Burman* designs require low number of experiments but as a two-level design method it is appropriate to generate only linear response surfaces.

For fitting the response to accurate RS models of higher than the second order polynomial, the *Three Level Factorial* and *Latin Hypercube* designs can be used to generate the experimental design matrix.

If the number of the factors is high then the *Random* design may provide the required DoE solution. In this case, the time required to evaluate an experimental point has to be small. Number of CPU's units (rather than using only one) can help DoE procedure, whatever the experimental methods are, if the computer analyses are utilized as the evaluation technique. A major advantage of DoE approach and RSM is the opportunity to distribute the computer simulations between the available CPU's units and to evaluate the experimental points simultaneously. In fact, this strategy can reduce dramatically the time to perform a simulation-based design process. These aspects are discussed in the application chapters of the thesis.

Table 3.10 provides information on some DoE computational modules and software packages with integrated routines for DoE analyses.

Table 3.10: DoE Commercial Packages

Developer	Package	Information on the WEB
SAS Institute	JMP Software	http://www.sas.com/otherprods/jmp/Home.html
Stat-Ease, Inc.	Design-Ease, Design-Expert	http://www.statease.com
ECHIP, Inc.		http://www.echip.com
UMETRICS AB	Modde	http://www.umetrics.com
Nutek		http://www.rkroy.com
MINITAB Inc.	MINITAB	http://www.minitab.com
S-Matrix Corporation	CARD	http://www.s-matrix-corp.com
Multisimplex KB	MultiSimplex	http://www.multisimplex.com
CAMO ASA	Unscrambler, Guideline+	http://www.camo.no/
StatSoft	Statistica	http://www.statsoft.com
Adept Scientific, Plc	DOE-PC IV	http://www.adeptsciense.co.uk/as/products/qands/qasi/doepciv/
RSD Associates	Matrex	http://www.rsd-associates.com
Design Computing	Gendex	http://www.designcomputing.net/gendex
Launsby Consulting	DOE Wisdom	http://www.launsby.com
Process Builder	STRATEGY	http://www.strategy4doe.com

3.5. Response Surface Methodology (RSM)

3.5.1. *Mathematical Background*

The Response Surface Methodology [142] is a technique used to create mathematical models for relationships that exist between one or more responses and a set of design variables. The most widely used model functions as response surface approximations are the first and the second-order polynomial (Equation 3.3)

$$y(x,\beta) = y(x_1, x_2, ..., x_n, a_0, ..., a_{nn}) = a_0 + \sum_{i=1}^{n} a_i x_i + \sum_{i=1}^{n} \sum_{j=i}^{n} a_{ij} x_i x_j \qquad (3.3)$$

where n denotes the number of design variables. The response surface is validated using the experimental data if available or the predicted data from computer simulations.

If m different experimental points (DoE) are used to obtain the response values y then the design variable values for all experimental points are used to build the m x n design matrix $D \in R^{m \times n}$ (see Section 3.2.2) and the m-dimensional vector of the response data (measurements or simulations), $\bar{y} \in R^{m}$:

$$D = \begin{pmatrix} x_1(1) & x_2(1) & \cdots & x_n(1) \\ x_1(2) & x_2(2) & \cdots & x_n(2) \\ \vdots & \vdots & \vdots & \vdots \\ x_1(m) & x_2(m) & \cdots & x_n(m) \end{pmatrix} \quad \text{and} \quad \bar{y} = \begin{pmatrix} y(1) \\ \vdots \\ y(m) \end{pmatrix} \tag{3.4}$$

where the $x_i(k)$ is the value of the i-th design variable at the k-th experimental point and the response value at the k-th experimental point is denoted as $y(k)$, $k=1,...,m$.

If the RS model function is a polynomial with p terms then the model matrix with m x p dimensions representing the model function can be defined from the multiplied values of the design variables related to the polynomial terms. For the second order RS model, this model matrix (see Section 3.2.3) M $(m \times p)$ is

$$M = \begin{pmatrix} 1 & x_1(1) & \cdots & x_n(1) & x_1(1)x_2(1) & \cdots & x_{n-1}(1)x_n(1) & x_1^2(1) & \cdots & x_n^2(1) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_1(m) & \cdots & x_n(m) & x_1(m)x_2(m) & \cdots & x_{n-1}(m)x_n(m) & x_1^2(m) & \cdots & x_n^2(m) \end{pmatrix} \tag{3.5}$$

The number of the experimental points (m) must be equal or greater than the dimension p of the RS model terms (i.e. the number of polynomial coefficients, components of the vector β). The construction of the RS approximation requires the unknown RS coefficients to be determined, i.e. the vector of regression coefficients β to be calculated,

$$\beta = (a_0, a_1, ..., a_n, a_{12}, ... a_{nn})^T \in R^p \tag{3.6}$$

Equation 3.3 and the model matrix with the vector of the RS terms for all experimental points are used to obtain the following matrix equation:

$$\bar{y} = M\beta$$

where $\bar{\hat{y}}$ is the vector of the approximated response values using the RS model. In the general case ($m>p$) there is no vector of regression coefficients β for which the polynomial function in Equation 3.3 matches the exact numerical responses obtained through the experiments or simulations. The error $\bar{e}$ between the real response values (from the experimental measurements or simulations) $\bar{y}$ and the approximated response values $\bar{\hat{y}}$ from the RS model is defined as

$$\bar{y} = \bar{\hat{y}} + \bar{e} = M\beta + \bar{e} \tag{3.7}$$

3.5.2. Illustrative Example for DoE and Response Surface Model

Consider the following design problem: using the underfill Young's Modulus and underfill CTE, construct a RS approximation to the thermal stress in silicon die of an underfilled flip-chip package subject to thermal load. The design task can be formulated as a DoE-RS problem if we define two design variables – the underfill Young's Modulus E (x_1) and the underfill CTE (x_2). The system response of interest is the value of the maximum stress in silicon die (y), and this value will be estimated based on stress analysis of flip chip package under thermal load. Each of the design variables is given some lower and upper limits that represent the feasible design space of underfill materials, e.g. the Young's Modulus is between 2 and 5 GPa, and the CTE is in the range 50 to 90 ppm/°C. For convenience, the two design variables can be scaled over the interval [-1, 1].

The next step is to specify the experimental points using a DoE method. For these design points in the 2-dimensional E-CTE design space finite element analyses are run and the maximum stress in die for each design point (i.e. an underfill material with the specific E and CTE) is evaluated. Figure 3.6 provides an illustrative example of 9 experimental points that correspond to Composite DoE and values of the maximum stress in die.

Once the data for stress values at all experimental points are available, a Response Surface model can be fit to approximate these stress responses. If linear polynomial is used, then the RS analytical model will have the form

$$y = a_0 + a_1 x_1 + a_2 x_2$$
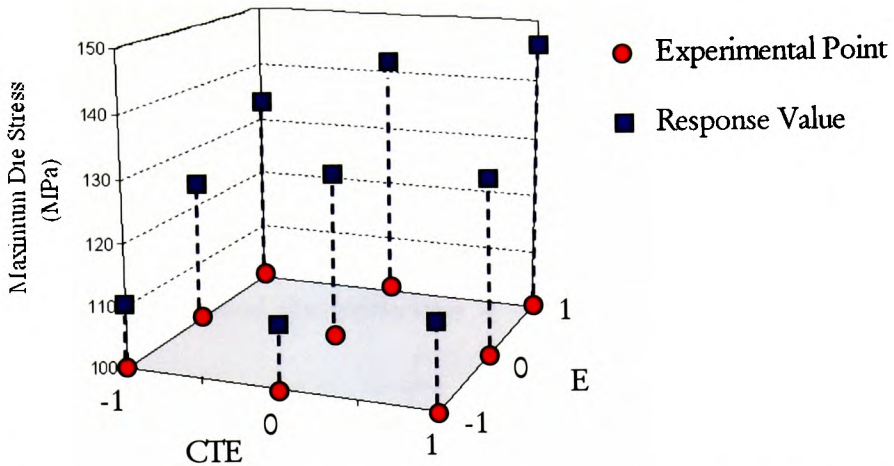
Figure 3.6: Illustrative Example of Composite DoE in E-CTE Underfill Design Space

If second order polynomial with mixed linear terms is chosen as RS, then the fit surface is

$$y = a_0 + a_1 x_1 + a_2 x_2 + a_{12} x_1 x_2 + a_{11} x_1^2 + a_{22} x_2^2$$

The number of coefficients in the above second order polynomial that need to be estimated is six and the number of experimental points is 9. As a result, the RS will not fit exactly the stress values at these points. There will be error in RS value for stress and the actual stress value obtained from finite element analysis. Figure 3.7 illustrates a second order polynomial fit for the above example.
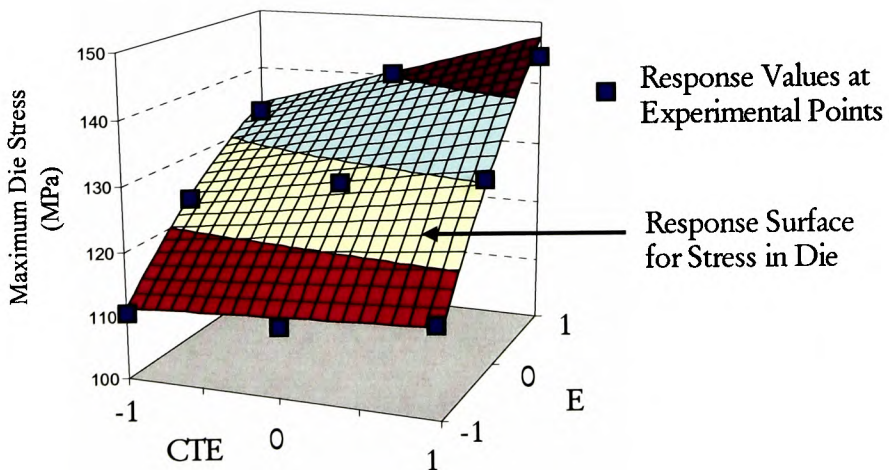
Figure 3.7: Illustrative Example of Second Order Response Surface

The following Section 3.5.3 provides details on the least square technique for estimation of regression coefficients.

3.5.3. *Estimation of Regression Coefficients*

The regression coefficients are determined – based on m experimental points - using the least squares method. In this method, the coefficients β are calculated to give a minimum root-mean-square of the prediction error $e_{mrs} = \sqrt{(\overline{e}^T \overline{e})/m}$.

To perform this calculation it is necessary to find the minimum of the prediction errors, i.e. to solve the problem

$$\min \overline{e}^T \overline{e} \tag{3.8}$$

The problem solution involves calculation of the first derivative of the prediction error and results in the following sequence of transformations:

$$\frac{\partial \overline{e}^T \overline{e}}{d\beta} = \frac{\partial}{\partial \beta}(-M\beta + \overline{y})^T (-M\beta + \overline{y}) = \frac{\partial}{\partial \beta}(\beta^T M^T M\beta - \overline{y}^T M\beta - \beta^T M^T \overline{y} + \overline{y}^T \overline{y}) = \tag{3.9}$$
$$= (M^T M + (M^T M)^T)\beta - (\overline{y}^T M)^T - M^T \overline{y} = 2M^T M\beta - 2M^T \overline{y}$$

Setting Equation 3.9 to zero gives the formula for calculating the unknown RS coefficients:

$$\beta = (M^T M)^{-1} M^T \overline{y} \tag{3.10}$$

The second derivative of the prediction error is

$$\frac{\partial^2 \overline{e}^T \overline{e}}{\partial \beta^T \partial \beta} = M^T M \tag{3.11}$$

If $B = M^T M$ (the matrix has to be non-singular) and $b = M^T \overline{y}$ then the vector of the RS coefficients can be calculated from Equation 3.12 by using any of the existing linear solvers,

$$\beta = B^{-1} b \tag{3.12}$$

3.5.4. *Dimensionality Limitations*

The second order polynomials do not give satisfactory results for the fit of the data values in some cases. The higher order polynomials can also provide the RS model approximation and to improve the fitting accuracy but these functions will need more experimental points. Table 3.11 illustrates the relationship between the polynomial type used to fit the response data and the minimum number of experimental points required to construct an empirical model with n design variables being observed.

Table 3.11: Polynomial Type and Required Minimum Experimental Points

Polynomial Type	Number of Experimental Design Points
Constant	1
Linear	$1 + n$
Pure Quadratic	$1 + 2n$
Quadratic with linear cross-terms	$1 + n + \dfrac{n.(n+1)}{2}$
Pure Cubic with second order cross-terms	$1 + n + \dfrac{n.(n+1)}{2} + n$
Cubic with second order and mixed quadratic-cubic cross-terms	$1 + n + \dfrac{n.(n+1)}{2} + \dfrac{n.(n+1)(n+2)}{2.3}$

Further increase in the polynomial order requires almost an impractical number of evaluations. For a relatively high number of design variables there is no other option but to use low order polynomial, or other type of function, to approximate the system response of interest. However, higher number of experimental points than the minimum required is still essential to benefit from the noise filtration capabilities of the least squares fitting approach.

The graph in Figure 3.8 is often refereed as "the *Curse of Dimensionality*" and shows the rapid growth in the minimum number of the experimental points for an increasing number of design variables.
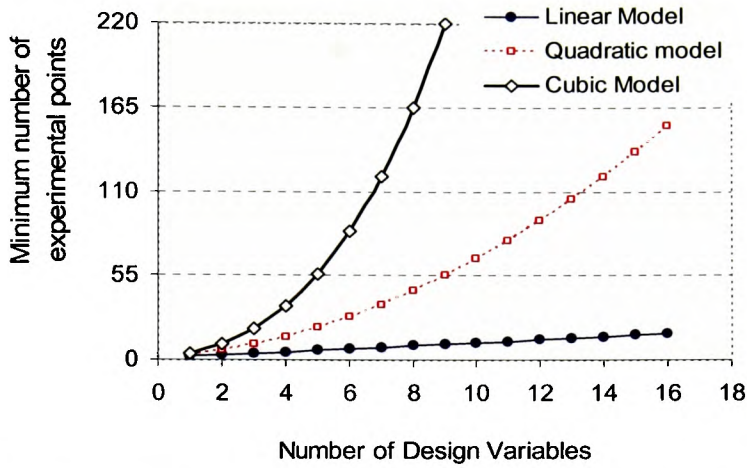
Figure 3.8: Experimental Points Required as Function of Design Variables Number

Table 3.12 provides information on the number of required experimental points as function of the number of design variables (factors) and different types of RS polynomial functions.

Table 3.12: Number of Experimental Points/ Design Variables for different RS models

Number of design variables	Linear model	Quadratic model	Quadratic with linear cross-terms	Cubic with second order cross-terms	Cubic with second-order and mixed quadratic-cubic cross-terms
2	3	5	6	8	10
3	4	7	10	13	20
4	5	9	15	19	35
5	6	11	21	26	56
6	7	13	28	34	84
7	8	15	36	43	120
8	9	17	45	53	165
9	10	19	55	64	220
10	11	21	66	76	286

For a specific combination of number of design variables and type of polynomial RS model, only some of the DoE methods are feasible because of the requirement for a minimum number of experiments. If the DoE methods outlined in Table 3.7 are observed in particular, then it is clear that for less than 7 design variables only the *Three Level Factorial* design can be used to create sufficient number of experimental points for any of the RS fit functions

stated in Table 3.12. For more parameters, the *Central Composite* design can be chosen. This simple example illustrates one of the most important observations required when DoE selection for a particular problem is made.

3.5.5. *Implementation of RS Methodology in Optimisation and Statistical Tools*

The Response Surface Methodology methods are implemented in tools such as VisualDOC [143], LMS Optimus [144], DEBORA [145], the TMA framework [146], the SILVACO framework [147], COMPACT [148] and others. In general, these and other similar software packages provide modules for both the DoE analysis and the RS Modelling.

3.6. Criteria for Evaluating and Comparing Experimental Designs and RS Models

3.6.1. *Efficiency of Experimental Designs*

3.6.1.1. D-Efficiency

The D-efficiency is an important criterion for DoE/RS evaluation based on the notion that the experimental design should ensure certain properties in the *moment matrix X:*

$$X = \frac{M^T M}{m} \tag{3.13}$$

where *m* is the number of experimental points and *M* is the model matrix (see Section 3.2.3). The determinant of the moment matrix X is an important norm with small value of $|M^T M|$ indicating poor estimation of the coefficients in the RS model. D-efficiency is defined as

$$D - efficiency = \frac{100}{m} |M^T M|^{1/p} \tag{3.14}$$

where *p* is the number of coefficients in the RS model and the efficiency is given in percent. The D-efficiency value shows the level at which the RS model terms are well defined.

3.6.1.2. A-Efficiency

The A-efficiency is a measure based on the sum of variances of the terms of the RS model. The parameters in the formula for the A-efficiency of an experimental design have the same meaning as the parameters used to define the D-efficiency.

$$A - efficiency = \frac{100\,p}{m \text{ trace} \left[(M^T M)^{-1}\right]} \tag{3.15}$$

The A-efficiency is also used to determine how well the coefficients in the RS model are defined. Bigger values for A-efficiency correspond to better definition of the coefficients in the RS approximation.

3.6.1.3. G-Efficiency

The G-efficiency criterion estimates the prediction capability of the RS model, i.e. the quantity is a measure of certainty of the approximate values produced by the RS model.

First, a measure of uncertainty of the predicted response at a particular point in the design space is defined as the **scaled variance** of the predicted response, i.e.

$$v_i(M) = M_i^T (M^T M)^{-1} M_i \tag{3.16}$$

In Equation 3.16, M_i is a vector containing the values of the RS terms at the particular design point i (i.e. a row in the matrix M).

The **estimated scaled error** of the predicted response is defined as the square root of the scaled variance of the predicted response:

$$s_i(M) = \sqrt{v_i(M)} \tag{3.17}$$

Finally, the G-efficiency of the experimental design is defined using the above estimated scaled error as

$$G - efficiency = \frac{100\sqrt{p/m}}{\max_{i=1,\dots,m}(s_i)} \tag{3.18}$$

Higher values of the G-efficiency indicate better prediction capability of the RS model.

3.6.2. *Estimated Error*

The efficiencies described in Section 3.6.1 are influenced by the parameters that refer to the size of the experimental design (m) and size of the RS model (p). As result, the experimental designs with rather small number of experimental points may appear more efficient but in the same time the RS coefficients may not be well defined, or there could be higher uncertainty in the predicted values. To deal with this potential drawback, the estimated scaled prediction errors (Equation 3.17) for all design points can be evaluated using a parameter called the Estimated Average Prediction Square Error, **EAPSE**:

$$EAPSE = \sqrt{\frac{\sum_{i=1}^{m} s_i^2}{m}} \tag{3.19}$$

3.6.3. *Residual Analysis*

The *Residual Analysis* is a procedure that helps to measure how well the RS model predicts the system response by comparing the actual and the predicted response values at the experimental points. If the number of the experimental points is exactly equal to the number of the RS coefficients then no errors will occur at these points because the approximation will fit exactly the response values. The residual analysis makes sense only if more than the number of RS terms experimental points are used. The error at an arbitrary design point can be estimated precisely only if the actual response at that design point is evaluated.

3.6.3.1. Definition of Basic Errors

If y_i is the actual response value at the experimental point i and $\hat{y}_i$ is the predicted response value from the RS model at the same design point, the residual e_i at this experimental point is defined as

$$e_i = y_i - \hat{y}_i \tag{3.20}$$

The *squared sum of errors (SSE)* is computed by summing the squared residuals (Equation 3.20) for all the m experimental points used to fit the RS model:

$$SSE = \sum_{i=1}^{m} e_i^2 \tag{3.21}$$

Another important statistical error in the residual analysis estimates the average variance of the response. The error, called *mean squared error (MSE)*, is defined as

$$MSE = \frac{SSE}{m - p} \tag{3.22}$$

where p is the number of RS model terms (i.e. number of the polynomial coefficients).

3.6.3.2. Scaled Residuals

The *scaled residuals* are often used instead of the un-scaled residuals to provide more information on the RS quality.

1. The *standardized residual*

The *standardized residual (SR, d_i)* is a scaled type of residual, useful to judge for *outliers* , and defined as

$$d_i = \frac{e_i}{\sqrt{MSE}} \quad , i = 1,...,m \tag{3.23}$$

The *outliers* are the experimental points that are not well captured in the RS model. Such points may exist either because of the numerical error in obtained actual response or because the RS model approximates poorly the real responses in that region of the design space.

2. The *studentized residual*

When calculating different residuals, it is useful to take also into account the location in the design space where this residual is estimated. One approach is to use the diagonal elements of the matrix $H = M(M^T M)^{-1} M^T$. When calculating different residuals, the diagonal

elements h_{ii} of the matrix H can be used as weighting factors. A residual that benefits this concept is the so-called *studentized residual* (r_i):

$$r_i = \frac{e_i}{\sqrt{(1 - h_{ii}).MSE}} \quad , i = 1,...,m \tag{3.24}$$

3.6.3.3. PRESS Residual

The *prediction error sum of squares* (*PRESS*) residual is a useful tool to detect and observe points with high influence in the RS model and points that are *outliers*. The *PRESS* residual is calculated using the following technique: an experimental design point, e.g. the i-th point, is selected and the RS model is fit to the remaining (m-1) experimental points. The obtained RS approximation is used to evaluate the response value at the selected point, i.e. at the i-th experimental point of the original DoE. If the predicted value is $\hat{y}_{(i)}$, the *prediction error* $e_{(i)}$, called also the i-th *PRESS residual* at design point i is defined as

$$e_{(i)} = y_i - \hat{y}_{(i)} \tag{3.25}$$

If this procedure is applied to each experimental point then m different PRESS residuals (prediction errors) can be calculated. The *PRESS statistics* is defined as

$$PRESS = \sum_{i=1}^{m} e_{(i)}^2 \tag{3.26}$$

The PRESS residual is a criterion for the influence of a design point on the least squares approximation with large values indicating high influence.

3.6.3.4. R-Student Residual

The *R-student* residual has definition similar to the studentized residual given in Equation 3.24. It is based on the idea used to obtain the *PRESS* residual, i.e. one experimental point is removed from the DoE set and the RS model is constructed. A new average variance of the response is calculated, $MSE_{(i)}$, and used to define the R-student residual:

$$t_i = \frac{e_i}{\sqrt{(1-h_{ii}).MSE_{(i)}}} \quad , i = 1,...,m \tag{3.27}$$

3.6.4. *Analysis of Variance (ANOVA)*

The results of the analysis of a set of experimental data can be summarized in a table called the *analysis of variance table (ANOVA)*. The entries in this table represent the variations of the response values. The table can be used to analyze and test the significance of the RS model. The analysis estimates the degree of similarity between the RS model and the actual relationship between the design variables and the response values and how well determined are the coefficients in the RS model.

3.6.4.1. *ANOVA* for the Whole RS Model

For a constructed RS model, the analysis of variance involves the estimation of the total sum of squares *SST* and splitting it into two components. The first is the *squared sum of errors* (*SSE*, the sum of squares due to the error, Equation 3.21). The second component is the *sum of squares due to the model* or regression, *SSR*. The total sum of squares is defined as

$$SST = SSE + SSR \tag{3.28}$$

The *sum of square due to model* can be calculated from *SSR=SST – SSE*, or in the explicit form:

$$SSR = \sum_{i=1}^{m}(\hat{y}_i - \tilde{y})^2 \tag{3.29}$$

where $\tilde{y} = \left(\sum_{i=1}^{m} y_i\right)/m$ (the average value of the actual responses at all experimental points).

The total sum of squares is also defined as:

$$SST = \sum_{i=1}^{m} y_i^2 - \frac{\left(\sum_{i=1}^{m} y_i\right)^2}{m} \tag{3.30}$$

where m is the total number of the experimental points and y_i is the actual response value at the i-th experimental point.

The analysis of variance test is summarized in the ANOVA-table. The table normally contains the entities shown in the sample ANOVA-table of Table 3.13.

Table 3.13: Sample ANOVA Table

Source of Variations	Degrees of Freedom, df	Sum of squares	Mean square	F_0	P
Model	p-1	SSR=SST-SSE	$MSR = \dfrac{SSR}{p-1}$	$F_0 = \dfrac{MSR}{MSE}$	P
Error	m-p	SSE	$MSE = \dfrac{SSE}{m-p}$		
Total	m-1	SST			

The number of degrees of freedom for the regression sum of squares is equal to the number of coefficients in the equation p minus 1 (df $(SSR) = p - 1$). The number of degrees of freedom for the error sum of squares is equal to the number of experimental points m minus the number of coefficients p (df $(SSE) = m - p$). The number of degrees of freedom for the total sum of squares is df (SST) = df (SSR) + df (SSE) = $m - 1$.

The *test of significance of the fitted model* is performed with the so-called F –test as exampled in Table 3.13. The value of F_0 is used to test the hypothesis that all coefficients in the RS model are equal to zero. The P value is the corresponding probability that all the coefficients in RS model are zero. To calculate the probability P based on the F_0 value F-distribution statistical table is used. Non-zero value for P indicates that at least one coefficient in the RS model is not zero. On the other side, high values of P show that the RS model does not fit well the response data.

The values of the F-Distributions are listed in tables [142] and can be found in the most statistical books.

Another statistical parameter is the *coefficient of multiple determination, R^2*. The R^2 is a measure of the amount of the reduction in the variability of the response obtained by using a

particular RS model. It shows how well the RS model fits the data at the experimental design points. The R^2 is defined as

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} \tag{3.31}$$

For example, if $R^2 = 0.98$ the RS model accounts for 98% of the variability in the response data at the points of the experimental design.

The R^2 always increases when extra terms are added to the RS model, e.g. moving from a linear to a quadratic polynomial model. A perfect fit is given by $R^2=1$ which corresponds to $m=p$. To take into account the above dependence, an adjusted R^2-statistic can be used. This parameter takes into account the number of the terms in the model:

$$R^2_{adj} = 1 - \left(\frac{m-1}{m-p}\right)(1-R^2) \tag{3.32}$$

If unnecessary terms are added to the RS model, the value of R^2_{adj} will often decrease, unlike the value of R^2-statistic. Dramatically different values of R^2 and R^2_{adj} indicate that insignificant terms have been included in the RS model.

The prediction capabilities of an RS model can be estimated using R^2 for prediction, computed from the *PRESS* statistic (equation 3.26) in the following way:

$$R^2_{predict} = 1 - \frac{PRESS}{SST} \tag{3.33}$$

The *coefficient of variation (CV)* is a statistical characteristic which defines how large the error is in comparison with the response value:

$$CV = \frac{\sqrt{MSE}}{average\ response} \tag{3.34}$$

3.6.4.2. *ANOVA* for the Terms in the RS Model

A similar analysis to the ANOVA for the whole RS model can be conducted for the individual coefficients of the RS model. The value of an individual coefficient in the RS model can be evaluated using the method of the least squares. The standard error (*se*) in the value of the individual coefficient b_j of the RS model is:

$$se(b_j) = \sqrt{MSE\ c_{jj}}\ , \quad j = 1,...,p \tag{3.35}$$

where c_{jj} is the diagonal element of the matrix $(M^T M)^{-1}$ corresponding to b_j. The coefficient of variations for an individual coefficient is $se(b_j)/b_j$, . The reciprocal values to these coefficients of variations provide the so-called *t*-statistic values. The values of the *t*-statistic show how well the individual coefficients of the RS model are defined.

CHAPTER 4

NUMERICAL OPTIMISATION METHODS AND INTEGRATED FEM-OPTIMISATION DESIGN TOOL

4.1. Introduction to Design Process

The process of *analysis* of a microelectronic system involves evaluations of its behaviour and response to particular operational or environmental conditions. Computer modelling, using finite element techniques, provides an efficient approach for predicting the physical behaviour of different systems and processes under specified conditions. To ensure the expected accuracy of the simulation predictions, a computer analysis must adopt reliable physics-based models that capture the complexity and the interactions between the modelled phenomena. As the characteristics of the analysed system are evaluated virtually in the computer simulations, a finite element analysis provides knowledge on the behaviour and performance of the analysed system before the stage of actual physical prototyping and testing. For example, the analysis of an electronic package under thermal cycling can predict the fatigue damage in solder interconnects and the thermal stresses across the assembly. Subsequent lifetime models can then relate these quantities to number of cycles to failure of the component.

The process of *design* of a system is different and has to be distinguished from the *analysis* process. The *design* process is a sequence of actions where the parameters and the features

of the system and the process are specified and re-defined in a way to satisfy some requirements. For example, a design process in electronic packaging may to identify the assembly materials for a component or some geometric sizes such as the stand-off height of an assembled chip or the thickness of the printed circuit board. Normally, any specific configuration of the observed parameters in the system requires some form of evaluation, i.e. it must be analysed. It is clear that an *analysis* is solved as a sub-problem in the process of *design* and used to qualify the degree at which the behaviour and the performance of the system satisfies the specified requirements. Based on the analysis results, the investigated system or process can be compared with already observed specifications and can be modified and improved further during the design process.

4.1.1. *General Scheme of the Design Process*

The design process is generally an iterative process of subsequent analyses, evaluations and modifications until a design with the acceptable performance and the desired characteristics is found. Traditionally, the process of design cycles (Figure 4.1) is performed by executing separately each of the involved tasks and controlled by the engineer at each of the stages. Obviously, such a design approach is not effective and in general is resource and time consuming. With developments in computer modelling it became possible to perform the design cycle in a much more effective and automated way. This automated approach, known as **Design Optimisation**, takes techniques from the numerical optimisation theory.
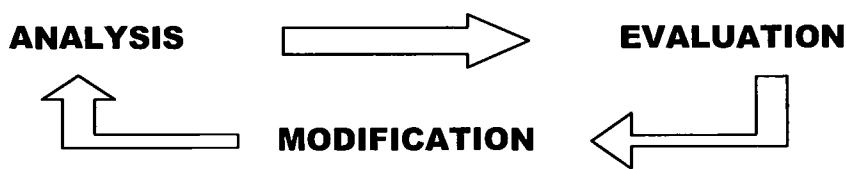
Figure 4.1: Design Process

The design process can be expressed naturally as an optimisation problem and solved as an optimisation task. The modification task is equivalent to changes in the values of the number of design parameters (design variables, DV's) that represent the system features of interest, e.g. the stand-off height of an assembled chip and the board thickness. The analysis results

(from experiments or FE analyses) can provide details on the behaviour and the design response of interest as a result of the particular configuration of the design variables. For instance, in the above example such a response of interest could be the thermal stress in the silicon die. These results are used in the *evaluation* stage of the design process to estimate how *good* the design is and to judge its *feasibility* (i.e. requirements satisfaction). In this example, an assembly design is better if the stress in the die is reduced while still satisfying all other potential requirements.

4.1.2. *Design Optimisation*

The evaluation of design quality (how good) from an optimisation prospective can be represented and based on the so-called *objective function* (the task goal). The objective represents those aspects of the design that a design engineer uses to judge its quality and wishes to improve to the maximum possible level (maximise or minimise), e.g. minimise the die stress. Design requirements normally exist (functionality, manufacturability, cost, limits on the design variations, etc.) and can be fitted into an optimisation problem using the *constraint* functions (e.g. PCB and die thickness must remain within some specified limits). The traditional design process of **analysis-evaluation-modification** cycles performed to aid identification of the optimal and feasible design specifications is equivalent to finding a solution for the related optimisation problem expressed in a mathematical form. The mathematical formulation of a design task is discussed in the next section of this chapter.

A design problem formulated as a mathematical/optimisation task can be solved using numerical optimisation techniques, i.e. the optimal values of the design variables can be found based on pure mathematical algorithms. The software implementation of these methods allows the design engineer to solve a design problem as an optimisation task. It offers an entirely automated and fast approach of performing the design process.

4.2. Optimisation Basic Concepts

In mathematical form, a general **constrained optimisation problem** states:

Find set of design variables $X = (X_1, X_2, ..., X_n)$ that will

$$\text{Minimize (Maximize)} \quad y = F(X) = F(X_1, X_2, ..., X_n) \qquad \text{objective function} \quad (4.1)$$

Subject to:

$$G_j(X) \le 0 \qquad j = 1, m_1 \qquad \qquad \text{inequality constraints} \quad (4.2)$$

$$H_k(X) = 0 \qquad k = 1, m_2 \qquad \qquad \text{equality constraints} \quad (4.3)$$

$$X_i^L \le X_i \le X_i^U \qquad i = 1, n \qquad \qquad \text{limit (side) constraints} \quad (4.4)$$

A set of n design variables is represented as an n-dimensional vector in R^n, $X = (X_1, X_2, ..., X_n)$. The objective function $F(X)$ given in (4.1) is minimised (or maximised) in the feasible region, i.e. the design space defined by the imposed constraints (4.2)-(4.4). The functions used to represent the objective $F(X)$ and the constraints, $G_j(X)$ and $H_k(X)$, are real valued defined on R^n and can be linear or nonlinear, explicit or implicit. The number of inequality constraints (4.2) $G_j(X) \le 0$ and equality constraints (4.3) $H_k(X) = 0$ is m_1 and m_2 respectively (m_1 and m_2 are positive integer numbers). The limit constraints (4.4) specify potential lower bound, $X_i^L \in R^1$, and upper bound, $X_i^U \in R^1$, on each of the design variables X_i ($i=1,n$). Any design point that satisfies the constraints (4.2)-(4.4) is called *feasible*. Thus, the solution of the above optimisation problem, i.e. the *optimal design point*, is the feasible design point that minimises (or maximises) the objective function.

If an optimisation problem is defined through the objective (4.1) and side constraints (4.4) only, i.e. there are no constraints of the form (4.2) and (4.3) in the formulation of the optimisation problem, then this optimisation task is classified as an **unconstrained optimisation problem**.

In general, the objective function and the constraints in the microelectronics design are dependent in an implicit manner on the design variables and/or the system responses. For example, in a design task of reducing the die stress (objective $F(X)$=die stress, to be minimized) by variations of the PCB thickness (X_1) and die thickness (X_2) (the vector of design variables is $X=(X_1, X_2)$), the objective functions is not in an explicit form. The value of the objective can be evaluated either through finite element analysis for stress in the die or through experiments. Because in many cases the experiments are impractical (expensive, time

consuming, dangerous to perform, impossible, etc.), in such design applications evaluation of optimisation functions can rely on the predictions from the computer simulations.

The optimisation theory observed in this chapter considers minimisation problems. A maximization problem with objective $F(X)$ can be represented in minimisation form by changing the sign of the objective function, i.e. minimizing $-F(X)$. By replacing the equality constraints by two equal and opposite inequality constraints, all constraints in an optimisation task can be given in the form of *less than or equal* constraints. This is a more general representation of the constraints and suits the general optimisation algorithms.

4.3. Direct Gradient-Based Iterative Search

In the electronic packaging design (and in the engineering design in general) some initial design configuration is usually available either based on experience or some preliminary design work. The available starting design configuration is referred to as an *initial design*. Because a set of design variables are given as an n-dimensional vector, in the optimisation modelling a specific design configuration will be referred as a design point in the R^n design space.

The numerical optimisation methods are often search procedures that use the following iterative modification of the design variables:

$$X^q = X^{q-1} + \overline{\alpha} S^q \qquad (4.5)$$

where q denotes the iteration number, X^{q-1} is the optimal design point obtained at the last, q-1, iteration of the iterative search procedure and X^q is the new modified design. The design point X^q is an improved design and corresponds to the optimal design at iteration q. In Equation (4.5), S^q is the search direction (**n-dimensional vector**) in the design space at iteration q, and $\overline{\alpha}$ is a **scalar** quantity (step size) that characterizes the distance of the "optimum" movement in the S^q direction.

The iterative procedure for design modification in Equation (4.5) updates the vector of design variables (i.e. the design point) subsequently, starting form an initially defined design point X^0.

Figure 4.2 illustrates the optimisation strategy that utilises the above iterative search approach for locating the optimal design point and provides schematic interpretation in the case of the 2D design space (X_1, X_2).
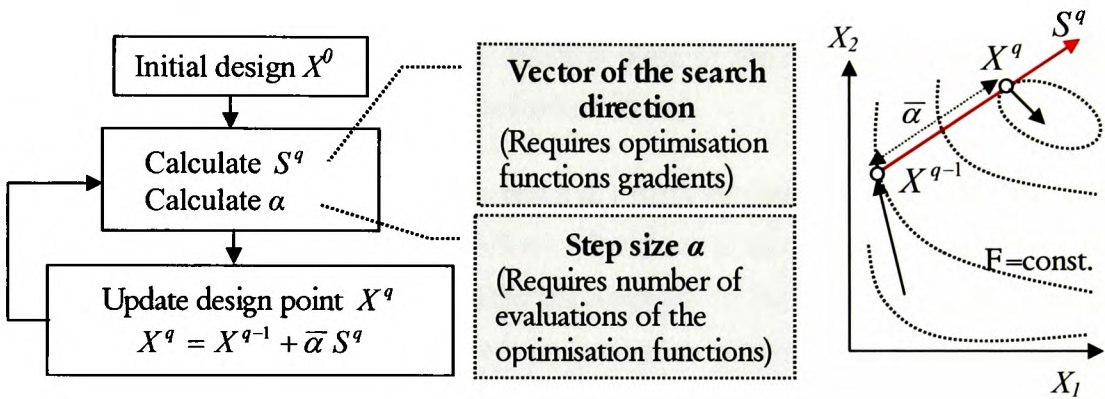
Figure 4.2: Optimisation Iterative Strategy

Two critical definitions are required for this procedure:

1. Definition of the *search direction* S^q ;

2. Efficient strategy for identifying the optimal step size of movement $\overline{\alpha}$ in the S^q direction at iteration q (the *one-dimensional search*).

4.3.1. *Search Direction and the One-Dimensional Search*

In a minimisation problem, **the search direction** could be any arbitrary direction if some small movement in this direction reduces the value of the objective function without violating any of the constraints (if there are any). The gradient of a function, in particular the gradient of the objective function, gives the direction of fastest increment known as the direction of *steepest ascent*. Optimisation problems with objective function requiring minimisation can utilise, for example, the negative of the gradient $S = -\nabla F(X)$ (the *steepest descent* direction) as a rule for the search direction

$$S^q = -\nabla F(X^{q-1}) \qquad\qquad (4.6)$$

As discussed later, there are more efficient methods that can be used to establish a search direction for the iterative gradient-based optimisation.

By definition, a search direction is

- **Usable**, if movement in that direction improves the objective function (objective decreases for a minimisation problem);

- **Feasible**, if by moving in this direction none of the constraints will be violated (definition relevant only to constrained problems).

Once a search direction S in the design space at a particular iteration is specified, the **one-dimensional search** in that direction follows. The idea is to move with an optimal step size $(\overline{\alpha})$ which will provide the best possible improvement to the objective function in that search direction without violating any of the constraints (if there are any). For example, a possible one-dimensional search strategy is to build a polynomial approximation to the objective in the specified search direction using evaluations of the objective (and constraints) for different step sizes α. The optimum value of the step size parameter $\overline{\alpha}$ is the one that minimises the objective approximation in the search direction without violating any potential constraints, i.e. $F(\overline{\alpha}) = \min F(X^{q-1} + \alpha S^q)$. Figure 4.3 illustrates the concept of the polynomial interpolation as a strategy for obtaining the solution of the one-dimensional search sub-problem.
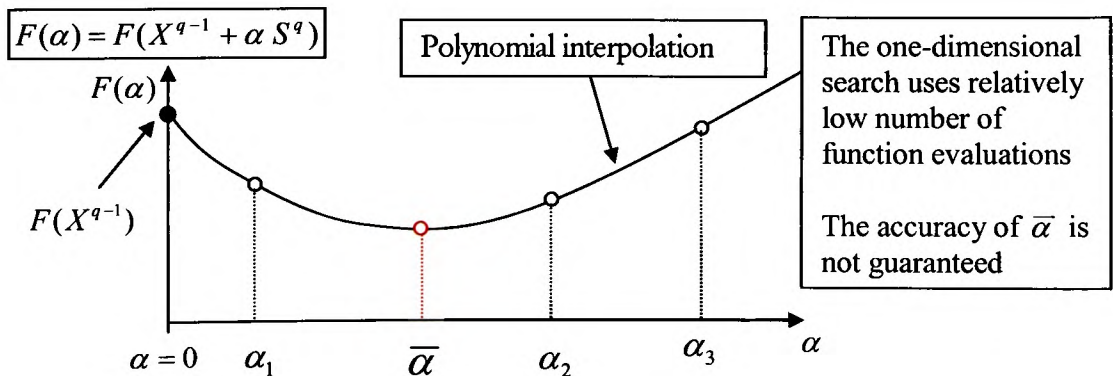
Figure 4.3: One-dimensional Search

At the beginning of the iterative optimisation procedure (iteration $q = 1$), the initial design point X^0 is updated to the new design X^1 by moving in S^1 direction with a calculated step $\bar{\alpha}$. At each of the following design iterations (design cycles) similar sub-problems of defining a new search direction and movement with a specific step are executed. Once the optimum in a particular direction is detected, the last available design is modified using Equation 4.5. This procedure terminates according to specified criteria for relative and absolute convergence or maximum number of design iterations. The optimisation procedure constructs a sequence of design points $X^0, X^1, \ldots, X^k$ such as $F(X^0) > F(X^1) > \ldots > F(X^k)$, where X^k denotes the last design point, i.e. the optimum design where the convergence of the optimisation process is met.

4.3.2. *Finite-Difference Approximations of First Derivatives*

The gradient-based search algorithms use objective/constraint function gradients to construct the vector of the search direction. The mathematical gradient of function $F(X) \in R^1$, where $X = (X_1, X_2, \ldots, X_n)$, is the vector of the partial derivatives with respect to the independent variables $X_1, X_2, \ldots, X_n$, i.e.

$$\nabla F(X) = (\frac{\partial F(X)}{\partial X_1}, \frac{\partial F(X)}{\partial X_2}, \ldots, \frac{\partial F(X)}{\partial X_n}) \tag{4.7}$$

Evaluations of the optimisation objective function and constraints in a design problem may require the relevant predictions for design response from a computer analysis program, i.e. the objective function may not be an explicit analytical function. For example, if the objective is to minimise the thermal stress in die then the objective function value can be calculated using the finite element predictions for stress. In this case, the optimisation using gradient-based techniques can be performed relying on the method of finite difference gradient calculations. The finite-differences are calculated using function evaluations at different design points calculated by small variations in the design variables values from the value at the observed design point.

Using the Taylor-series approximation of twice continuously differentiable univariate function $f(x)$ ($x \in R^1$) the *forward-difference* formula can be obtained as

$$\gamma_{FD}(f,h) = \frac{f(x+h)-f(x)}{h} = f'(x) + O(h)$$

$$O(h) = \frac{1}{2}h f''(x+\theta h) \text{ where } 0 < \theta < 1 \tag{4.8}$$

The $\gamma_{FD}(f,h)$ is the approximation of the first derivative of $f(x)$ using step length h for the argument $x \in R^1$.

For higher accuracy, the *central-difference* approximation with two function evaluations around x can be used:

$$\gamma_{CD}(f,h) = \frac{f(x+h)-f(x-h)}{2h} = f'(x) + O(h^2)$$

$$O(h^2) = \frac{1}{6}h^2 f'''(x+\theta h) \text{ where } 0 < \theta < 1 \tag{4.9}$$

When the function has n independent parameters, $2n$ function evaluations are required using the *central-difference* approximation. If a single evaluation takes several minutes to hours, then the *forward-difference* formula is preferable to reduce the overall calculation time. To keep the error of the approximation in an acceptable range the step size is adapted for each input parameter.

The true gradient of a multivariate function $F(X) \in R^1$, where $X = (X_1, X_2,...,X_n)$, is approximated as

$$\nabla F(X) = (\frac{F(X_1 + \delta X_1, X_2,...,X_n) - F(X)}{\delta X_1}, \frac{F(X_1, X_2 + \delta X_2,...,X_n) - F(X)}{\delta X_2},...,$$
$$\frac{F(X_1, X_2,...,X_n + \delta X_n) - F(X)}{\delta X_n}) \tag{4.10}$$

The Equation (4.10) for finite difference calculation of gradient of multivariate function can be used for gradient calculation of objective and constraint functions that define an optimisation problem.

4.4. Constrained Optimisation

The general optimisation problem given earlier by (4.1)-(4.4) can be redefined in terms of a constrained minimization problem with inequality constraints of type "less than or equal" as

Find set of design variables $X = (X_1, X_2, ..., X_n)$ that will

Minimize $F(X)$

Subject to $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ (4.11)

$$g_j(X) \le 0 \qquad\qquad j = 1, m$$

$$X_i^L \le X_i \le X_i^U \qquad i = 1, n$$

The direct methods for constrained minimisation consider the side constraints separately from the other types of constraints in the above problem definition.

The list of the available direct algorithms varies from simple methods such as the Random Search [26] to more complicated techniques such as the Method of Feasible Directions [26].

4.4.1. *Modified Method of Feasible Directions*

The **gradient-based solution procedure** for the general non-linear constrained optimisation problem (4.11) is outlined in Figure 4.4. As a representative numerical technique that follows this solution approach, the Modified Method of Feasible Directions by Vanderplaats [26] is briefly reviewed in the sub-sections of Section 4.4.1. In general, any gradient-based method for constrained optimisation relies on similar iterative search rules.

INITIALIZATION
Start with an initial X-vector, $X = X^0$ at iteration number $q=0$

$q=q+1$

Evaluate function value at X^{q-1}: $F(X^{q-1})$ and $g_j(X^{q-1})$, $j=1,m$

Identify the set of critical/violated constraints, J

Calculate gradients at X^{q-1}: $\nabla F(X^{q-1})$ and $\nabla g_j(X^{q-1})$ $j \in J$

Determine a USABLE-FEASIBLE search direction S^q

Perform one-dimensional search to find $\overline{\alpha}$

Update the set of design variables,
$X^q = X^{q-1} + \overline{\alpha}S^q$

Convergence to the optimum

NO

YES

EXIT

Figure 4.4: Constrained Optimisation Approach

4.4.1.1. Search Direction in the Modified Method of Feasible Directions

In constrained optimisation, the search direction must ensure that at least for a small movement in that direction the objective function improves. In addition, it must also ensure that moving in that direction none of the constraints will be violated.

Let observe the optimisation problem specified in (4.11). The first step of finding the search direction is to identify the sets of **active**, **inactive** and **violated** constraints. A constraint is called **active** at a specific design point if its value is zero within specified tolerances:

$$g_j(X) = 0 \qquad \text{Active Constrained} \qquad (4.12)$$

In addition to the active constraint definition, **inactive** and **violated** constraints are formulated based on the following inequality relations:

$$g_j(X) < 0 \qquad \text{Inactive Constrained} \qquad (4.13)$$

$$g_j(X) > 0 \qquad \text{Violated Constrained} \qquad (4.14)$$

Once the set of all active and violated constraints are available, their gradients and the gradient of the objective function are calculated. There are different rules to find the search direction in each of the following scenarios:

1) **Case 1**: there are no active or violated constraints;

2) **Case 2**: there are active constraints but none violated;

3) **Case 3**: there are violated constraints.

Only the first two scenarios will be explained in the following sections. In the case of initially violated constraints (Case 3) the goal is first to move into the feasible design region without considering the status of the objective (increasing or decreasing). Once a feasible design point is obtained, then the optimisation is based on the rules described under Case 1 and Case 2. Details on how to deal with problems with initially violated constraints can be found in references [26, 153].

- **Case 1: No active or violated constraints**

An iterative search optimisation procedure, as discussed previously, requires an initial design point (initial set of values for the design variables). Normally, the starting point is within the **feasible** design space, i.e. this design point satisfies all requirements imposed through the constraints.

If at the beginning of the optimisation there are no violated and no active constraints and if the requirement for feasibility is satisfied, the only requirement applied to the search direction at the first step is the direction to be *usable*. The reason is because a small movement in any direction will not violate any of the constraints. One of the preferable choices in the case of inactive constraints only is to start the optimisation process with the *steepest descent* search direction (Equation 4.6).

Let emphasize that if the initial design point is feasible then during the optimisation process the feasible design space will never be violated and only feasible design points will be accepted. A result of this is that only inactive or active status of constraints can occur.

If after the first movement in a steepest descent direction there are no active constraints, the new *usable* direction must be determined. Normally the choice is for a conjugate direction instead of using the steepest descent again. One of the options is to use the Fletcher-Reeves conjugate direction (outlined in Section 4.5.1.1). The advantage of using conjugate direction is that instead of moving in a direction perpendicular to the last one, the search proceeds in the *steepest descent* direction but with some fraction of the previous search direction. As it will be discussed, this small modification gives enormous improvement in the convergence rate.

- **Case 2: Active constraints but non violated**

A major search direction problem in the constrained optimisation is how to determine the next search direction if the design from the last iteration is feasible but there are active constraints. In such a scenario a **usable-feasible** direction is required to improve the objective (*usable*) and to move in parallel/away from the active constraints (*feasible*).

If the set of the active constraints is J, the solution of the following **direction finding problem** (4.26)-(4.28) gives the search direction S^q at an iteration where active constraints are encountered:

$$\text{Minimize} \quad \nabla F(X^{q-1})^T S^q \tag{4.15}$$

Subject to:

$$\nabla g_j(X^{q-1})^T S^q + \delta \le 0.0 \qquad j \in J \qquad \text{Condition for Feasibility} \tag{4.16}$$

$$(S^q)^T S^q \le 1.0 \qquad\qquad\qquad \text{Limits on the Search Direction} \tag{4.17}$$

The objective in (4.15) is condition for *usability* and brings the requirement to obtain search direction that will improve the objective function $F(X)$ in the original problem. The condition for *feasibility* (4.16) guarantees that the S^q direction points towards the feasible

region, i.e. it is a direction which will not violate any of the current active constraints at the X^{q-1}. The small positive scalar δ ensures that the search direction will not be exactly perpendicular to the gradient of the constraint and will always provide direction away from the boundaries $g_j = 0$, $j \in J$. As the problem (4.15)-(4.16) itself may have infinite number of solutions, a restriction that imposes limits on the search direction (4.17) is essential to obtain a unique solution of the direction finding problem.

The algorithm of solving the problem (4.26)-(4.28) can be found in Reference [26].

Figure 4.5 illustrates schematically the requirement for usable and feasible search direction in a constrained optimisation problem in the 2D design space.
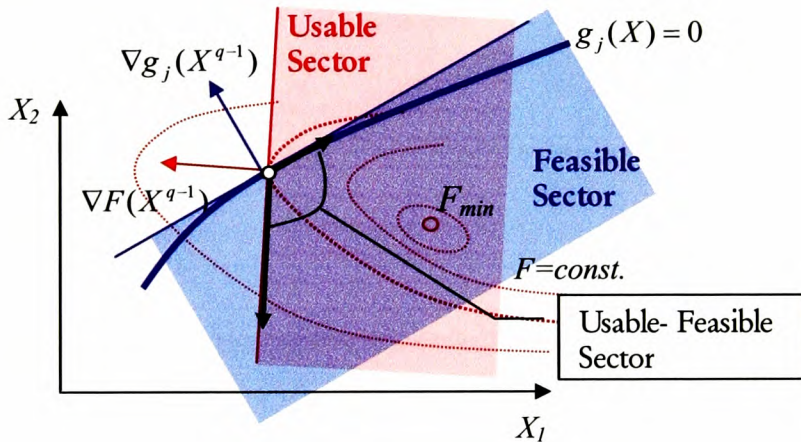
Figure 4.5: Usable-Feasible Concept for Search Direction

4.4.1.2. The One-Dimensional Search

Once the *usable-feasible* direction is identified, the next task is to perform the one-dimensional search in this direction and to find the step size $\bar{\alpha}$ that minimises the objective function in that direction, i.e. to solve the task:

$$\text{Minimize } F(\alpha) = F(X^{q-1} + \alpha S^q)$$

It is important that the design point must be kept also feasible ($g_j(\alpha) = g_j(X^{q-1} + \alpha S^q) \leq 0$).

As discussed in Section 4.3.1, a common strategy for a one-dimensional search is to use polynomial interpolation for the scalar, $\overline{\alpha}$, relying on the objective function evaluations in the search direction. Similar polynomials are used for the constraint functions to provide the constraint optimum in the search direction. Having these polynomial approximations, the step size that provides the minimum of the objective without violating any constraints can be found. Notice, that in some cases the optimum step size can be associated with the unconstrained minimum of the objective in the search direction, but in others with a constraint minimum (Figure 4.6).
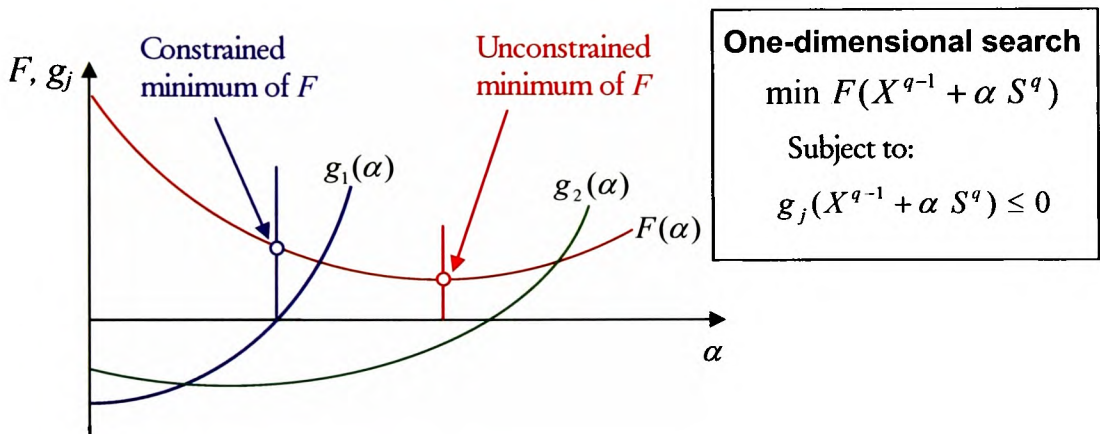
Figure 4.6: Constrained and Unconstrained One-dimensional Search

The rest of the section details the procedure for obtaining the set of step size values used for creating the polynomial approximations to the optimisation functions.

The one-dimensional search strategy must start with testing an initial value for α . Using the rule for the design update,

$$F(X^q) = F(X^{q-1} + \overline{\alpha} S^q) \tag{4.18}$$

an approximation to the objective function (same for the constraint functions) at the design point X^q can be written as

$$F(X^q) = F(X^{q-1}) + \sum_{i=1}^{n} \frac{\partial F(X^{q-1})}{\partial \overline{\alpha}} \overline{\alpha} = F(X^{q-1}) + \sum_{i=1}^{n} \frac{\partial F(X^{q-1})}{\partial X_i} \left(\frac{\partial X_i}{\partial \overline{\alpha}} \right) \overline{\alpha} \tag{4.19}$$

The term $\dfrac{\partial F(X^{q-1})}{\partial X_i}$ is the i-th component of the gradient vector of the objective at X^{q-1}

(notice that the gradient vector is available). From the relationship $X_i^q = X_i^{q-1} + \overline{\alpha} S_i^q$ the

term $\dfrac{X_i}{\partial \overline{\alpha}}$ can be also evaluated: $\dfrac{X_i}{\partial \overline{\alpha}} = S_i$ (S_i - the ith component of S^q is also available).

Using equation (4.18), the slope of the objective (same for the constraints) at $\alpha = 0$ can be obtained by

$$\frac{\partial F(X^{q-1})}{\partial \overline{\alpha}} = \nabla F(X^{q-1})^T S^q \tag{4.20}$$

If at the beginning of the search in a particular search direction the objective is expected to be reduced with some fraction, e.g. 10%, then

$$F(X^q) = F(X^{q-1}) + \sum_{i=1}^{n} \frac{\partial F(X^{q-1})}{\partial \overline{\alpha}} \overline{\alpha} = F(X^{q-1}) - 0.1 \left| F(X^{q-1}) \right| \tag{4.21}$$

This relationship gives a possible estimation for the value of the step size α as a first attempt to move in the one-dimensional search which will reduce the objective function with certain percentage (e.g. 10%):

$$\overline{\alpha}_{est} = \frac{-0.1 \left| F(X^{q-1}) \right|}{\left(\dfrac{\partial F(X^{q-1})}{\partial \overline{\alpha}} \right)} \tag{4.22}$$

The same procedure can be applied to a constraint (gradient information for constraints is also available). If a constraint is not active then the above equation provides a linear approximation that can be used to define the step size. The step size will be such that it will make the constraint active, i.e. $g_j(\alpha) = 0$ (instead of reducing the objective). Mathematically this requirement can be written as

$$g_j(X^q) = g_j(X^{q-1}) + \sum_{i=1}^{n} \frac{\partial g_j(X^{q-1})}{\partial \overline{\alpha}} \overline{\alpha} = 0.0 \tag{4.23}$$

This relation provides estimation for the step size α which will drive the constraint to its active status:

$$\overline{\alpha}_{est} = \frac{\left| g_j(X^{q-1}) \right|}{\left(\dfrac{\partial g_j(X^{q-1})}{\partial \overline{\alpha}} \right)}$$

(4.24)

The first step size in the search direction, α_1, is the smallest positive proposal among the estimations made for the objective (Equation 4.22) and the inactive constraints (Equation 4.24). In fact, similar approximations with respect to the lower and the upper bounds of the design variables are also used to obtain the step size which will not violate their design limits. With this step size, the objective and the constraints are evaluated ($F(\alpha_1), g_j(\alpha_1)$)

After the first step, α_1, in the one-dimensional search is estimated, the next stage is to find bounds on the optimal step size $\overline{\alpha}$ that in the general case will correspond to the constrained minimum of the objective in the S^q direction. Only positive values of the step size are possible and need to be considered because of the search in a usable-feasible search direction. Notice that the objective and constraint values at $\alpha = \alpha_0 = 0$ (the design point X^{q-1}) and at the first step $\alpha = \alpha_1$ are available at this stage of the one-dimensional search.

If we start initially with a feasible design, the design point X^{q-1} is also feasible, i.e. all constraints are satisfied ($g_j(\alpha_0 = 0) \leq 0$). The search direction S^q is determined to improve the objective function. Number of possible scenarios for identifying bounds on $\overline{\alpha}$ can be observed. These cases and related step size decisions are:

- **Case 1**: If at $\alpha = \alpha_1$ the objective function is greater than the objective at $\alpha = \alpha_0 = 0$ the upper bound on $\overline{\alpha}$ is found, and this is $\alpha = \alpha_1$ (Figure 4.7)

- **Case 2**: If at $\alpha = \alpha_1$ any of the constraints becomes violated, i.e. $g_j(\alpha_1) > 0$ the upper bound on $\overline{\alpha}$ is again $\alpha = \alpha_1$ because we move from a feasible design to an infeasible design (Figure 4.7)

In both **Case 1** and **Case 2** we need the objective and constraint values for one more step size $(\alpha_{mid} = \alpha_1/2)$. With these function values at α_0, α_{mid} and α_1 the objective and the constraints can be approximated. The approximations then can be used to find the optimal step size movement $\overline{\alpha}$ in the direction S^q.

- **Case 3**: Objective function at $\alpha = \alpha_1$ is smaller than at $\alpha = \alpha_0 = 0$ and none of the constraints are violated at $\alpha = \alpha_1$. In this case α_1 is not upper bound for $\overline{\alpha}$ and a movement further in the search direction can be done. A simple rule is to perform the next movement with step size two times greater than the previous one. It may take many steps until upper bound is found, i.e. Case 1 or Case 2 takes place. In such scenario, only the last three (or four) solutions are retained - α_{i-2}, α_{i-1} and α_i. If more than two steps are undertaken than the lower bound will not have a value $\alpha_0 = 0$ but instead will be carried out and assigned to α_{i-2}. Similarly to the previous cases, all objective and constraint values for those retained step sizes are stored and used for $\overline{\alpha}$ interpolation (Figure 4.7)
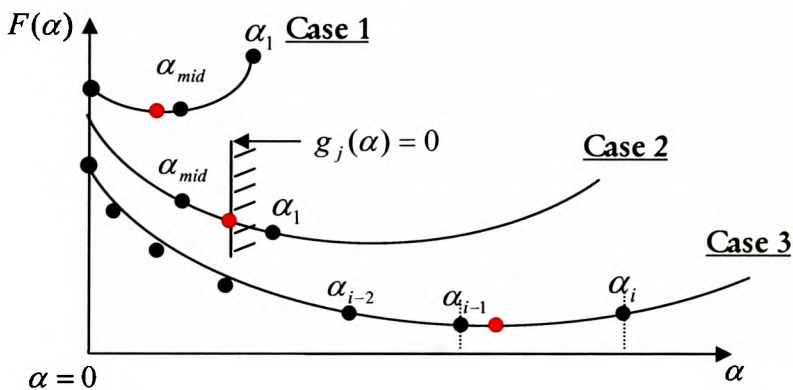
Figure 4.7: Optimal Step Size - Bounds and Points for Polynomial Fitting

Let for convenience denote the lower bound as α_L and the upper bound as α_U which both specify the limits on the solution of the one-dimensional search problem. To obtain the one-dimensional search solution, the objective function and constraints are approximated using polynomial curve fits. The order of the polynomial depends on the number of the stored

function values – second if three different step-sizes are evaluated or cubic if stored data is for four points. The polynomial that approximates the objective is used to identify the minimum of the function.

If the polynomial approximates a constraint, then the curve fit is used to calculate the value of the step size which will make it zero (in terms of the one-dimensional search it means to drive the constraint to become active). Using basic mathematical observations, upper bounds on $\bar{\alpha}$ based on the active status of constraints are derived. The minimum upper bound from all constraints indicates the proposed solution of the one-dimensional search from the *feasibility* point of view.

With no violated constraints at the $\alpha = 0$, the final estimate for the optimal step size $\bar{\alpha}$ is the minimum between the step size which minimizes the objective and the step size acting as the upper bound for the constraints.

4.4.2. *Convergence Criteria for the Optimum Design*

One of the important and critical parts in an optimisation iterative search technique is the termination of the design optimisation procedure. Different criteria can be used to judge for convergence and used in conjunction to navigate the optimisation process. The mathematically proven criterion includes satisfaction of the necessary (and for some type of problems also sufficient) Kuhn-Tucker conditions. The maximum number of iterations and the diminishing returns can be used also to terminate the optimisation process.

The most common criteria are explained under the following sub-section.

4.4.2.1. Kuhn-Tucker Necessary Conditions

The necessary Kuhn-Tucker conditions are well described in any book on optimisation theory. If a design X_{opt} is optimal for an optimisation problem these conditions state that the following three conditions are true:

- X_{opt} is feasible $\hspace{8cm}$ (4.25)

- $\lambda_j g_j(X_{opt}) = 0 \qquad j = 1, m \qquad \lambda_j \geq 0$ $\hspace{3cm}$ (4.26)

- $\nabla F(X_{opt}) + \sum_{j=1}^{m} \lambda_j \nabla g_j(X_{opt}) = 0$ (4.27)

  $\lambda_j \geq 0$

In the above conditions λ_j are called Lagrange multipliers. The first condition states that the optimal design needs to satisfy all imposed constraints while the second determines the Lagrange multiplier zero for each inactive constraint at the X_{opt}, i.e. if $g_j(X_{opt}) < 0$ then $\lambda_j = 0$.

The third condition is a statement for usability and feasibility. It states that at the optimal point there is a linear combination (the Lagrange multipliers) of gradients of the objective and the active constraints which vanish. For unconstrained minimisation problems this condition simply states that at the optimum design the objective function gradient must vanish.

4.4.2.2. Objective Relative and Absolute Convergence (diminishing returns)

The first of these convergence criteria requires that the relative change in the objective between two iterations is less than a specified tolerance ε_{rel}

$$\frac{\left| F(X^q) - F(X^{q-1}) \right|}{\left| F(X^{q-1}) \right|} \leq \varepsilon_{rel}$$ (4.28)

The second criterion states that the absolute change in the objective between two iterations is less than some absolute tolerance ε_{abs} and it is satisfied if

$$\left| F(X^q) - F(X^{q-1}) \right| \leq \varepsilon_{abs}$$ (4.29)

Using both of these convergence criteria it can be assured that continued iterations will be avoided if the optimum is approached asymptotically. If the values of the objective function are large, then the relative convergence acts. If the objective values are small, then the termination of the optimisation process is due to the absolute convergence criteria. These

criteria need to be met on a number of consecutive iterations in order to apply and accept termination of the optimisation task.

4.4.2.3. Maximum Number of Iterations

Similarly to other iterative processes, the maximum number of iterations can act as termination criterion. In some circumstances this may prevent excessive computations. The gradient based search techniques normally approach the optimal point within relatively small number different search directions. Thus the maximum iteration convergence criteria have mainly preventive purposes.

4.4.3. *Sequential Linear Programming (SLP)*

The sequential linear programming approach uses linearization of a non-linear problem and solves the linear approximation with some of the well established methods in the linear programming (e.g. simplex method). Having the approximate optimal point, a new linearization around that point is used to construct and solve a new optimisation linear problem. These steps are performed repeatedly until acceptable precision of the solution is obtained. The technique is called SLP (Sequential Linear Programming) and is also known as the Kelley's cutting plane method [151].

The first step in the algorithm is to create a linear approximation to the objective function and constraint functions using the Taylor's series expansion rule. If X^{q-1} denotes the point about which the expansion is performed and the vector δX is $\delta X = X - X^{q-1}$, then the linear approximated problem becomes:

$$\text{Minimize} \quad F(X) \approx F(X^{q-1}) + \nabla F(X^{q-1})^T \delta X \tag{4.30}$$

Subject to

$$g_j(X) \approx g_j(X^{q-1}) + \nabla g_j(X^{q-1})^T \delta X \leq 0 \qquad j = 1, m$$

$$h_j(X) \approx h_k(X^{q-1}) + \nabla h_k(X^{q-1})^T \delta X = 0 \qquad k = 1, l \tag{4.31}$$

$$X_i^L \leq X_i^{q-1} + \delta X_i \leq X_i^U \qquad i = 1, n$$

In the above linear optimisation problem δX is the vector of the design variables and the values of the functions and gradients at the point X^{q-1} are known (constants) together with the coefficients.

In the case of underconstrained problems, with less active constraints at the optimal design than the number of the design variables, it is possible to construct an unbounded linear approximation problem. The way to deal with such difficulties is to define "move limits" on the design variables. These move limits can be imposed as a percentage from the variables and will provide a solution of the problem within these bounds. Generally, the move limits are adjusted and reduced at the subsequent design iteration during optimisation until the desired convergence to the exact solution is obtained. The critical part in SLP algorithm is how to define the move limits and what strategy for the factor of the move limits reduction will ensure the highest possible efficiency [26].

The idea behind sequential optimisation and the move limits approach is exploited also in other search techniques such as the classical sequential quadratic programming (SQP) [26] and the recently developed by the Centre for Quantitative Methods (CQM, see [148]) new sequential optimisation approach [152].

4.5. Unconstrained Optimisation

In the optimisation theory, a problem is classified as **unconstrained** if no constrains of type (4.2)-(4.3) are imposed. The lower and upper bounds on the variables are still represented and taken into account in the unconstrained optimisation. In the area of electronic packaging many design problems can be represented as unconstrained optimisation design tasks, e.g. identification of the optimal material properties in specified ranges for an electronic assembly for maximised lifetime and reliability. The form of an unconstrained optimisation problem is:

Find set of design variables $X=(X_1, X_2, ..., X_n)$ that will

Minimize $F(X)$ (4.32)

Subject to: $X_i^L \leq X_i \leq X_i^U$ $i = 1, n$ (4.33)

The necessary condition for a vector X^{opt} to be an extreme point for an n-variable objective function $F(X)$ is the gradient vector $\nabla F(X^{opt})$ to vanish at that point, i.e.

$$\nabla F(X^{opt}) = 0 \qquad (4.34)$$

The above condition is only necessary to judge a point for optimality. The sufficient condition states that the matrix of the second partial derivatives of $F(X)$, the Hessian matrix, is positive defined (has positive eigenvalues). If this condition is true for the Hessian at the point X^{opt} then this point is at least a local optimum. The sufficient condition for global optimum requires at all points in the design space the Hessian to be positive defined. It is almost impossible to verify this condition for real design applications.

The process of iterative numerical optimisation of the unconstrained problem (4.30)-(4.31) is described in Figure 4.8 flowchart.
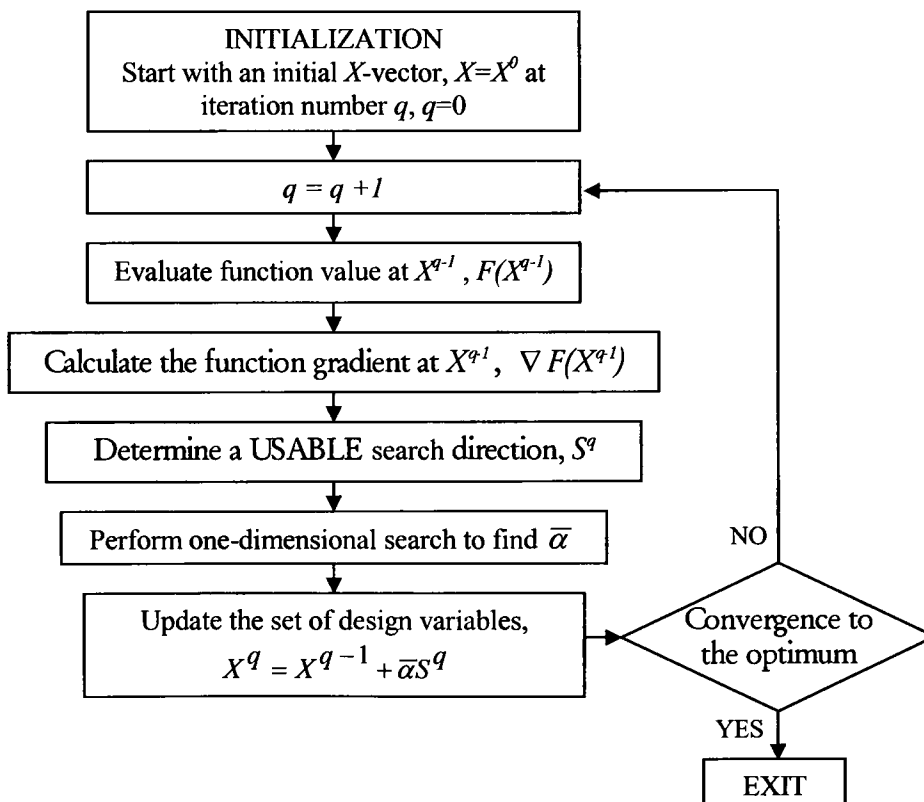
Figure 4.8: Iterative Search Procedure for Unconstraint Optimisation

The iterative search procedure for unconstrained optimisation is similar to those observed in the constrained optimisation:

1. Determine a usable search direction at each of the search iterations, i.e. S^q;

2. Perform the one-dimensional search and estimate $\bar{\alpha}$, i.e. find the step size which gives the minimum of the objective function in the present search direction;

3. Formulate criteria for convergence to the optimal solution and termination of the optimisation process.

The following sub-sections of the chapter provide further discussion of the above steps in the unconstrained case for optimisation.

4.5.1. *Search Direction in Unconstrained Minimisation*

A common strategy for determination of the search direction at the first iteration ($q=1$) is to use the direction of *the steepest descent*.

The steepest descent method can be adopted as a search direction rule for the whole optimisation process, i.e. to be used at any design iteration. In this case, the search directions at two subsequent iterations are perpendicular. Although the method is one of the best known first-order algorithms, it is not efficient and has poor rate of convergence [26]. Most effective algorithms use the steepest descent to establish the search direction only at the first iteration. It can be used also at any iteration with poorly established search direction caused by nonlinearities.

4.5.1.1. Fletcher-Reeves Conjugate Search Direction

The Fletcher-Reeves conjugate method [11,26] is a simple modification of the steepest descent method which improves dramatically the optimisation search. The search direction is defined as

$$S^q = -\nabla F(X^{q-1}) + \beta_q\, S^{q-1}\,, \qquad \text{where} \quad \beta_q = \frac{\left|\nabla F(X^{q-1})\right|^2}{\left|\nabla F(X^{q-2})\right|^2} \qquad (4.35)$$

The Fletcher-Reeves search direction takes into account the optimisation history. It requires n (number of design variables) or fewer iterations for convergence if the problem is truly quadratic. Generally it is beneficial for optimisation if the Fletcher-Reeves method is restarted with the steepest descent when the search process slows down.

4.5.1.2. BFGS Search Direction

This method [15-18] is known as a quasi-Newton method. It creates the search direction using an approximation matrix H to the inverse matrix of the second derivatives of the objective function (Hessian). The search direction is given by

$$S^q = -H \, \nabla F(x^{q-1})$$ (4.36)

At the first iteration the search direction is set to the steepest descent by setting H to be equal to the identity matrix, I. After the first iteration, the H matrix is updated in the following way:

$$H^{q+1} = H^q + D^q$$ (4.37)

where

$$D^q = \left(\frac{\sigma + \tau}{\sigma^2} \right) pp^T \frac{1}{\sigma} \left[H^q y p^T + p(H^q y)^T \right]$$ (4.38)

$$\begin{aligned} \sigma &= p^T y & p &= X^q + X^{q-1} \\ \tau &= y^T H^q y & y &= \nabla F(X^q) - \nabla F(X^{q-1}) \end{aligned}$$ (4.39)

The BFGS method has similar performance to the Fletcher-Reeves method and for strictly quadratic functions also converges in n or fewer iterations. Since most optimisation problems are non-quadratic, the method may have to be restarted by setting $H=I$.

4.5.2. One-Dimensional Search

The general strategy of the one-dimensional search was explained in Section 4.4.1.2 under the constrained optimisation discussion. The one-dimensional search in the unconstrained

minimisation case is similar and can follow the constrained one-dimensional search with the only difference that there are no constraints to be observed.

Once the optimum value $\bar{\alpha}$ is determined, the design point is updated using Equation 4.5.

4.5.3. *Convergence Criteria*

Details on the convergence criteria were given in Section 4.4.2 of the chapter. They are similar for both unconstrained and constrained problems with small modifications due to the presence of constraints in a constrained optimisation task.

4.6. Non-Gradient based Optimisation

Apart from the class of the gradient-based optimisation methods, there is a large variety of other methods for optimisation. A class of novel optimisation algorithms based on a probabilistic (evolutionary) search of the design space has become very attractive in the recent years [31-33, 154-157]. These algorithms are zero-ordered and use only the function values (non-gradient optimisation) to find the optimal solution. They require extensive function evaluations which is their major disadvantage compared with the gradient methods for optimisation. Thus, they are unpractical and hardly can be used in simulation-based design optimisation process if the evaluation of the optimisation function required is time consuming. If the design approach uses the strategy of the Response Surface Modelling, i.e. the optimisation function evaluations are based on explicit analytical function then these evolutionary techniques become very efficient. Important advantages that the non-gradient optimisation techniques obey are:

- Easy to be programmed and incorporated as a module into a design framework;

- Do not require continuity of the optimisation function over the design space;

- Best probability performance to find the global, or near the global optimum.

4.6.1. *Genetic Algorithms (GA)*

The Genetic Algorithm (GA) [31] offers the opportunity to solve multidimensional optimisation problems efficiently using relatively high number of function evaluations. The GA is effective and applicable for problems where optimisation functions are defined explicitly or where the evaluations are inexpensive.

Genetic algorithm is a model that mimics the behaviour of the mechanisms of evolution in nature. This is done by creating individuals in a population with slightly different properties, the so-called *chromosomes*. These individuals in the population then go through a process of simulated evolution.

The implementation of genetic algorithms usually consists of the following steps:

1. Evaluation of the fitness (objective) of all of the individuals (designs) in the population.

2. Creation of a new population (set of designs for further evaluation) by performing operations such as the **mutation** on the individuals whose fitness has just been measured.

3. Replace the old population with the new population in the next iteration.

Genetic algorithms are often described as global search methods that do not use gradient information. The advantage of their use is that they are more likely to find the global optimum of the problem while the gradient-based methods always terminate at the nearest local optimal design point. On the other side, these algorithms need a significant number of evaluations of the optimisation functions.

Genetic algorithms are generally highly parallelizable algorithms because the required evaluations of the population in each generation are independent and can be performed separately. They are also very general, robust algorithms and simple to implement.

4.6.2. *Particle Swarm Optimisation (PSO)*

The PSO was developed by Kennedy and Eberhart [32] as an evolutionary algorithm based on a simplified social model. It mimics the natural behaviour of a swarm, e.g. swarm of bees, searching for a food source. In this interpretation the search of the best available food source

(i.e. optimum) is navigated based on the own memory of each particle of the swarm (a bee) as well as the knowledge of the swarm as a whole. While most of the other evolutionary algorithms are suited for discrete optimisation problems, the PSO is naturally applicable to the continuous design spaces.

In Chapter 7 of the thesis, a demonstration of the class of non-gradient based optimisation algorithms is made using the PSO optimisation method. A section in that chapter will also provide more details on the PSO algorithm.

4.6.3. *Other Evolutionary Algorithms*

Many other heuristic search algorithms for solving optimisation problems using only the function values are developed recently:

1. **Ant Colony Optimisation** (ACO) [157] can be used to solve discrete optimisation problems;

2. **Simulated Annealing** (SA) [159] mimics the mechanics physical phenomenon of equilibrium of atoms during an annealing process;

3. **Tabu Search**, proposed by Glover [154], explores the design space taking advantage of search history;

4. **Differential Evolution** (DE) [33] is a stochastic-based algorithm. It is demonstrated recently for electronic packaging design optimisation in one of the latest work by Leon *at al.* from Nokia Mobile Phones Inc. [158].

4.7. Optimisation Examples

4.7.1. *Example of a Non-Linear Unconstrained Optimisation Problem*

The first example demonstrates the minimisation process of an unconstrained function of two variables x and y. The Fletcher-Reeves method is used to perform the gradient-based search to locate the optimal point of the function:

$$F(x,y) = 20\,(x^2 - 2y)^2 + 50\,(x-1)^2 + 1$$

The necessary condition for a minimum of the function (gradient of the function equal to zero) states that

$$\left| \begin{array}{l} \dfrac{\partial F(x,y)}{\partial x} = 80x^3 - 160xy + 100x - 100 = 0 \\ \dfrac{\partial F(x,y)}{\partial y} = -80x^2 + 160y = 0 \end{array} \right.$$

Simple analytical calculations provide the exact solution of the optimisation problem as $X_{opt}^{uncon} = (x_{opt}, y_{opt}) = (1, 0.5)$. This function has a similar form to the famous Rosenbrock function which is used often as a test case for the robustness of an optimisation algorithm.

Optimisation process for this problem using the Fletcher-Reeves method results in a search process of 11 optimisation iterations. Each of these iterations corresponds to a search in one specific search direction in the 2D x-y design space. The last 5 search directions are for obtaining precise convergence based on the optimisation convergence settings. Figure 4.9 illustrates the minimisation process. The search starts from an initial point $X^0 = (-2, 15)$ with corresponding objective value at that point $F(X^0) = 13971$.
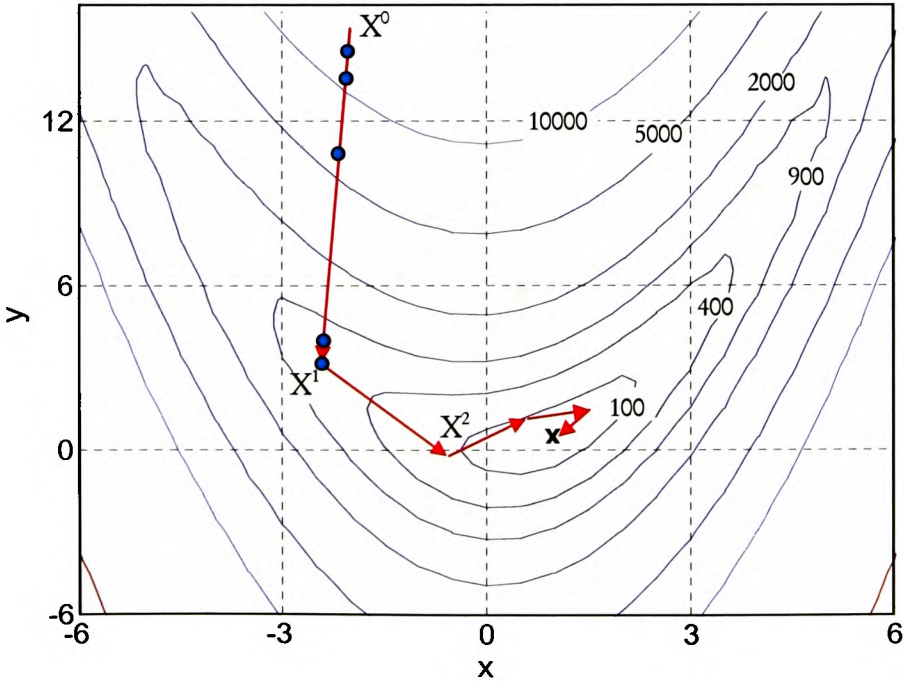
Figure 4.9: Unconstraint Minimization Process using Fletcher-Reeves Method

Table 4.1 provides the numerical results for the evaluated design points and the related function values.

Table 4.1: Optimisation History for Unconstraint Optimisation

Design Point	x	y	$F(x,y)$
X^0 *(initial)*	-2.0	15	13971.0000
X^1	-2.39787	2.940042	578.616606
X^2	-0.59269	-0.42284	156.487892
X^3	0.421117	0.779083	55.8889944
X^4	1.41442	1.315893	17.5555150
X^5	1.484622	1.19764	13.4739017
X^6	1.046628	0.504059	1.26118297
X^7	1.014866	0.520971	1.01392443
X^8	1.002715	0.500292	1.00084006
X^9	0.9999905	0.4999838	1.00000001
X^{10}	0.9999905	0.4999888	1.0
X^{11} *(optimal)*	0.9999905	0.4999888	1.0

Detailed results on the one-dimensional search for this example will be given only for the first iteration. In this case, the initial design X^0 is improved by moving in the S^1 search direction with step $\overline{\alpha}$, and ultimately the design point X^1 is obtained. Figure 4.9 presents the evaluated design points during the first search direction corresponding to different step sizes α in the one-dimensional search. Table 4.2 contains the exact numerical data for the one-dimensional search performed in the first search direction.

Table 4.2: Optimisation History for Unconstraint Optimisation

Step size in 1-D search in S^1	Design Point	x	y	$F(x, y)$
	X^0	-2	15	13971
α^1	$X^1_1 = X^0 + \alpha^1 S^1$	-2.02088	14.36707	12609.92
α^2	$X^1_2 = X^0 + \alpha^2 S^1$	-2.05467	13.34297	10560.43
α^3	$X^1_3 = X^0 + \alpha^3 S^1$	-2.14312	10.66184	6093.3
α^4	$X^1_4 = X^0 + \alpha^4 S^1$	-2.3747	3.64256	624.6109
α^5	$X^1_5 = X^0 + \alpha^5 S^1$	-2.98097	-14.7342	30214.75
$\overline{\alpha}$	$X^1 = X^0 + \overline{\alpha}\, S^1$	-2.39787	2.940042	578.6166

The statistics of the performance of the Fletcher-Reeves method on this example shows 136 analysis calls in total with 11 gradient requests.

4.7.2. *Example of a Non-Linear Constrained Optimisation Problem*

This optimisation example uses the same objective function as the one in Section 4.7.1 but the optimisation is subject to one inequality constraint. The optimisation problem is

$$F(x, y) = 20\,(x^2 - 2y)^2 + 50\,(x-1)^2 + 1$$

$$\text{Subject to: } g(x, y) = 15x - 4y + 40 \le 0$$

Figure 4.10 presents a graphical interpretation of the search process using the Modified Method of Feasible Directions (MMFD) by Vanderplaats [26].

The search process is started from the same initial point as the one used in the unconstrained minimization example ($X^0 = (-2, 15)$). The optimal design point in this case lies on the linear constraint being imposed. Notice that at the optimal point the constraint is active. Summary on the optimisation process and the function improvement plus the constrained values at the design iterations are given in Table 4.3.

Table 4.3: Optimisation History for Constraint Optimisation

Design Point	x	y	F(x,y)	G(x,y)
X^0 (initial)	-2	15	13971.0000	-50.0000000
X^1	-2.39787	2.940042	578.616606	-7.72826613
X^2	-2.06328	2.262689	471.624210	0.000000000
X^3	-2.06328	2.262689	471.623948	1.50000000e-005
X^4	-2.0299	2.387879	468.601909	-1.92329708e-010
X^5 (optimal)	-2.0299	2.387879	468.601712	1.49998077e-005

The constrained optimum is (-2.0299, 2.3879) and the optimal function value and the constraint value are respectively 486.6 and 1.5e-5 (i.e. 0 within numerical tolerances of the method). The overall constrained optimisation in the problem performs in 23 analysis calls in total with 5 gradient requests.

$$g\,(x,y) = 15x\text{-}4y\text{+}40 \le 0$$
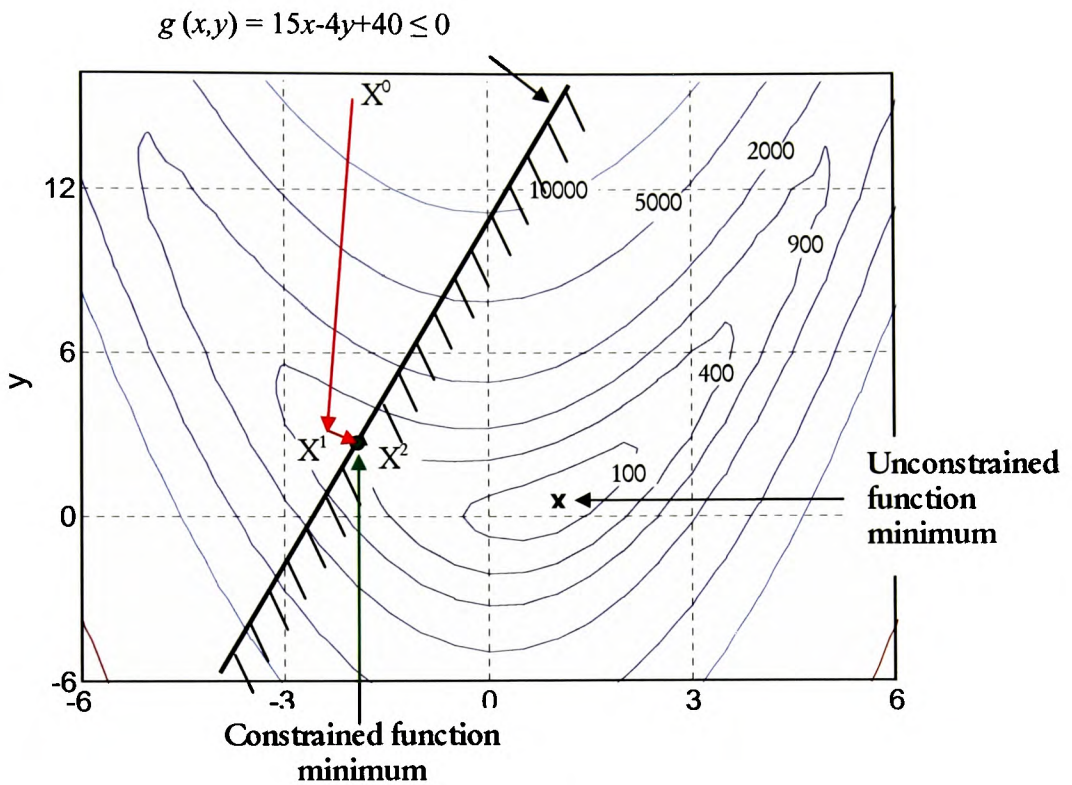
Figure 4.10: Constraint Minimization Process using MMFD

4.8. FEM-Optimisation Integrated Approach

4.8.1. *Demand for a General Optimisation Framework*

In the area of the design and packaging of electronics components and systems, different problems related to the reliability, performance or the cost of a product need to be addressed. From the design point of view, these issues normally require some design development. In this process, the objectives and the design may have completely different nature in the different microelectronics applications.

Based on the above observation, it is very important to develop a general design approach for microelectronics applications which is not problem dependent. This approach has to take advantage of the numerical optimisation capabilities to navigate the design process towards the optimal solution (in an automated way) and from simulation-based prediction results used to evaluate virtually the design aspects of interest. Thus, our intention is to base the

design process on a coupled *simulation-optimisation* software framework which can provide a tool for automated design process using the power of the modern computer.

The general design optimisation framework requires the integration of a simulation tool – in our case based on multiphysics Finite Element Analysis (FEA), and optimisation modules for constrained and unconstrained numerical optimisation. This design tool will provide the capabilities to formulate practically any design problem as an optimisation task and to solve this task using computer models for analysis. To obtain the intended software environment for these design purposes, software modules for finite element analysis and optimisation with appropriate interfaces for integration are required.

The rest of this chapter explains the integrated approach and the software environment used to develop a general design tool. The framework is demonstrated in the following chapters for a number of microelectronics applications.

4.8.2. *VisualDOC – a Flexible Optimisation System*

The *VisualDOC* [143, 161, 162, 165] (Visual Design Optimisation Control program) is a general purpose optimisation tool with modules for numerical non-linear optimisation and statistical analysis. In the following sub-sections some aspects of the optimisation system *VisualDOC* are detailed.

The design process for any system or process starts with a definition of the design project. Initially, all aspects of the design are observed carefully, and the major objectives and the design parameters that can vary are stated. These initial decisions will also involve observations for all important requirements that the developed design must obey. The design project specification then can be entered in VisualDOC in terms of definitions of the design variables, objective function and constraints. All the data related to an optimisation task and its formulation is stored in the VisualDOC design database. Once the design task is specified and completed, the problem can be solved using a particular optimisation module from the VisualDOC optimisation library. After obtaining the solution of the design task, optimisation results are stored in the database and can be used for post-processing activities or for further processing. The task can be solved again using a new functional module and the same or different control optimisation options. The optimisation module provides options to restart

the optimisation process from the last optimal design, which is a useful feature to refine an available solution.

The VisualDOC capabilities can be used not only to find the solution of an optimisation problem, but it provides also the opportunity for understanding design features and the design variable-objective relationship through the post-processing modules.

4.8.2.1. VisualDOC System

The integration of the collection of software programs and libraries in VisualDOC to an analysis program will add design optimisation capabilities to that simulation tool. This integrated environment will provide the whole functionality for a general design discussed previously – it can be used to analyze and predict a system response to the imposed conditions, and to explore the generated data from the analysis to calculate optimal designs and to perform design studies. The integrated design tool can explore the design space, to test "what-if" scenarios, etc. The *PHYSICA* multiphysics finite element code, introduced in Chapter 2, is the analysis tool linked to the *VisualDOC* optimisation routines.

The components in VisualDOC are the database, the design modules, the Graphical User Interface (GUI) and the response program interface used to link a 3rd Party Analysis. Figure 4.11 demonstrates the interaction and relationship between the first three components.
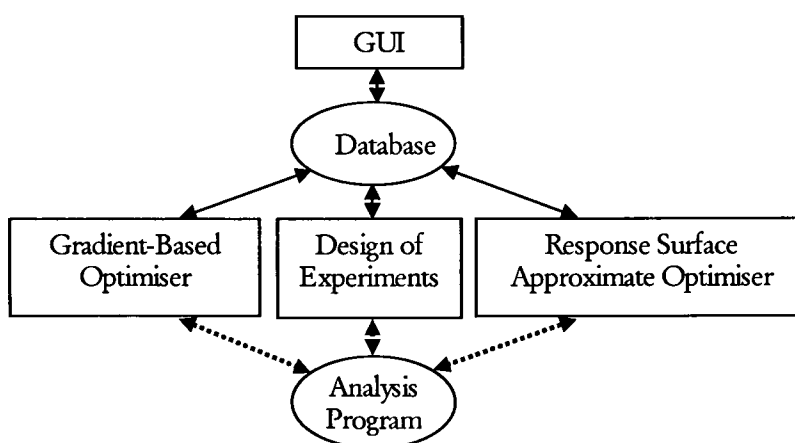
Figure 4.11: VisualDOC System

- **Graphical User Interface (GUI).** The GUI provides user friendly capabilities for design information pre- and post- processing. The definition of the design problem using the GUI is based on convenient spread-sheet format. When an optimisation task is defined, the related data about the analysed system in terms of finite element analysis design inputs (design parameters that can vary) and related output parameters are specified. For example, consider an electronic package to be analysed with respect to solder joint life time and die stress under thermal cycling, and let the package parameters allowed to vary be the underfill Young's modulus (E) and the CTE. Then, the first step of the design task definition in VisualDOC is to define the E and CTE as input parameters (the **Inputs** spreadsheet window) and the die stress, solder plastic strain and lifetime as response parameters (the **Responses** spreadsheet window). The type for both the input parameters (e.g. continuous, discrete, integer) and the responses (independent or synthetic) must be specified. A response is an independent parameter if its value is obtained directly as a calculated quantity from the FEA, e.g. the stress in the die or the solder plastic strain. The synthetic responses are defined as explicit functions of the inputs or responses in the Synthetics spreadsheet of VisualDOC, i.e. the solder life time is calculated from an explicit relation using the plastic strain value.

The next step in the design task definition is to specify the design variables, constraints and the objective function from an optimisation problem prospective. This specification will also require input of their specific descriptions and control characteristics. For example, the design variables can be equivalent to the input parameters detailed above, i.e. underfill E and CTE. Upper and lower limits can be entered. The objective is to maximise the lifetime of solder joints (given by the synthetic response) subject to upper stress limit for silicon die stress.

The information entered through the GUI is stored in the database and used for later processing. Apart from the design/optimisation task definitions, the graphical environment allows the user to set the design task (e.g. gradient-based optimisation), and algorithm control parameters (e.g. Modified Method of feasible directions + convergence tolerances), to run the optimisation tasks and to post-process the design results (e.g. the history of the objective improvement). Figure 4.12 shows a view of the graphical user interface.
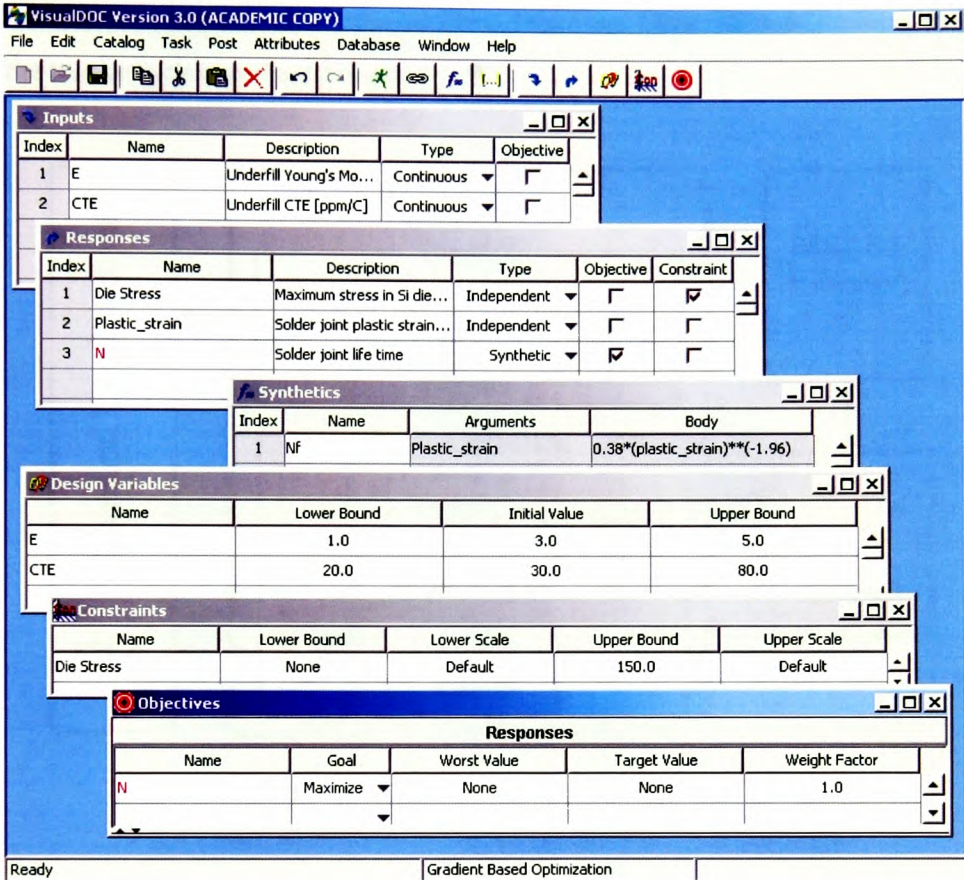
Figure 4.12: VisualDOC GUI

- **VisualDOC Database.** The database contains all design optimisation data and acts as a primary communication interface between the components.

The information for each design problem is stored in its own database, but may contain many separate tasks. The database consists of three parts — the interface section, the task definitions and the task results section. These are shown in Figure 4.13.

Figure 4.13 also illustrates also the interaction scheme between the design modules and the database. The interface objects in the database are created through the VisualDOC GUI. The same GUI helps to post-process the design results. Inside the database, all three components interact and transfer information. The design modules interact with

the task definitions to obtain the problem definition. Once the task is solved, they store the results back into the database.
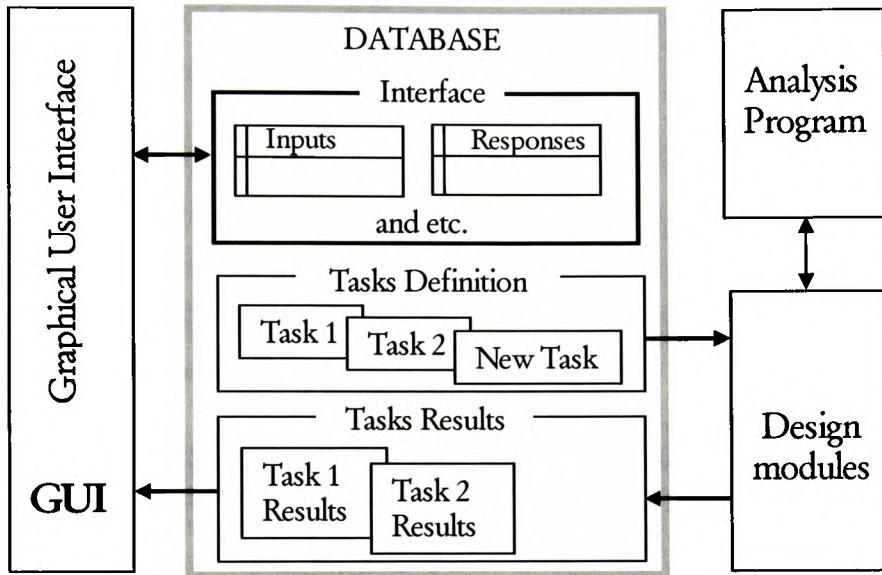
Figure 4.13: Interaction between Components and Database

The design modules and the database are the VisualDOC parts that interact with the analysis program. Details on the interacting procedure between *VisualDOC* and the analysis program *PHYSICA* will follow.

4.8.2.2. Design Modules

Design modules encapsulate the optimisation, approximation and statistical algorithms. Most of the numerical algorithms and techniques for Response Surface approximations and DOE available in the VisualDOC design modules were explained in Chapter 3. Optimisation related methods were detailed earlier in the present chapter.

o *Gradient-based Optimiser*

The *VisualDOC* optimisation routines are based on numerical algorithms that require the gradient information for each of the observe responses (i.e. the objective and constraint functions). If no gradients are provided, the optimiser calculates them using the finite

difference approach (Section 4.3.2). If the gradients for a particular problem can be calculated analytically, then they can be passed to the *VisualDOC* optimiser. In this case, the finite difference approach will be omitted. For computationally expensive analyses supplying the gradients reduces the overall time to perform the optimisation process. The error due to the approximated gradients is also eliminated.

The collection of the optimisation algorithms inside the *VisualDOC* tool offers a numerical robustness of these methods which make them applicable to wide range of applications. Six different numerical optimisation algorithms are available to perform the gradient based optimisation (Figure 4.14). These methods have been extensively tested and developed, and proven to perform efficiently.
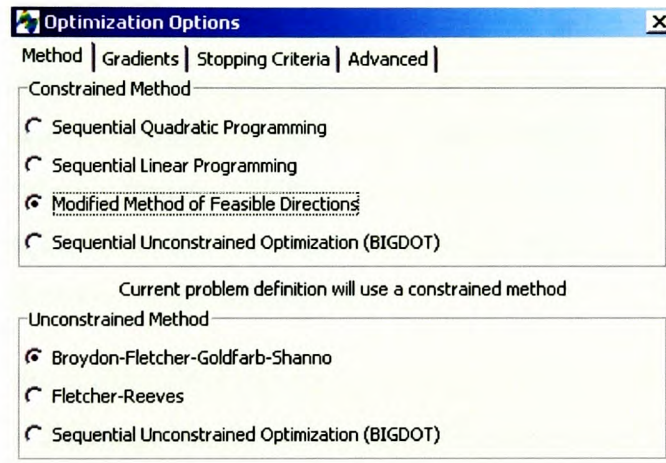
Figure 4.14: Optimisation Algorithms

For the constrained optimisation, the method options include the Modified Method of Feasible Directions (MMFD), the Sequential Linear Programming (SLP) and the Sequential Quadratic Programming (SQP). For the unconstrained optimisation, the Broydon-Fletcher-Goldfarb-Shanno (BFGS) and the Fletcher-Reeves algorithm can be selected. The design variables can be defined as continuous, discrete or integer. Apart from the tasks with a single objective function, tasks with multi-objective optimisation are also possible.

Efficient optimisation algorithms for parallel applications [163] and very large scale optimisation [164] are also developed in VisualDOC. For large scale optimisation problems, the Sequential Unconstrained Optimisation method (BIGDOT) is the recommended option

for unconstrained and constrained problems. For the latter case, the problem is transformed initially to an unconstrained optimisation task.

o *Response Surface Approximate Optimisation*

The Response Surface (RS) modelling in *VisualDOC* utilizes the polynomial approximations and related RS analysis. As emphasized in Chapter 3, this approach is effective for relatively small number of design variables. The RS models can be constructed for each of the responses specified in the optimisation problem. These approximations can be used by the gradient optimiser afterwards as an alternative approach to the FEA evaluations of the objective and constraint functions during the optimisation process.

o *Design of Experiments*

The *VisualDOC* has a functional module to perform Design of Experiments (DoE) analysis. The DoE can be used as a separate program for statistical analysis of the design or can be used as a tool to construct Response Surfaces and to perform DoE based RS optimisation. As described in Chapter 3, the design of experiments is a powerful tool to identify important design variables and/or to construct RS models.

Figure 4.15 shows the available DoE models for response surface construction, related DoE executable actions and the experimental designs implemented in *VisualDOC*.
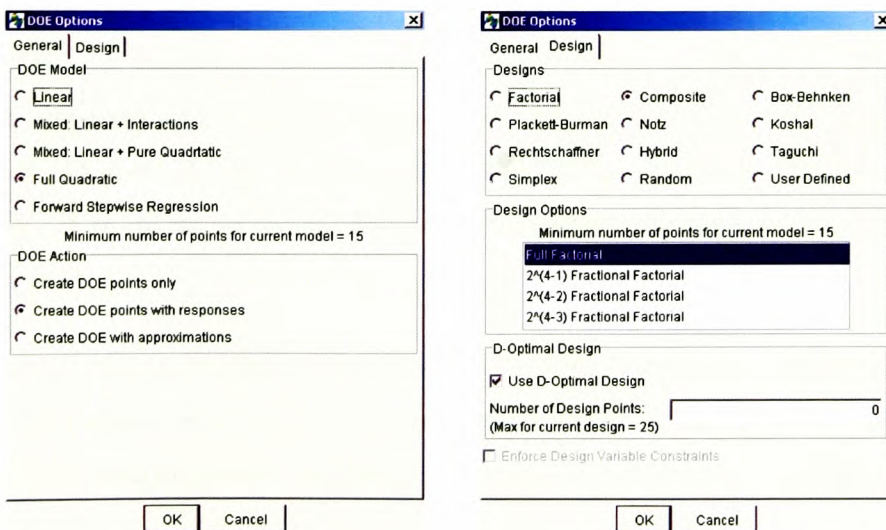
Figure 4.15: DoE Models and Designs Methods

4.8.3. *PHYSICA-VisualDOC Integrated Framework*

The *PHYSICA* finite element software tool can be referred to as a *standalone* program. It has an open source structure and the user has access to the software modules so he or she can implement his own code. This is a highly beneficial feature of the analysis tool as it allows the user to link the optimisation routines of *VisualDOC* to the simulator program in a relatively straightforward manner. The software integration is implemented using *link routines*. These software coupling routines are implemented inside the *PHYSICA* and are intended to manipulate the optimisation data transfer between *PHYSICA* and *VisualDOC*. The *link routine* module reads and writes the transfer data into text ASCII files accessible by both the optimiser and the simulator.

The first ASCII file with a default name *dvar.vef* passes the design variable values from *VisualDOC* to the *PHYSICA* analysis program. Each of the rows in this file contains the value of one particular design variable. The design variables appear in the file in the same order as they are defined in the **Inputs** window of the design project when they are initially defined through the *VisualDOC* GUI. All the entities from the **Inputs** parameters window, i.e. all design variables, must appear in the *dvar.vef* file.

Similarly, the *PHYSICA link routine* module has to write the relevant design response values from the FE analysis into another ASCII text file. The default name of this file is *resp.vef.* The data in this file is read by *VisualDOC*. Each of the lines in this file contains the value of one response (e.g. maximum stress in the die, creep energy in solder, etc.). The order in which the responses appear in the *resp.vef* file is exactly the same as they appear in the **Responses** window of the design project (the design problem definition through VisualDOC GUI). All the entities detailed in the **Responses** window have to be presented in the *resp.vef* file.

Figure 4.16 provides an example of the *dvar.vef* and *resp.vef* files for a design problem. The design variable values, being input parameters for simulation, are written in the file *dvar.vef.* The response values of interest - die stress, plastic strain and life-time, are listed in the file *resp.vef*
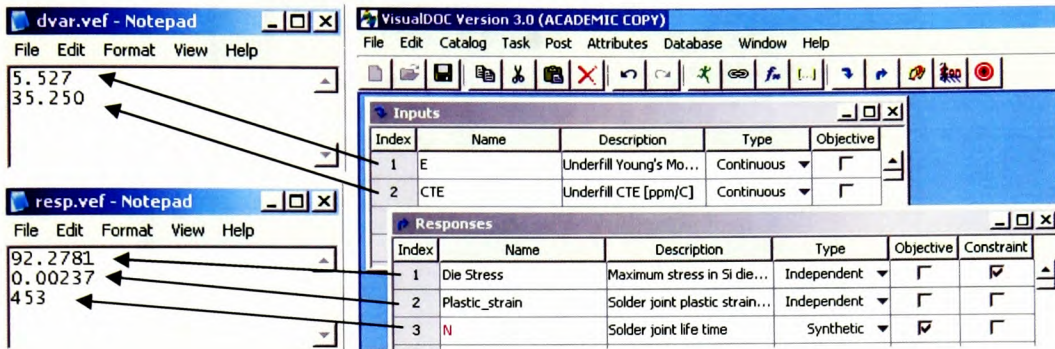
Figure 4.16: Example of *dvar.vef* and *resp.vef* files

This approach assumes a direct transfer of the numerical values of the design variables from the optimiser to the analysis tool, i.e. the optimiser supplies the simulator with the designs that need to be evaluated. Similarly, the observed design responses from the analysis tool are passed back to the optimiser, so the optimiser can evaluate quantitatively the corresponding design point from an optimisation point of view. The major efforts to couple the PHYSICA and VisualDOC are concentrated in obtaining and post-processing the desired analysis output results. Normally, this is implemented within the PHYSICA source code and requires some software development in terms of adding user code as appropriate.

Optimisation is entirely dependent on the accuracy of the FE model representation and the level of capturing the real behaviour of the simulated process. Reliable analysis outputs are critical for a successful and representative design optimisation process. The results on the optimal design are relative and must be observed only with respect to the design task definition, i.e. what objective and constraints have been applied and used as a base for the design optimisation.

4.8.4. *Optimal Design Procedure and Design Process Strategies*

The optimal design procedure in the coupled *PHYSICA-VisualDOC* software tool is represented in the Figure 4.17 flowchart.
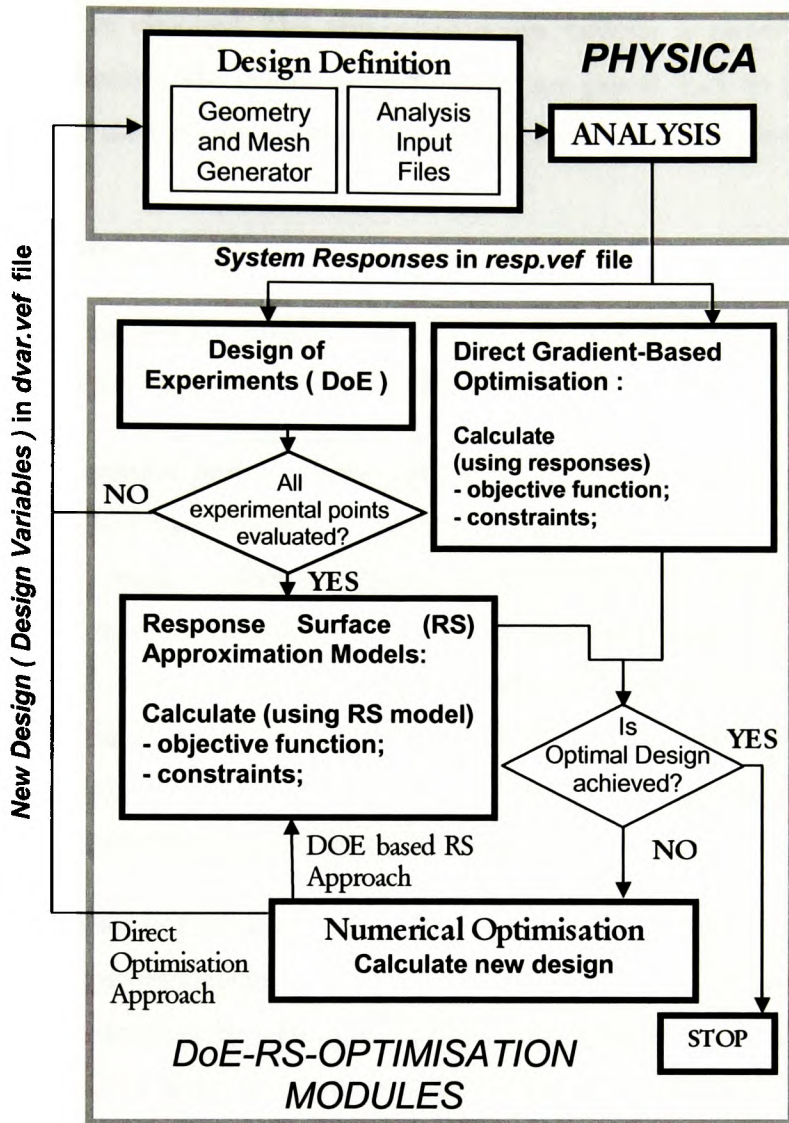
Figure 4.17: Optimal Design Procedure and Software Coupling

4.8.4.1. Gradient-Based Design Optimisation

The Gradient-Based (Direct optimisation) design module is a first order search approach. It evaluates the objective function and the constraints for a design point based on the system responses from the analysis tool (*PHYSICA*). Numerical optimisation search techniques rely on an iterative improvement of the design as described previously. The procedure starts with an initial design and converges to the optimal design. If the currently evaluated design is not the optimal one, then the numerical optimisation search technique proceeds and a new

design point will be observed. This new set of design variables is passed to the analysis program for evaluation. The design response values are passed back to the *VisualDOC* design module and used to carry out optimisation (Figure 4.17) at the next iteration.

4.8.4.2. DoE Analysis and RS

The Design of Experiments module creates an experimental design based on the design definition method and the specified DoE controls.

Each of the experimental points is then passed to the *PHYSICA* analysis program and the executed simulation is used to evaluate that design point. The FE responses are returned back to the DoE design module and stored on the database for later processing. This continues until all experimental points are evaluated as shown in Figure 4.17. This is a single processor approach. After that the experimental points and the related FE responses can be used to construct the RS model. Once the RS models the responses of interest are available, they can then be used to run a sensitivity analysis or to perform design optimisation using the explicit response approximations (no FE analyses required).

The DoE methodology allows distributed usage of the available computer resources. Instead of evaluating experimental points in sequence one by one, they can be passed to different processors and evaluated at the same time. This is a great advantage of the approach since it allows the user to safe design time and to obtain design results in shorter period of time. Figure 4.18 illustrates distributed CPU's approach in DoE analysis.
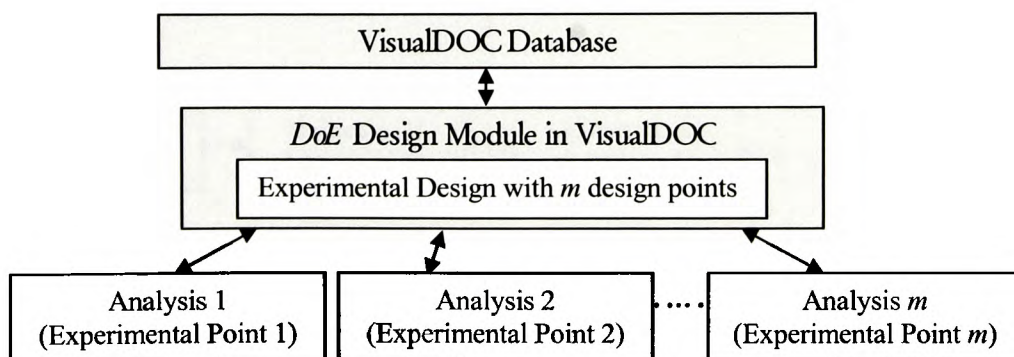
Figure 4.18: Distributed Computer Resources with DoE

4.8.4.3. Incremental Response Surface Modelling (IRSM)

The IRSM starts with a set of initial design points (e.g. based on DoE) sufficient to construct a linear response surface approximation. After that, at each of the subsequent design cycles a new design point is added and the FE responses for that point are obtained. The design point is then used to update the response surface approximation. As the process continues, quadratic terms in the polynomial response surface model can be specified. This approximation refinement carries on until, for example, a full quadratic model is determined. If the optimisation process continues after this point (of having all possible quadratic terms in the approximation) the addition of more points leads simply to better determination of the terms in the already available response surface model.

At each of the design cycles, the current response surface is used to perform the optimisation. The optimisation strategy applied can be, for example, the gradient-based or any other optimisation approach that can take advantage from the special form of the approximation (e.g. linear, and later on quadratic). The optimisation process provides an optimal point for the particular RS approximation and the corresponding optimal approximate objective value (Obj_A). At this current optimal point the analysis code is run and the FE results are used to evaluate the exact value (Obj) of the objective for this design point. If Obj_A and Obj are close enough within the specified relative and absolute tolerances, the process of refinement of the RS approximation (and respectively optimisation process that uses this approximation to find the optimal design) terminates. Otherwise, this current optimal design is the point added in the subsequent design cycle in the approximation model (Figure 4.19).
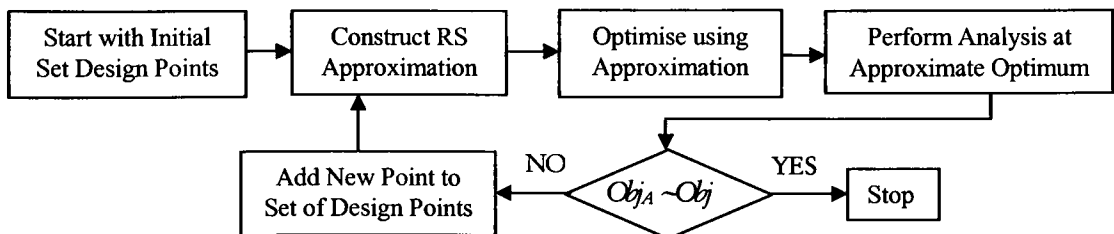
Figure 4.19: IRSM within Integrated Design Strategy

The IRSM approach has some advantages comparing with the classic use of the RSM. It offers an ordered and clever distribution of the design points near the optimum. As a result, the approximation has higher accuracy in this region and the optimisation process is capable of locating the best design using relatively low number of design evaluations.

4.8.5. *Design Optimisation Packages*

The *VisualDOC* optimisation and DoE routines are chosen to be integrated with the *PHYSICA* analysis framework because the algorithms available in the package are robust and proven to perform very efficiency. The "easy-to-couple" structure of optimisation modules is an advantage that was also taken into account.

Different design optimisation packages exist and are commercially available to the design engineer community. Two software tools with similar functional capabilities as VisualDOC will be mentioned here. First, the COMPACT [148] design tool is developed by the CQM BV in the Netherlands and used for design purposes by Philips Research/ Philips Semiconductor for the implementation of their virtual prototyping design strategy [83-85]. The CQM package is also used with FLOTHERM, the thermal analysis software by Flomerics Ltd. It provides the optimisation routines for their CFD tool FLOTHERM as part of the sequential global optimisation methodology [160].

The second package that incorporates the DoE, RSM and optimisation routines for design purposes, mentioned here, is the LMS OPTIMUS [144] developed by LMS Int. The LMS OPTIMUS is an environment for exploration of the design space to gain the critical insights into the dynamics of the problem. The OPTIMUS package can be used to find the optimal design using a combination of gradient methods for a quick analysis or the novel non-gradient genetic algorithms.

CHAPTER 5

EXPERIMENTAL AND MODELLING PARAMETRIC ANALYSIS OF NO-FLOW UNDERFILL FLIP-CHIP ASSEMBLY

This chapter presents a comprehensive study of important aspects related to a novel bonding technology known as *no-flow underfill packaging*. This study was performed in close collaboration with the City University of Hong Kong. The results from this project were disseminated in the 4-th ISEPT Electronic Packaging Symposium (2001) [166] and also contributed to a *Microelectronics Reliability* journal publication [167].

The study combines experimental and modelling analysis using the finite element method. It adopts the traditional approach of problem investigation for different "what-if" scenarios using computer simulation tools (methodology known as the *parametric study* approach).

The route towards improvement of the investigated packaging technology is based on FE analysis predictions on important characteristics and conditions by verifying the packaging process input parameters of interest. The no-flow underfill study provides invaluable knowledge on the investigated problem in different aspects and helped to get insight the novel process. The parametric study described in this chapter is performed without using any of the statistical or optimisation techniques available through the integrated *VisualDOC-PHYSICA* design tool. The idea was to provide a base for comparison between the

traditional and widely used parametric study analysis and the proposed novel simulation-based optimisation approach for virtual design process (demonstrated in the next chapters). This chapter also provides some background and discussion on the major concerns regarding reliability of the electronic packages. In particular, the fatigue failure of solder interconnects under thermal cycling is detailed.

5.1. Introduction to Flip-Chip and No-Flow Underfill Technologies

The face-down attachment of silicon dies to organic substrates, known as *flip-chip assembly*, has a number of important advantages, compared to traditional surface mount assembly: greater performance, low cost, high I/O number and distribution, low weight, etc. [34]. While this technology is continuously driven to satisfy the market demands for faster, smaller and cheaper products, the reliability of such components is a major concern for the electronic manufacturing companies. During the component testing under thermal cycling, high levels of stress and strain develop across the solder joints due to the extreme thermal miss-match (*CTE* difference) between the materials used in the package. As a result, the solder joints degrade over time and are very likely to fail due to thermal fatigue [34]. The success of flip-chip relies in part on improving thermo-mechanical reliability offered by applying underfill encapsulation technology in electronic assemblies [34, 35]. The method commonly used to dispense underfill into the gap between the die and substrate is based on the capillary flow (surface tension phenomenon) technology.

Figure 5.1 illustrates both the *traditional* (fast flow) and *no-flow* underfill flip-chip assembly approach. The advantage of the no-flow process is that it needs a fewer number of processing steps.
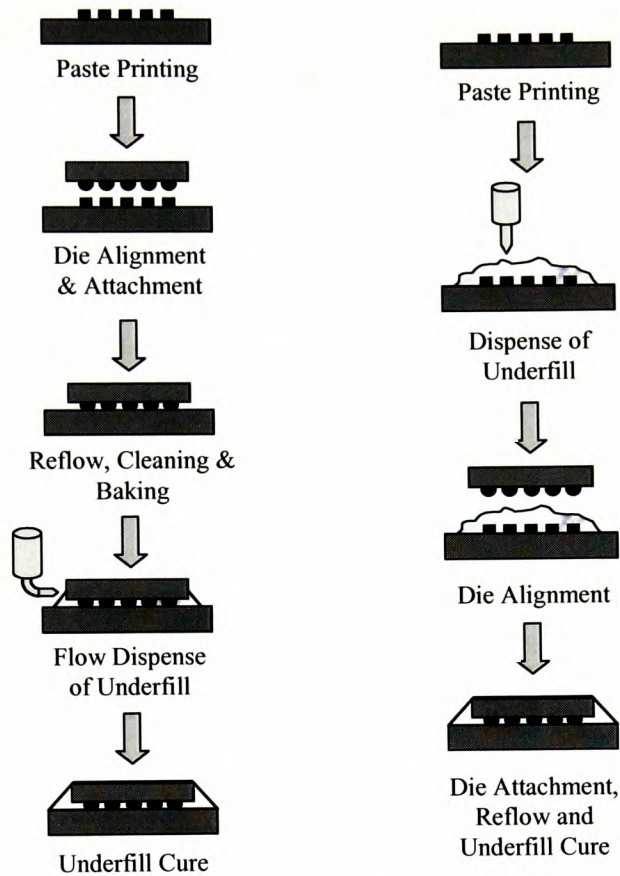
Figure 5.1: Underfilling Processes

The application and curing of the no-flow underfill materials before and during the reflow process in the flip-chip assembly process is a particular and important advantage over traditional underfill. This advantage can be exploited to increase the flip-chip manufacturing throughput. However, adopting a no-flow underfill process may introduce reliability issues such as underfill entrapment, de-lamination at interfaces between underfill and other materials, and lower solder joint fatigue life. The modelling study here focuses on the assembly process and the reliability of flip-chips with no-flow underfill [166, 167, 176].

In the flip-chip assembly process underfill materials are applied to meet the reliability requirement for the product. The application and curing of the underfill material for flip-chips are carried out traditionally after the reflow process. Compared with other surface mount assembly procedures, this is an additional process that compromises high throughput

and low cost production. Recently, the so-called *no-flow* underfill process has been invented [168]. This new process enables underfill addition and curing to occur before and during the reflow process. To ensure that the solder material connects to the pads on the substrate, these underfill materials normally do not contain particulate filler material and hence have a higher coefficient of thermal expansion (CTE) than those used in the traditional capillary flow process. This may lead to lower solder joint reliability and thus create a technical barrier in adopting the no-flow underfill technology [169]. The characterization and development of no-flow underfill encapsulants as a practical approach towards real manufacturing applications of this technology is discussed by Wang *et al* in a number of publications [170, 171].

Traditional underfills contain an epoxy with added inorganic silica fillers [172]. These materials have a CTE value in the order of 30 ppm/°C. Unfortunately these underfills can not be used in a no-flow application as the filler particles will hinder contact between solder and pad [172].

Recently, research work at City University Hong Kong [173] has studied the reliability of a no-flow underfill μBGA assembly using thermal shock and cyclic bending tests. The results have shown that this no-flow underfill can greatly enhance the mechanical fatigue lifetime, but cannot improve the thermal reliability of μBGA assembly because the CTE of no-flow underfill is high. Modelling predictions a study on flip-chips with two different no-flow underfill materials in the assembly have been used to analyze the solder joint quality and its resistance to cyclic thermal shocks.

5.2. Experimental Data

The flip-chip sample preparation using no-flow underfills and the subsequent interconnects reliability test were undertaken at the City University of Hong Kong [173]. The modelling work at the University of Greenwich [166, 167] was intended to provide knowledge on important issues related to the successful implementation of this novel no-flow underfill technology. The experimental data was used to validate the modelling predictions on the influence and sensitivity of different package parameters with respect to reliability and life time of solder interconnects.

5.2.1. *Flip-Chip Assembly – Sample Preparation*

Table 5.1 gives the geometric data for the flip-chip component being investigated. For both the traditional and no-flow underfill materials the geometric data is the same except for the measured stand-off height (obtained from experiments).

Table 5.1: Geometric Parameters of Flip-Chip Package

FLIP CHIP GEOMETRY DATA	
Die dimensions (mm)	6.3 x 6.3 x 0.6
Standoff height (10^{-3} mm) (Traditional Underfill Model)	118
Standoff height (10^{-3} mm) (No-Flow Underfill Model)	130
Substrate thickness (mm)	1.5
Pad diameter (10^{-3} mm)	150
Cu-Ni pad thickness (10^{-3} mm)	35
Bump pitch (10^{-3} mm)	457
Number of bumps	48

The flip chip package has a bump diameter and pitch of 190 and 457 μm respectively, and the solder bump material used is eutectic 63Sn/37Pb. The test board for flip chip assembly is composed of FR-4 substrate. The metallization on the substrate is plated copper with nickel and gold flash finish. The diameter of the test board pads is 150 μm.

A daisy chain design is used to test electrical conductivity through all solder joints. Figure 5.2 shows the layout of the daisy chain. The flip-chip package analyzed in this study has 48 peripheral interconnects connecting the pads on the die to the pads on the substrate.
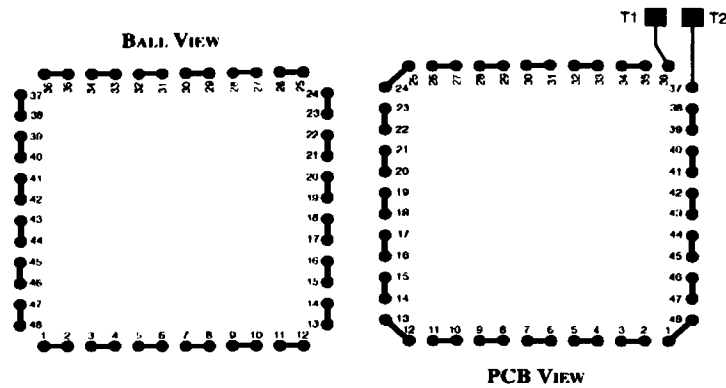
Figure 5.2: Daisy Chain

The properties of the flip-chip assembly materials are summarized in Table 5.2. There are two categories of no-flow underfill materials, NF1 and NF2. Both of them do not contain any filler particles. The no-flow underfill properties are detailed in Table 5.3. Both no-flow underfill materials (NF1 and NF2) are cured during a modified reflow cycle and no additional post-cure is required.

Table 5.2: Flip-Chip Material Properties

	CTE [$10^{-6}/°C$]	E [GPa]	ν
Substrate	18	22	0.28
Cu pad	17	121	0.35
63Sn/37Pb Solder	21	10	0.40
Si (Die)	2.8	131	0.30
Traditional Underfill	30	6	0.35

Table 5.3: Detailed Properties of No-Flow Underfill Materials

Property	NF1	NF2
Glass transition temperature Tg (°C)	170	100
CTE ($10^{-6}/°C$)	70 ~80	90
Storage Modulus @ 25 °C (GPa)	2.5 ~3.0	2
Viscosity (Pa.s)	4 ~6	20
Moisture absorption (%)	3.7	<1.5

The work associated with the assembly of chip samples, reflow process and finally the thermal shock test was carried out by the experimental team at the City University of Hong Kong. The test boards were baked at 150°C for 1 hour before assembly so as to remove the risk of moisture absorption. The substrate and chip were placed and aligned using a Karl Suss Manual flip-chip bonder. Accurate volume of no-flow underfill was then dispensed by syringe onto the bond site to form a dome-shaped underfill layer using a LY9801D dispensing machine. The chip was then placed onto the substrate. The bonded samples were then reflowed using a 5-zone air convection oven (BTU VIP-70N) in a compressed air environment. The time-resolved temperature during reflow between the component and the test board was then measured using a wireless profiler (Super M.O.L.E, E31-900-45/10). An optimized "rapid ramp" reflow profile was used as shown in Figure 5.3.
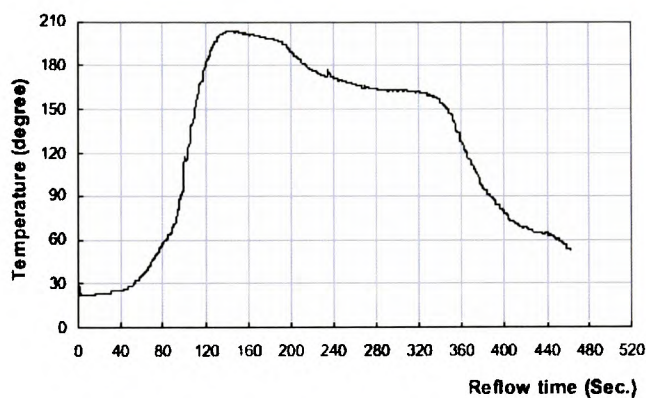
Figure 5.3: Measured Reflow Profile

After reflow, the no-flow flip chip assemblies are inspected with an X-ray inspection system and Scanning Acoustic Microscope (SAM) for voids, de-lamination and other defects. Solder joint size and standoff height were obtained by cross-sectioning a few samples.

5.2.2. *Thermal Shock Test*

In order to investigate the effect of no-flow underfill on the fatigue failure of flip chip assemblies, samples with two different no-flow underfills, NF1 and NF2, were prepared at the City University of Hong Kong. These samples were subjected to 1000 thermal shock cycles and were performed in an air-to-air condition in a TABAI TSA-70L thermal shock

chamber. The temperature range of the thermal cycle was from -25 to 125°C and the samples were held at each temperature for 15 minutes. Eleven samples for each no-flow underfill are used. Electrical conductivity of the no-flow flip chip samples was monitored throughout the reliability test. The detection of open solder joints through electrical interruption is used as the failure criterion [173]. The number of cycles to failure of solder joints was recorded. Finally, failed joints are analyzed using scanning electron microscope (SEM).

5.3. Computational Modelling

The methodology adopted to study the above no-flow underfilled packages is a combination of experimental and computer-modelling methods. A comprehensive parametric study approach is intended to provide details on the relationship between the material properties and reliability, and on how underfill entrapment may affect the thermo-mechanical fatigue life of flip-chips with no-flow underfill.

Computer modelling analysis predicts the evolution of thermal stresses and strains in the whole package during a thermal cycle. The creep strain values from the analysis are then used in an appropriate lifetime model to estimate the number cycles to failure (lifetime). The analysis software package PHYSICA [118] is used to model and predict the performance of the no-flow underfilled packages. None of the integrated statistical or optimisation routines are used. Through a wide range of Young's Modulus, E and CTE values the flip-chip models are investigated to indicate relationships between flip-chip characteristics and performance. Solder joints with trapped underfill are also investigated.

5.3.1. Analysis Specifications and Creep Model

The computer modelling analysis calculates the stress and strain profiles across the package during a thermal cycle. The cyclic temperature profile observed in this analysis consists of four stages: ramp down, hold at lower temperature, ramp up and hold at higher temperature. Each of these stages lasts 15 minutes. The minimum and maximum temperatures are –25 and 125 degrees respectively.

Because of its low melting point (T_m=183°C), the Sn-Pb eutectic solder normally operates at high homologous temperature (*HT*), typically $HT > 0.5$. The homologous temperature is dimensionless quantities given by the ratio between the operating (or working) temperature T_o and the melting point temperature T_m ($HT = T_o/T_m$). In such circumstances the major solder deformation mechanisms are the time dependent plasticity and creep. A *sinh* strain rate equation (Equation 5.1) has been used to model the solder material creep deformation [45]. The inelastic strain rate $\dot{\varepsilon}_{ij}^{cr}$ using *sinh* equation is given as

$$\dot{\varepsilon}_{ij}^{cr} = A(\sinh(\alpha\,\sigma_{eff}))^n \exp\left(-\frac{Q}{RT}\right)\frac{3}{2}\frac{s_{ij}}{\sigma_{eff}} \tag{5.1}$$

where R is the gas constant, T is the temperature in Kelvin, s_{ij} is the deviatoric stress tensor, σ_{eff} is the Von Mises stress, and all other symbols represent material properties. For 63Sn/37Pb solder alloy the creep constants have values as follows: $A = 9.62x10^4 s^{-1}$, $\alpha = 0.087 MPa^{-1}$ and $Q = 8058.37R$ [45].

The time dependent plasticity and creep deformation are highly non-linear, thus transient analysis is used to simulate solder behaviour. In such an analysis, the time domain of a thermal cycle is divided into many time steps (a time step size of 10s is used in this modelling). The thermal load at any time step is the temperature change between this step and the previous step.

The CPU time to run a single simulation associated with the above analysis takes 5 hours 28 minutes and 30 sec on 1GHz Compaq Alpha ES45 processor. The linear solver specified in *PHYSICA* was the bi-conjugate gradient (BICG) method [118].

5.3.2. *Analysis Predictions and Life-Time Model*

The database with the analysis results is used to extract useful information about the models. At the end of an analysis the stress and strain histories are available as the output. In the present work the most important output quantity is the accumulated effective creep strain. This damage parameter is used to provide thermally induced fatigue in the solder joint.

Let us denote the incremental creep strain tensor during one time step Δt as $d\varepsilon_{\Delta t}^{cr}$. The components of $d\varepsilon_{\Delta t}^{cr}$ tensor are $d\varepsilon_{\Delta t\,ij}^{cr}$ ($i,j=x,y,z$). The accumulated effective creep strain $\varepsilon_{accum}^{creep}$ at the end of a thermal cycle is calculated from:

$$\varepsilon_{accum}^{creep} = \sum_{\Delta t}\sqrt{\frac{2}{3}\,d\varepsilon_{\Delta t}^{cr}\,d\varepsilon_{\Delta t}^{cr}} =$$
$$\sum_{\Delta t}\sqrt{\frac{2}{3}\left(\left(d\varepsilon_{\Delta t\,xx}^{cr}\right)^2 + \left(d\varepsilon_{\Delta t\,yy}^{cr}\right)^2 + \left(d\varepsilon_{\Delta t\,zz}^{cr}\right)^2 + \left(d\varepsilon_{\Delta t\,xx}^{cr}\right)^2 + 2\left(d\varepsilon_{\Delta t\,xy}^{cr}\right)^2 + 2\left(d\varepsilon_{\Delta t\,yz}^{cr}\right)^2 + 2\left(d\varepsilon_{\Delta t\,zx}^{cr}\right)^2\right)}$$
(5.2)

The sum in Equation 5.2 is over all time steps Δt completing a thermal cycle. Since stress and strain are not uniformly distributed across the solder joints, a volume weighted average value of accumulated effective creep strain in the solder $\varepsilon_{accum}^{creep}$ is calculated over a range of elements with maximum value of the quantity. This value is denoted as $\varepsilon_{sum}^{creep}$ and used to define the damage in solder joints. High values of $\varepsilon_{sum}^{creep}$ indicate lower reliability and vice-versa.

This quantity is used as a damage indicator in the Coffin-Manson empirical relationship [48, 174] to predict the number of cycles to failure (N_f). The Coffin-Manson life-time model is

$$N_f = \theta(\varepsilon_{sum}^{creep})^\alpha$$
(5.3)

It correlates $\varepsilon_{sum}^{creep}$ to life-time in terms of number of cycles to failure, N_f (reliability indicator). In Coffin-Manson relationship the θ and α are eutectic solder material related parameters ($\theta = 0.38$ and $\alpha = -1.96$) obtained by fitting the function in Equation 5.3 to flip-chip experimental data [175].

The present parametric study is over a range of Young's Modulus (E) and CTE values for the underfill and provides fatigue lifetime data for the *no-flow* underfilled flip-chip component under cyclic thermal loading.

5.4. Experimental Results

One of the major challenges associated with the no-flow underfill assembly process is the adjustment of the bonding (or placing) force of the head. Well-adjusted bonding force ensures good contact between solder bumps and the substrate pads, and the smooth compressive flow of no-flow underfill. Too little force may lead to no contact between solder bumps and substrate pads. On the other side too much force may cause serious deformation of solder bumps. Excessive force may also result in underfill trapping under solder bumps after reflow. This phenomenon is illustrated in Figure 5.4.
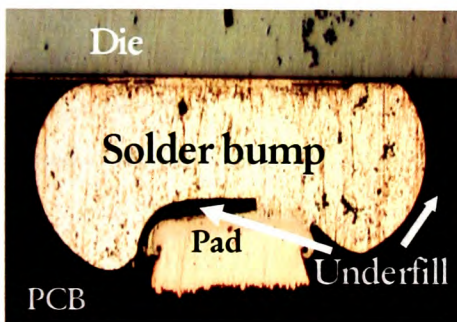
Figure 5.4: Underfill Trapping

Figure 5.5: Proper Solder Joint Bonding with no Underfill Entrapment

Figure 5.5 shows the cross-section of a no-flow flip chip solder joint, which presents good bonding between a solder bump and a substrate pad. In this case there appears to be no underfill trapped at the interface of the solder joint.

Before the thermal shock test, the samples are inspected using SAM to see whether there are any voids inside the underfill layer. Figure 5.6 shows the C-scan image from the SAM for one of the samples after reflow. It is found that most of the no-flow flip-chip samples for both NF1 and NF2 show no voids inside the underfill layer. Only a few samples contain voids (white spots) inside the underfill layer as indicated in Figure 5.6. The voids that appeared inside underfill layer in our samples are quite small and located at the centre but not near the solder joints. Normally, this type of void should not affect long-term reliability.
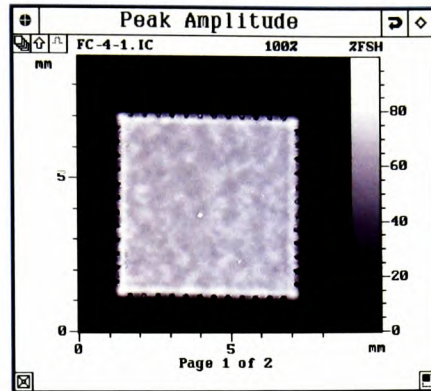
Figure 5.6: C-scan Image of No-Flow Flip-Chip Sample

The numbers of cycles to failure for NF1 and NF2 samples are summarized in Table 5.4. Eleven samples for each underfill have been used in the tests. The number of cycles to failure is defined as the lifetime of the first sample to fail. The results from the work undertaken by Wong's group (W1, W2 & W3) have also been presented in Table 5.4 for comparison [169].

Table 5.4: Material Properties and Thermal Shock Cycles to Failure of Different No-Flow Underfill Materials

Material	Tg	CTE	Cycles to failure
NF1	170	70 ~80	<500
NF2	100	90	<300
W1#	74.7	94.7	<200
W2#	152	71.5	<600
W3##	141	48.7	>1000

\# Under –55 to 125 °C thermal shock condition

\#\# Addition of filler inside the no flow underfill

The plots in Figure 5.7 detail further the experimental results from the work undertaken by the Wong's group [169]. The figure gives also details about the different underfill properties. Note, that the underfill modulus of these underfills is ranked: the higher the modulus is, the smaller the ranking number is given. Normally, lower CTE of these underfills is accompanied by a higher Young's Modulus, thus it is difficult to separate the effect of these two factors. However, with respect the interconnection reliability the graphs in Figure 5.7 indicate that lower CTE and higher Young's Modulus are desired. It is evident from these plots and from Table 5.4 that an increase in CTE correlates closer to the decrease in the number of cycles to failure, i.e. the CTE of underfill is a more influential material property than the Young's Modulus with respect to interconnects reliability.

	Underfill				
	U1	U2	U3	U4	U5
Tg (°C)	132.1	142.0	151.9	171.0	141.1
CTE (PPM/°C)	111.0	129.0	71.5	75.7	48.7
Modulus Ranking	5	4	3	2	1
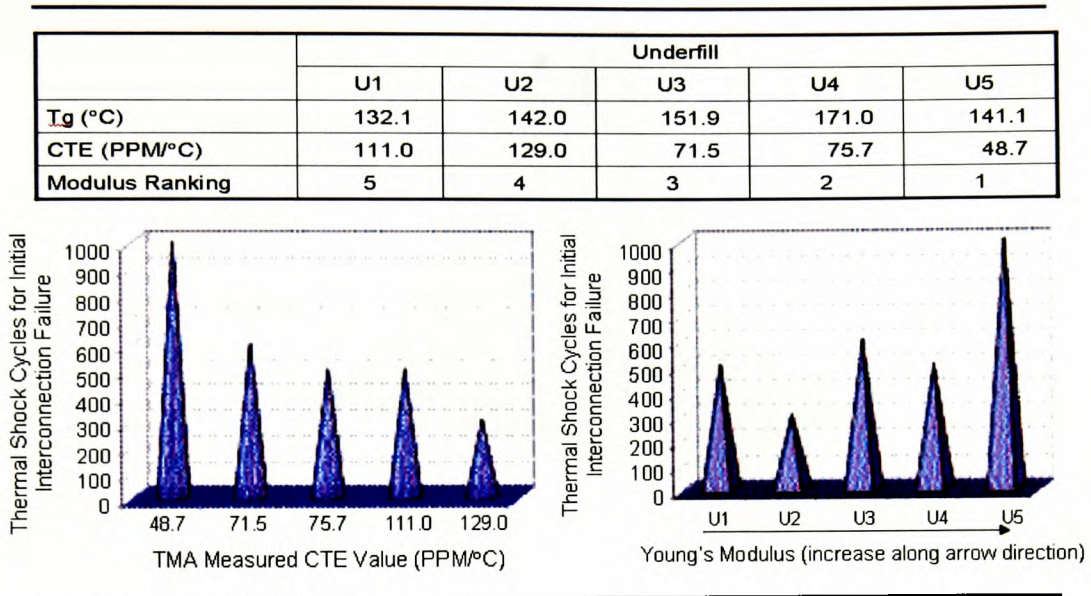
Figure 5.7: Correlation of Interconnection Reliability with CTE and Modulus
(Experimental Results from Wong's Group, Reference [169])

The trends in reliability, from the underfill properties prospective, observed through the experimental results by both the project collaborator (City University of Hong Kong) and the Wong's group were used as a base of comparing and validating the related modelling results. The next section (Section 5.5) of this chapter details the modelling results.

5.5. Modelling Results

5.5.1. Underfill CTE and E (traditional underfilled flip-chip assembly)

Figure 5.8 details a parametric reliability study results for the underfill material properties of CTE and E. The results are based on the traditional underfill model, which has a lower stand-off height than the no-flow underfill model. Within the parameter range that has been studied in this work, the highest reliability is achieved with the combination of E=9 (GPa) and CTE=15 (ppm/°C). Clearly we can see that as the CTE becomes larger and E becomes smaller (i.e. as we approach the material properties associated with a no-flow underfill) the number of cycles to failure decreases.
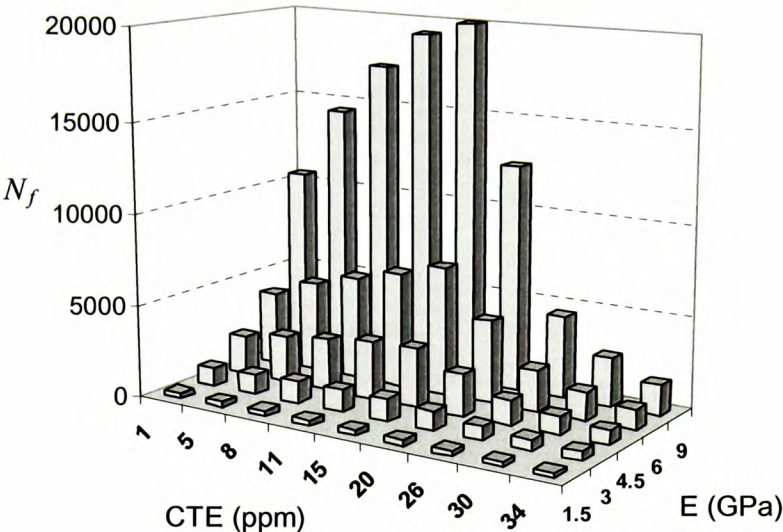
Figure 5.8: Parametric Study Results in CTE-E Design Space (Flow Underfill)

5.5.2. *Underfill CTE and E (no-flow underfilled flip-chip assembly)*

The main geometric difference between the flip-chip models for traditional and no-flow underfill is the stand-off height observed in the experiment, but the major factor resulting in a significant difference in reliability is the underfill properties. No-flow underfill typically has a much higher CTE and lower E than its flow counterpart.

Figure 5.9 details results from a parametric study for Young's Modulus of the underfill, which varies from 1Gpa to 6Gpa. This range covers possible values for no-flow underfill. The CTE of the underfill is kept constant at 80ppm, which is a possible value for no-flow underfill. Clearly we can see that increasing the Young's Modulus results in an increase in cycles to failure.

Figure 5.10 shows the results of s similar analysis in terms of CTE, which varies from 50ppm to 100ppm. In this parametric study the value of Young's Modulus was kept constant at 3Gpa. From this graph it is evident that high CTE of underfill degrades the thermo-mechanical reliability of solder joints.
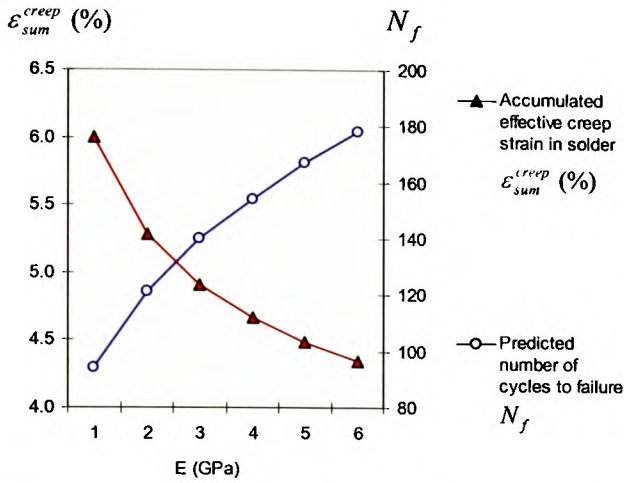
Figure 5.9: Predicted Number of Cycles to Failure and Creep Strain in Solder vs. E of Underfill (CTE fixed at 80ppm)
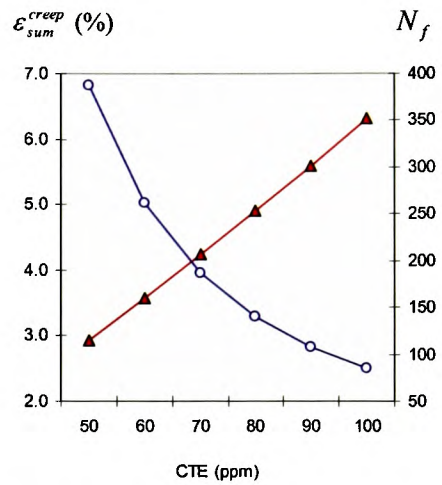
Figure 5.10: Predicted Number of Cycles to Failure and Creep Strain in Solder vs. CTE of Underfill (Young's Modulus E fixed at 3GPa)

Table 5.5 details the creep strain at the critical corner solder joint of this flip-chip package using different no-flow underfills. The underfill CTE is investigated in the range from 50 to 90 ppm/°C and the underfill Young's Modulus varies from 1 to 6 GPa. Figure 5.11 illustrates the data presented in Table 5.5 in terms of number of cycles to failure. As with the traditional underfill process we can see that increasing Young's Modulus and decreasing CTE will, in general, result in an increase in cycles to failure of this flip-chip package.

Table 5.5: Creep Strain in Solder Joint [%] as Parametric Function of No-Flow Underfill Properties

UNDERFILL	E (GPa) 1.0	E (GPa) 2.0	E (GPa) 3.0	E (GPa) 4.0	E (GPa) 5.0	E (GPa) 6.0
CTE [ppm/°C] 50.0	4.2129	3.3172	2.9202	2.6901	2.5329	2.4193
CTE [ppm/°C] 60.0	4.7678	3.9443	3.5643	3.3344	3.1716	3.0522
CTE [ppm/°C] 70.0	5.3674	4.6022	4.2295	3.9973	3.8223	3.6951
CTE [ppm/°C] 80.0	6.0009	5.2809	4.9078	4.6683	4.4812	4.3435
CTE [ppm/°C] 90.0	6.6602	5.9748	5.5946	5.3420	5.1442	4.9953

It is interesting to note that the magnitude of these cycles to failure is a magnitude less than observed with the traditional underfill material. This is due to the significant difference in the material properties that the traditional and no-flow underfills obey.
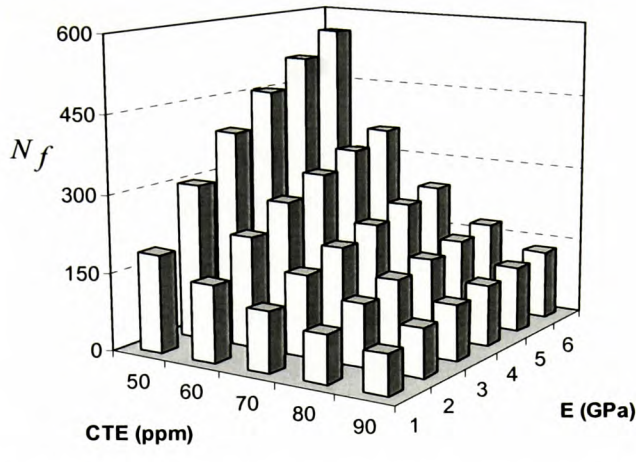
Figure 5.11: Predicted Number of Cycles to Failure N_f as Parametric Function of Underfill E and CTE (No-Flow)

5.5.3. *Parametric Study on No-Flow Underfilled Flip-Chip Geometry*

The reliability of a no-flow underfilled package is also dependent on some geometric configurations. One of the major differences between the packages assembled via the traditional fast-flow underfilling process and the novel no-flow process is in the obtained stand-off height parameter. Figure 5.12 presents what is the effect of solder stand-off height on the flip-chip reliability. The analysis shows increase in lifetime with increase of stand-off height.
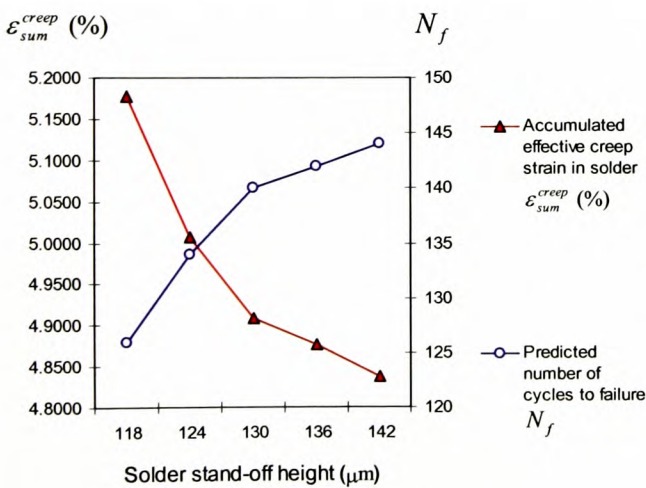
Figure 5.12: Predicted Number of Cycles to Failure and Creep Strain in Solder Joint vs. Solder Stand-off Height
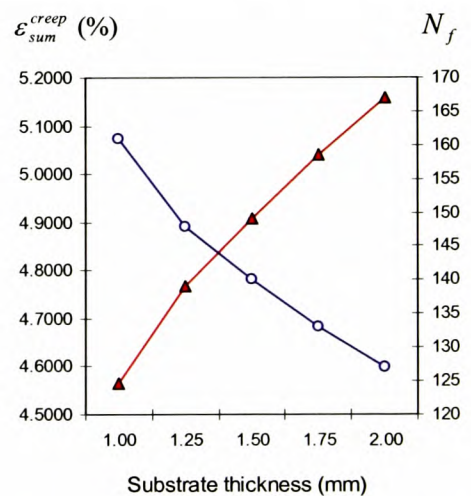
Figure 5.13: Predicted Number of Cycles to Failure and Creep Strain in Solder Joint vs. Substrate Thickness

Figure 5.13 deals with another geometry parameter, the substrate thickness. An increment of the value of this geometric parameter leads to a decrease in the solder joint life time.

The modelling results have shown clearly the trend of lifetime changes with material properties. Modelling predictions for underfill properties influence on solder joints reliability agree with the experimental observations from the project partner and Wong's group quoted in Table 5.4 and illustrated via the plots in Figure 5.7. However, quantitatively there is a discrepancy between the modelling and experimental results. One of the possible reasons is that the lifetime model is originally developed based on experimental data in significantly lower range for creep strain in solder joints.

5.5.4. *Underfill Entrapment Modelling Results*

It has been shown in Figure 5.4 that underfill may become trapped between the solder pad and solder bump. When solder joints are formed with underfill trapped between solder and pad, the stress state in the solder will change. Crack path and failure mechanism may also change too. To see how underfill entrapment affects reliability the following simulations have been undertaken. These models still assume that the failure mechanism remains that of the crack growth in solder.

Figure 5.14 (left) shows a model (Model-1) with trapped underfill. The trapped underfill has a doughnut shape with the center is still filled with solder and therefore the joint still has good electrical contact. Figure 5.14 (right) shows another model (Model-2) which is similar to Model-1 but the trapped part is isolated by solder and pad, and it is twice as thick as that found in Model-1.
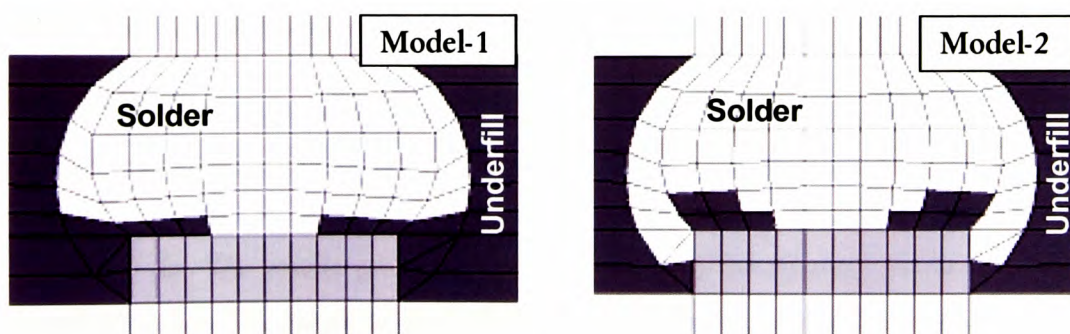
Figure 5.14: Modelling Variations for Underfill Entrapment

When these models are used it has been found that the lifetimes of Model-1 and Model-2 are 14% and 24% higher than that of the model without trapped underfill. This phenomenon can be understood by examining the creep strain distribution as depicted in Figure 5.15. For the flip-chip model without trapped underfill in Figure 5.15 (a) the maximum accumulated creep strain is close to the PCB pad. In general, material properties and solder joint geometry determine the strain distribution. It is possible that the location of the critical part may change when geometry and/or material properties have changed.

a) No trapped underfill   b) Model-1   c) Model-2
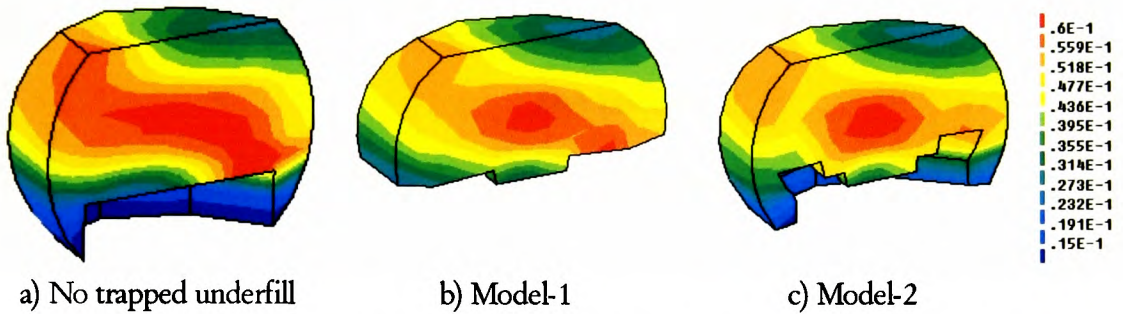
Figure 5.15: Creep Strain Distribution for No-Trapped and Trapped Underfill Models

In Figure 5.15 (b), the creep strain distribution for Model-1 is shown. It can be seen that the location and magnitude (being lower) of creep strain has changed. Therefore, due to a small amount of underfill entrapment we expect an increase in fatigue lifetime. For Model-2 (Figure 5.15 (c)), the underfill layer is thicker and this will give even higher reliability. From these results we can conclude that trapped underfill between the solder and pad can actually increase the number of cycles to failure. This assumes that the failure mechanism is the same for a solder joint with trapped underfill and one without.

5.5.5. *Modelling Results on Stress in Silicon Die*

The solder joint fatigue failure is just one of the many failure modes a flip-chip may suffer. While the accumulated creep strain determines the thermal-mechanical fatigue life of solder joints of a flip-chip, the stress level is the most important factor in causing brittle fracture of the silicon die. The results presented here are based on the nominal geometry properties described earlier. The values of underfill CTE and Young's Modulus E are 80 ppm/°C and 3GPa respectively. As shown in Figure 5.16, the maximum stress in die changes with time during the thermal cycle, at about the mid-cycle the stress is at the highest. At the end of a

full cycle only a negligible residual stress is present. In the following, stresses at mid- cycle, when stresses are near their maximum, are presented.
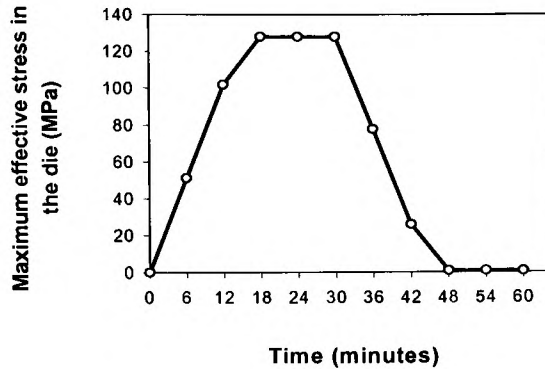
Figure 5.16: Maximum Die Stress

Figure 5.17 shows the effective stress distribution in the die of a flip-chip with no-flow underfill. On the active side and the corner the stress is in general higher than elsewhere. Figure 5.18 shows the tensile stress in the X direction (in the plane of the top or bottom surface). The stress near the bottom surface (the active side of the chip) is compressive but near the top surface the stress is tensile. The maximum is just above the corner solder joints. In Figure 5.19, the tensile stress along the Z direction (perpendicular to the top and bottom surfaces) is shown. It can be seen here that the stress peaks are at the corners of the die. These are caused by the contractions of the PCB and the underfill fillet. Because the no-flow underfill has a higher CTE than the traditional flow underfill, the high stress level at the die corner is more prominent for no-flow underfill assembled flip-chips. Since high stresses are associated with brittle die cracking the maximum stress in the die may be used as an indicator of this failure mechanism.
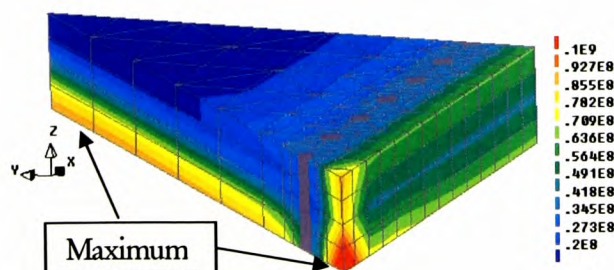
Figure 5.17: Effective Stress Distribution in Silicon Die [Pa]
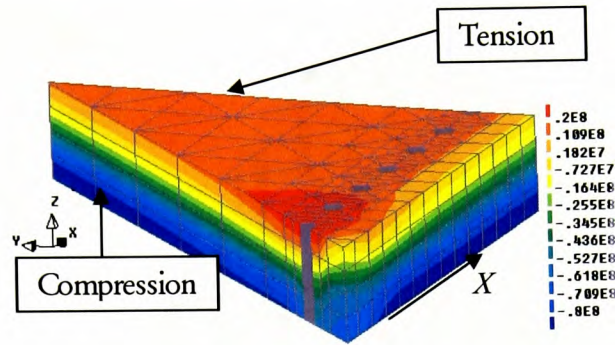
Figure 5.18: Normal Stress (*X*-direction) in Silicon Die [Pa]
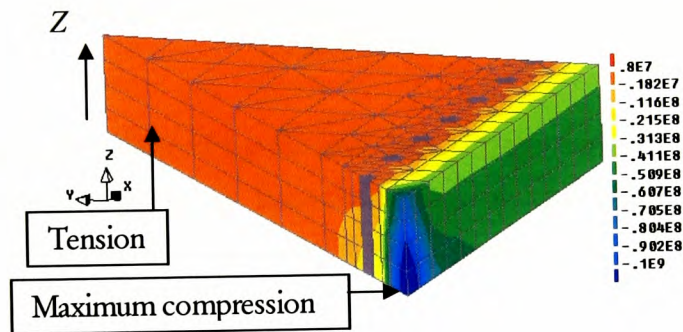
Figure 5.19: Normal Stress (*Z*-component) in Silicon Die [Pa]

The maximum effective stresses of the die (Von Mises definition), calculated as volume weighted average quantity over mesh elements with highest stress, for a range of underfill CTE and E values are plotted in Figure 5.20. It is evident from this picture that the combination of high CTE and high E results in the highest die stresses, whereas low CTE and low E keep the stresses low. If the value of underfill E is in the low range of 1-3 GPa, however, the die stress becomes insensitive to changes in underfill CTE value.
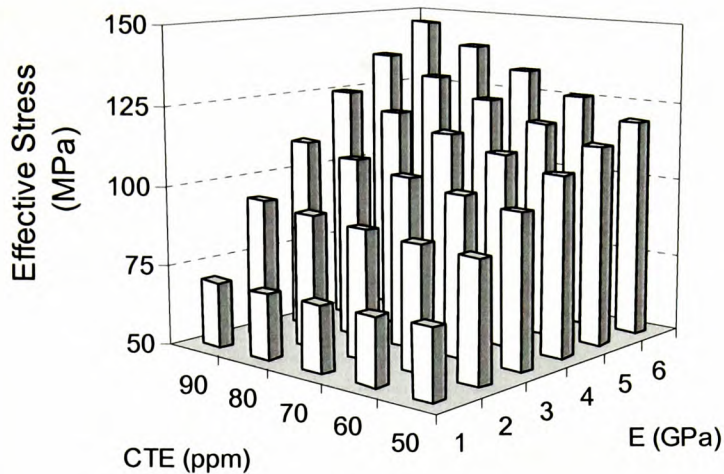
Figure 5.20: Effective Stress in Silicon Die as Function of
Underfill E and CTE

Similar trends were also observed for the maximum Z component of the compressive stress at the bottom surface of silicon die. Since the stresses with high magnitude are found at the interface between the die and the underfill, these results could be used to evaluate the possibility of a die-underfill de-lamination failure.

5.5.6. *Deformation of Silicon Die*

Other useful information from this parametric study is the knowledge of the overall deformation of the die. Because of the CTE mismatch, the contraction of the PCB and the underfill forces the die to bend when the flip-chip is under a thermal load. Figure 5.21 shows the vertical displacement distribution along the line from the die center to its corner.
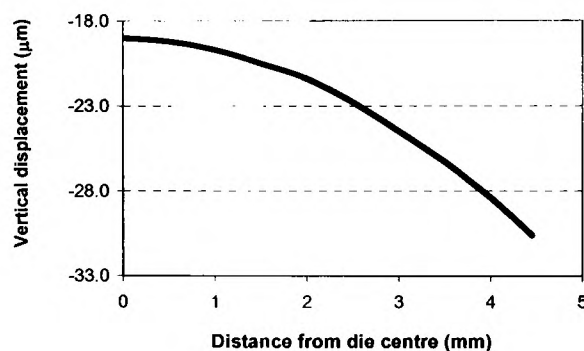
Figure 5.21: Relative Vertical Displacement from Die Center to Die Corner

The displacement of the corner relative to the center reflects the extent of die bending. This displacement is shown in Figure 5.22 as a function of underfill CTE and E. By comparing this figure and Figure 5.20, it can be seen that the highest die bending does not correspond to the highest die stress. The reason of this mismatch is that die bending only gives information about the overall global deformation whereas the maximum stress occurs at the die corner because of a local effect - the contraction of the underfill fillet.
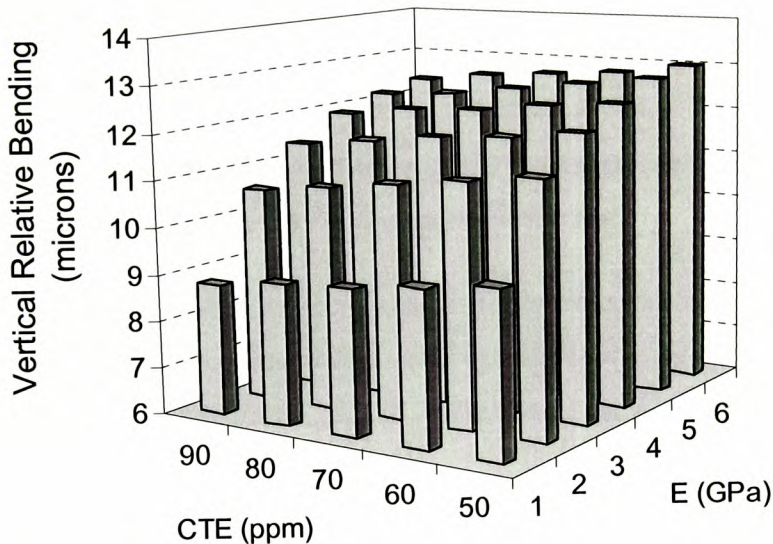
Figure 5.22: Relative Displacement of Die Corner

5.6. Conclusions on Performance of the No-Flow Underfilled Flip-Chip Package

No-flow underfill materials can be used in the flip-chip assembly process. The thermo-mechanical reliability of these packages has been investigated using computer simulations and the parametric study design approach.

The modelling studies were undertaken in close collaboration with the experimental microelectronics group at the City University of Hong Kong. The experimental tests for flip chip solder joint reliability performed at the City University were used to validate the predictions from the finite element simulations. The modelling work, on the other side, has involved the investigation of a number of flip chip design parameters and provided knowledge on different aspects of the assembly process.

Comprehensive computer modelling through parametric study of different what-if scenarios has confirmed that:

- High CTE and lower Young's Modulus of underfill lowers the fatigue lifetime of solder joints with underfill CTE being the major factor;

- Higher stand-off height increases lifetime;

- Thicker substrate reduces lifetime;

- Trapped underfill can increase lifetime;

However, when it comes to choosing the 'best' material properties, conflict of interest exists: what is good for the solder joint life may be detrimental to the die.

In the experiments, fewer voids have been found in the product but underfill trapping is a common phenomenon. The modelling results on this problem have shown that the effect of the underfill trapping is a complicated one. In some cases, trapped underfill may act as a crack stopper.

The modelling strategy used to analyze the problems discussed in this chapter was the parametric study approach. A major disadvantage of the parametric study approach is that it does not provide detailed information on the sensitivity of design variables on lifetime and their interactions. It can not detect efficiently possible important interactions between assembly parameters and normally requires significant number of analysis to aid the optimal design solution.

It is essential to adopt a more efficient and preferably fully automated approach that is capable to account for parameter interactions in a design problem and can provide the optimal solution quickly and reliably. The following sections will present the proposed optimisation modelling approach for different microelectronics applications and will demonstrate the design process using the coupled *VisualDOC-PHYSICA*. The route of applications being presented follows the degree of analysis complexity – from single-physics to increasing multi-physics coupling.

CHAPTER 6

OPTIMISATION MODELLING FOR FLIP-CHIP SOLDER JOINT RELIABILITY BASED ON SINGLE PHYSICS PHENOMENON

This chapter details and demonstrates the integrated optimisation-modelling approach for predicting and optimizing the reliability of solder joints associated with electronic packages [177-182]. The integrated *VisualDOC-PHYSICA* design tool is used to identify efficiently and quickly the most suitable design parameters for flip-chip solder joint performance during thermal cycling. Number of electronic packaging conference papers [178-180, 182], a *Springer-Verlag* conference book publication [177] and a *Soldering & Surface Mount Technology* journal publication [181] focus on this integrated design methodology. The publications detail the coupled software tools and discuss the reliability optimisation of flip-chip packages. These investigations and the relevant optimisation studies are on different design parameters and assembly technologies and also observe the performance of different optimisation strategies and their efficiency.

The flip-chip components assembled using the *no-flow* underfills as discussed in Chapter 5 are now designed and optimised using the automated design software system *VisualDOC-PHYSICA*. The design process focuses on the optimal material selection and determination of those geometric configurations of the flip-chip package that will enhance its thermal performance and will provide longer life-time of the electronic products.

6.1. Problem Specifications

The *no-flow* underfill alternative applies underfill material before the reflow process and uses the reflow process to cure the underfill and fuse the solder joints at the same time as shown in Figure 6.1. Despite the clear advantages over the traditional assembly process, no-flow technology brings a number of concerns for the thermo-mechanical reliability of these packages. This is primarily due to the change in the material properties of current *no-flow* underfill materials, and the experimental results indicate that the thermal fatigue lifetime decreases.
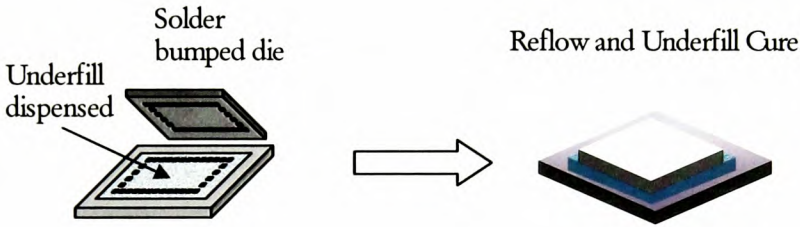
Figure 6.1: Novel *No-Flow* Underfill Assembly Process

Understanding and defining appropriate materials and geometric parameters for flip-chip packages assembled using this novel no-flow technology is crucial for subsequent product reliability.

6.2. From Parametric Study towards Optimisation Modelling

Finite Element Modelling techniques have been used extensively to predict thermal stress in solder joints. Coupled with suitable fatigue models these tools provide the ability to gain an insight into the effect that changes in material/geometric properties have on product reliability [47-49].

Although the parametric study on *no-flow* underfilled flip-chip package, as presented in Chapter 5, provides critical data and insight on the effects of changes to design parameters, this design approach can be very time consuming. Also, the parametric study strategy cannot investigate more than two parameters simultaneously and the important aspect of parameter interaction and their combined influence cannot be observed. Obviously, to identify optimal

characteristics of a system or process, all design variables have to change simultaneously in optimisation process.

Tools, based on numerical optimisation techniques, overcome the undesirable features of the parametric approach by offering a very logical and automated approach in identifying the best model and process parameters for different design problems [69, 73, 74, 181]. Applying such optimisation strategies to microelectronics component design can offer a fully automated, time efficient and systematic approach to achieve the most appropriate design characteristics.

The time efficiency of this approach compared to the parametric study approach is achieved as a result of the reduction in the number of simulations used to obtain the best design variable values and the design response sensitivities. For example, the parametric study on underfill Young's Modulus and CTE in Chapter 5 (see Table 5.5) uses 30 finite element simulations (total CPU of 165 hours on 1GHz Compaq Alpha ES45 processor). In this chapter it will be shown how the same problem but for doubled number of design variables (total of four design variables) is analyzed using the optimisation modelling approach with no more than 20 simulations. In this case, although the number of the investigated design variables is increased twice, the optimisation approach procedure takes 33% less simulation time (equivalent to 55 hours) than the parametric study.

6.3. Flip-Chip Reliability and FEM Computational Modelling

6.3.1. *Geometry and Mesh Model*

The flip-chip assembly under investigation has the same nominal geometric specification as outlined in Chapter 5. It consists of a silicon die attached to an FR-4 board and connected via solder joints. Underfill epoxy material fills the gap between the chip and substrate. Table 6.1 provides details on the dimensions of the nominal package and specifies ranges for variations for those parameters that will be investigated and optimised.

The connection of the chip to the substrate is accomplished using 48 uniformly and peripherally arranged solder joints. The distance between any two adjacent solder joints, pitch

size, is 457 µm, and the distance between the substrate and die (stand-off height) is 130 µm for the initial flip-chip design.

Table 6.1: Initial Geometric Parameters and their Variation Ranges

Geometry Parameter	Initial value [mm]	Variations	
		Absolute range [mm]	Range from Initial value
Die dimensions	6.3 x 6.3 x 0.6	none	none
Substrate thickness	1.5	1.2 1.8	± 20 %
Stand-of height	130 µm	104 µm - 156 µm	± 20 %

Pad
    Diameter
    Thickness
	
	150 µm
	35 µm
		
		none
		none
			
			none
			none

Bump pitch size	457 µm	none	none
Number of bumps	48	none	none

Figure 6.2 details the computer model used to represent this package. One-eight symmetry is assumed to lower the compute times. The most important requirement in building the finite element model is parameterization. Simulation-based virtual design optimisation is an automated process thus any characteristic under investigation (optimisation) has to be expressed as parameter in the corresponding input files. This aspect of the design tool was discussed in Chapter 4. The no-flow flip-chip parametric modelling of the geometry is adopted now in *VisualDOC-PHYSICA* software tool.
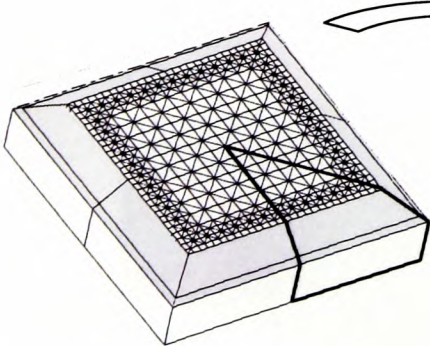
Figure 6.2: Top View of Flip-Chip Assembly and Modelled Part
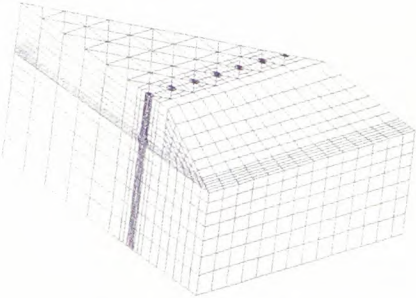
Figure 6.3: Flip-Chip FE Model - Geometry and Mesh

Once the package geometry and all model components are parametrically defined, they are automatically meshed. Generated mesh across the initial 3-D flip-chip assembly model has 9 212 elements. This mesh density ensures reasonable computational time for single FE analysis. In the same time this mesh is significantly improved compared with the mesh used in Chapter 5. Mesh elements consist of linear wedges and bricks (Figure 6.3). All solder interconnects inside the modelled one-eight part of the whole package are taken into account and included into finite element model representation.

6.3.2. *Material Model*

Figure 6.4 details the materials used in the flip-chip package. Under temperature cycling test or real operational environment the solder material experiences deformations due to time-dependent creep and stress relaxation [34, 45]. A reliable and extensively used model for solder joint thermal fatigue is the *sinh* constitutive law of creep deformation in the solder discussed previously in Chapter 5 (Equation 5.1). This visco-plastic model is assigned to solder material and used in simulations while the other materials are assumed to behave in an elastic manner.

All five materials used in the package have temperature dependent properties. The material data for temperature dependent material properties for this flip-chip assembly is taken from Le Gall, Qu and McDowell [51] and Hong [183]. As a result of this, the material data in the model is improved compared with the study presented in Chapter 5.
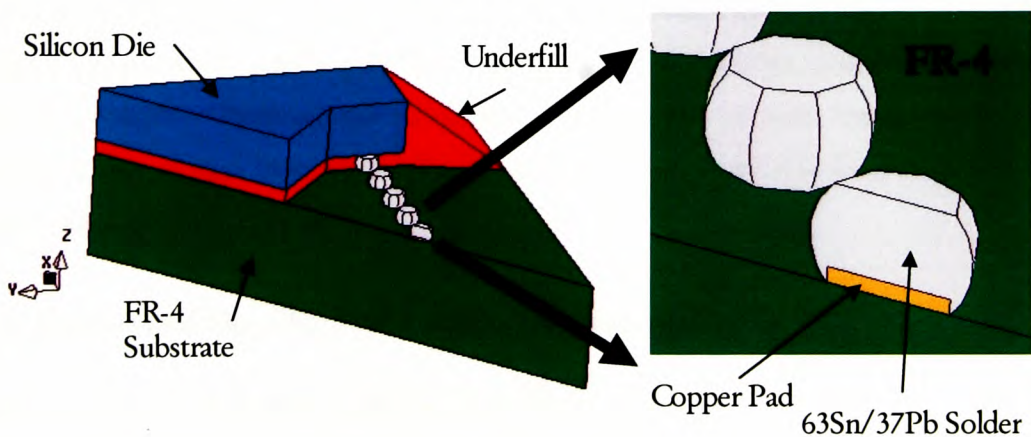
Figure 6.4: Flip-Chip Assembly Materials

With respect to the coefficient of thermal expansion CTE, the organic composite materials (FR-4) for package carrier substrate is modelled as an orthotropic material. The detailed properties of all materials are summarised and reported in Table 6.2.

Table 6.2: Package Material Properties and their Variation Ranges

Material	Temp. (°C)	CTE (10^{-6}/°C)	Young's Modulus E (GPa)	Poisson ratio v
Substrate : FR4 Card	30	20.0 (xy) , 86.5 (z)	22.4	0.1425
	95	20.0 (xy) , 86.5 (z)	20.68	0.1425
	125	20.0 (xy) , 86.5 (z)	19.3	0.1425
Pad : Copper	-25	16.0	128.0	0.34
	0	16.0	127.0	0.34
	50	16.0	124.88	0.34
	125	16.0	121.45	0.34
Solder : Sn63/Pb37	0	25.2	26.45	0.360
	50	26.1	12.52	0.365
	100	27.3	6.91	0.378
Die : Silica	0	2.8	162.1	0.23
	50	2.8	150.0	0.23
	100	2.8	140.0	0.23
Underfill	-----	80.0	3.0	0.35
Underfill Variations	-----	50.0 - 90.0	2.0 - 5.0	none

6.3.3. *Modelling Assumptions*

Time dependent non-linear stress/strain analyses are undertaken in the optimisation process to evaluate the thermo-mechanical flip-chip responses for different design configurations. A number of simplifications and assumptions in the analysis are imposed. The stress-free state for the entire assembly, including the die component, is at 125°C, which is the starting thermal cycle temperature. Modelling the temperature cycling test assumes isothermal loading throughout the package. In addition, all initial stresses are neglected. Finally, perfect adhesion between all materials is considered.

6.3.4. *Analysis and Flip-Chip Initial Design Model Response*

Material properties, boundary conditions and temperature profile are the required analysis inputs. The thermal load that has been used consists of four stages: temperature decrease, dwell at low temperature extreme, temperature increase, and dwell at high temperature

extreme. Each of these temperature phases is equal to 15 minutes and the temperature range is from -25 to 125 °C (Figure 6.5). The single-physics analysis of stress/strain behaviour of flip-chip assembly based on the prescribed temperature loading will provide also the important quantity of accumulated in the solder creep strain per cycle (respectively creep energy) as the damage parameter. As discussed in the previous chapter this type of analysis requires transient simulation that follows the temperature profile.
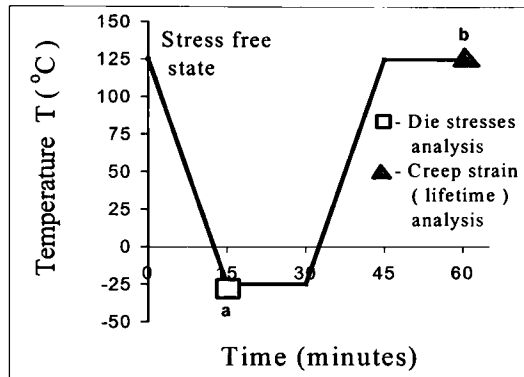
Figure 6.5: Thermal Cycle Specification

The potential for die failure is estimated by monitoring the maximum levels of Von Mises stress in the die at the end of cooling down stage in the thermal cycle (Figure 6.5, point a) which is associated with highest induced stresses in flip-chip package. It is known that the highest Von Mises stress in the die will be at its bottom surface, especially those areas at the center and corner of the die. The reason for this is that the low value of CTE for the die compared to the other materials used in the assembly results in high compressive stresses at the die bottom side and tensile stresses at the die top surface. Figure 6.6 shows the Von Mises stress contours in silicon die for the initial flip-chip package design configuration (geometry and material) as reported in Tables 6.1 and 6.2.

In this study the Coffin-Manson correlation methodology for lifetime predictions as described in Chapter 5 (Equations 5.2 and 5.3) is adopted to predict number of cycles to solder joint thermal fatigue failure. In the process of studying different flip-chip assemblies it was detected that the corner solder joint is always the one with the highest accumulated effective creep strain. Figure 6.7 illustrates the creep strain distribution across corner solder joint for the initial flip-chip design.
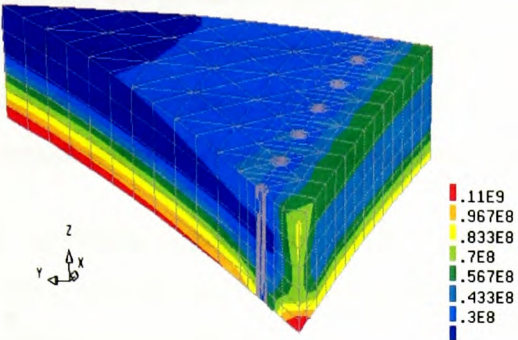
Figure 6.6: Silicon Die: Von Mises Stress [Pa] Distribution and Deformation (Initial Design)
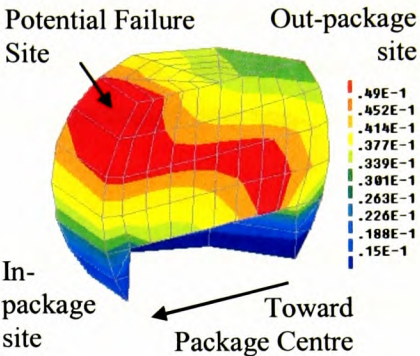
Figure 6.7: Accumulated Effective Creep Strain Levels in Corner Joint (Initial Design)

6.4. Flip-Chip Design Optimisation Strategies

The design tool based on coupled finite element code *PHYSICA* with the optimisation tool *VisualDOC* is explored to solve the above design problem. To perform the optimal design procedure, the optimizer uses FE evaluations (i.e. the flip-chip design responses) for different design configurations (i.e. different sets of values for specified design variables in the package) that the optimisation technique iteratively generates. Two optimal design strategies are tested and compared in this optimisation modelling. First, the optimal procedure adopts the direct first-order search techniques based on gradient calculations [143] as described in Chapter 4. The second approach demonstrates the Design of Experiments (DoE) combined with Response Surface (RS) approximations [143] strategy detailed in Chapter 3.

6.5. Flip-Chip Design Optimisation: Simulations and Results

6.5.1. Design Optimisation Task

Previous studies for flip-chip assemblies have identified the importance of underfill material properties and some geometric parameters such as substrate thickness and stand-off height for solder joint fatigue reliability [46-49, 177-180]. With respect to *no-flow* underfilled assemblies the underfill Young's modulus *(E)* and its coefficient of thermal expansion *(CTE)* have been experimentally investigated [168, 184] and it was observed that they had a major impact on reliability. In this optimisation modelling both material properties of the *no-flow*

underfill and geometric design parameters are investigated and optimally designed to provide improved thermo-mechanical reliability of the whole electronic package. As part of the design methodology, predictions for the effect of design parameters and their degree of impact (sensitivity) on reliability can be made.

Four design variables are used to parameterize the flip-chip geometry and material design. These variables are the design specifications optimised in such a manner to ensure enhanced package performance:

1. Young's Modulus of underfill *(E, GPa)*

2. Coefficient of Thermal Expansion of underfill *(CTE, ppm/°C)*

3. Stand-off Height *(SOH, μm)*

4. Substrate Thickness *(ST, mm)*

The objective in the optimisation study is to maximize solder joint reliability in terms of number of cycles to fatigue failure N_f.

Within the optimisation procedure the above four design parameters are allowed to vary within specified ranges. The variation limits for the design variables under investigation are given in Tables 6.1 and 6.2. Another important constraint used in the design specification limits the maximum Von Mises stress in the silicon die $\bar{\sigma}$ at the end of the cooling down period in the thermal load (high stress cycle point). By imposing upper allowable limit on die stress, the design optimisation results will ensure that stress induced failures associated with silicon die (e.g. die cracking) are avoided. The upper allowable limit value for Von Mises stress in the die σ^{max} is taken to be 155.5 MPa. This value is entirely acceptable because it is significantly less than the failure stress magnitude associated with silicon which has values in the range of 220-250 MPa and higher [174].

The constrained optimisation task can be expressed in the following form:

Find values for the design variables *(E, CTE, SOH, ST)* that will

Maximize Lifetime, $N_f = 0.38(\varepsilon_{sum}^{creep})^{-1.96}$

Subject to:

$$\bar{\sigma} \leq \sigma^{max} \quad , \text{where} \quad \sigma^{max} = 155.5 \; MPa \tag{6.1}$$

and lower and upper limits on E, CTE, SOH and ST
as given in Tables 6.1 and 6.2.

The lifetime model in above optimisation task adopts the Coffin-Manson relationship (see Chapter 5, Equation 5.3).

Solutions of the above design optimisation task for analysed flip-chip model are provided in the following sections using two different strategies:

1. Direct Optimisation using finite difference gradient calculations;

2. Design of Experiments (DOE) and Response Surface Models based optimisation.

6.5.2. Flip-Chip Design using Direct Optimisation

Direct Optimisation is based on gradient search techniques that are capable of finding the exact optimum design. Finite difference techniques are exploited to calculate gradients of the objective and constraint functions and this process requires calls to the analysis program. These gradients are used to specify search directions in the design space as described in Chapter 4. Other part of the calls to *PHYSICA* package evaluates the objective and the constraint in the optimisation search process once a search direction becomes available. This optimization process is slowed down due to the transient visco-plastic analysis of the flip-chip thermo-mechanical behaviour under thermal cycling. The benefit of applying this approach is in obtaining the exact solution of the optimization problem. This can be used subsequently to compare with other optimization strategies.

The solution of the optimisation task relies entirely on the finite element analysis predictions for package performance. In this design problem all sub-problems are executed automatically with no further user interaction with the integrated design tool. Once the design task is defined and the design problem executed, the final optimal design configuration is achieved in a fully automated manner. A new design evaluation involves two important sub-tasks:

- Updating model geometry input files and generating the new model. This sub-task will be executed if geometric design variable is used, e.g. solder joint stand-off height. Also, a subsequent meshing of the domain is required. To minimize the mesh change effects, mesh element ratio is kept close to a constant value. Also, volume weighted average values over mesh elements of interest for the analyzed response are normally observed. Within the integrated design tool, this sub-task is executed using the CAD/Mesh tool *FEMGV* [119];

- Updating *PHYSICA* analysis input files. These new settings representing the new design are used then in finite element simulation. The analysis input files are updated if process condition or material properties are used as design variables, e.g. underfill properties.

The above two sub-tasks and the rest of the interaction procedures between *VisualDOC* and *PHYSICA* are fully automated and do not require any interaction from the user as discussed in Chapter 4.

The gradient-based numerical optimisation technique used to solve the constrained optimisation problem (6.1) is the Modified Method of Feasible Directions (MMFD) [26, also Chapter 4].

Figure 6.8 illustrates the design optimisation history. This contained five search directions, with each direction containing a number of calls to the analysis code *PHYSICA*. Clearly we can see that the optimiser has changed the design variables in such a way as to ensure the cycles to failure increase from 136 to 589. To obtain the optimal values of design variables the optimiser has made 15 analysis calls to the finite element programme *PHYSICA*.

The changes in each design variable throughout this process are given in Figure 6.9. These show a decrease in *CTE* and substrate thickness and an increase in Young's modulus and stand-off height from the values at the initial design.
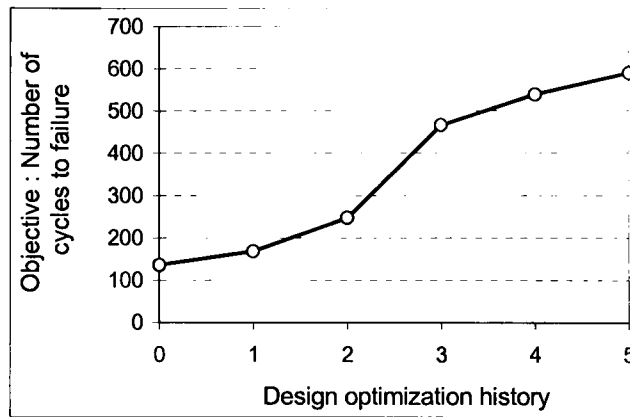
Figure 6.8: Objective Optimisation History
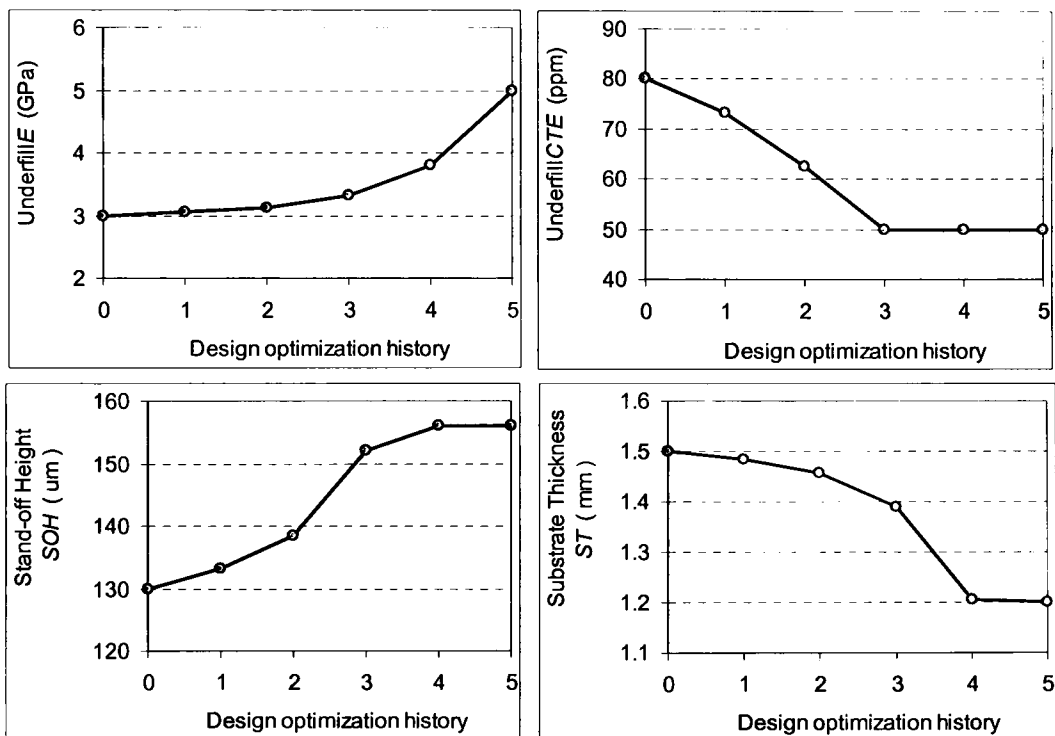
Figure 6.9: Design Variables Optimisation History

The increased lifetime for the optimal design configuration is a result of significantly reduced levels of accumulated effective creep strain in the solder joins at the end of a thermal cycle. Figure 6.10 illustrates the history of creep strain accumulation across the most critical corner solder joint for the initial and optimal designs throughout a full thermal cycle.
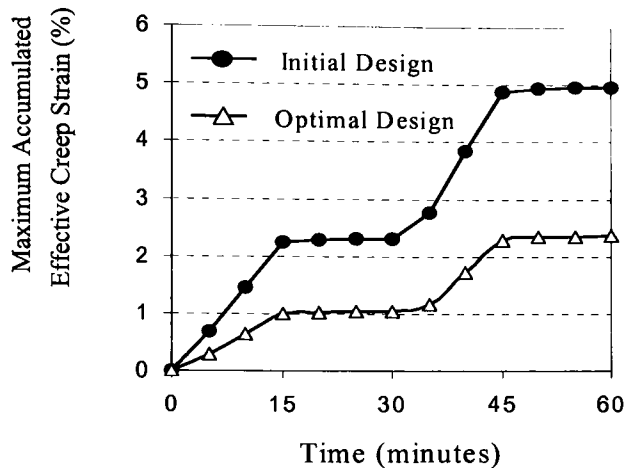
Figure 6.10: Maximum Accumulated Effective Creep Strain as
Function of Time during a Complete Thermal Cycle

The above optimisation results have very good agreement with the trends identified in the parametric study outlined in Chapter 5. No-flow underfills with low CTE and high Young's Modulus improve the reliability of package. For the geometric design variables, higher stand-off height and thinner substrate will also enhance the reliability of solder joints. The value of constraint, the maximum stress in silicon die, increases slightly at the optimal flip-chip design from 120MPa to 134MPa.

Further analysis of the optimal results will follow at the end of this chapter.

6.5.3. *Flip-Chip Design using Design of Experiments (DoE) and Response Surface (RS) Modelling*

DoE and RS modelling are exploited within the sensitivity and optimisation analysis by:

1. Identifying the experimental design points.

2. Undertaking a finite element analysis at each design point to obtain the system responses (experimental data) at these points.

3. Using the response data to construct response surface (RS) approximations.

4. Perform sensitivity and optimisation analysis using the generated response surfaces.

This section will demonstrate the above procedure as an alternative approach to direct optimisation (Section 6.5.2) for solving the same optimisation task (6.1). Additional sensitivity analysis will also be provided.

6.5.3.1. Design of Experiments (DoE) Method

The starting point in a DoE analysis is to establish the set of experimental design points at which the finite element analysis will be undertaken to provide flip-chip response data (i.e. stress and strain values). The choice of polynomial order for intended Response Surface approximation is full quadratic with mixed linear terms. In our case we have four design variables and such a polynomial approximation will have 15 coefficients. To build the RS model, i.e. to find the values of these coefficients, at least 15 (exact fit of experimental data) or more (for least squares fit of experimental data) experimental points are required.

In this study the Three-Level Taguchi Orthogonal Array L9 and the Factorial Notz design techniques are used [143, Chapter 3] and these result in 19 design points. These two experimental design techniques are suitable to construct accurate full quadratic approximations to the system response data using a relatively low number of design points provided by both DoE methods. Their combination gives a better construction of the response surface model. The central location of the design space is also added to the design points in the final experimental design. This extra point provides greater accuracy in the second order polynomial approximations to solder creep strain and die stress response of flip-chip assembly. All experimental points are determined based on design variables limits as specified in Tables 6.1 and 6.2:

1. Underfill Young's Modulus (E) from 2 to 5 GPa;

2. Underfill CTE from 50 to 90 ppm/°C;

3. Solder joint stand-off Height (SOH) from 104 to 156 µm;

4. Substrate Thickness (ST) from 1.2 to 1.8 mm.

Once the design points have been identified, the next step is to calculate the flip-chip design responses at these points. Executing DoE task in *VisualDOC* runs the finite element analysis (*PHYSICA*) for experimental points. Post-processing calculations provide the Von Mises

stress in the die at the end of cooling down phase (constraint value) and the accumulated effective creep strain in the solder at the end of the thermal cycle (used to identify the objective, number of cycles to failure N_f, where $N_f = 0.38(\varepsilon_{sum}^{creep})^{-1.96}$). Table 6.3 shows the set of experimental points, the predicted responses (stress in the die and accumulated creep strain in solder joints) and cycles to failure at each design point.

Table 6.3: Design of Experiments

Design Point No.	E (GPa)	CTE (ppm)	SOH (μm)	ST (mm)	Die Von Mises Stress in Silicon Die $\bar{\sigma}$ (MPa)	Accumulated Effective Creep Strain in Solder $\varepsilon_{sum}^{creep}$ (%)	Cycles to Failure : Lifetime N_f
1	5	90	130	1.2	147.89155	5.0527405	132
2	5	70	104	1.8	127.05443	4.1510113	194
3	5	50	156	1.5	125.33030	2.4542704	544
4	3.5	90	104	1.5	128.68752	5.7681266	102
5	3.5	70	156	1.2	129.30603	3.8058821	230
6	3.5	50	130	1.8	113.37784	2.9955601	368
7	2	90	156	1.8	97.64297	5.8739781	98
8	2	70	130	1.5	108.06065	4.7173206	151
9	2	50	104	1.2	120.97282	3.6309626	252
10	2	50	156	1.8	96.85377	3.3120830	302
11	2	90	156	1.2	113.63030	5.5149138	111
12	2	90	104	1.8	106.59825	6.7686215	74
13	5	90	156	1.8	130.82172	5.1244829	128
14	5	90	104	1.2	148.09805	5.2085061	124
15	5	50	156	1.2	133.41422	2.3565151	589
16	5	50	104	1.8	122.07200	2.8720692	400
17	3.5	50	130	1.5	119.59298	2.8529394	405
18	3.5	70	104	1.5	125.39785	4.4147812	172
19	3.5	70	130	1.2	131.55486	4.0000290	209
20	3.5	70	130	1.5	122.82088	4.1535586	194

6.5.3.2. Response Surface (RS) Modelling

After obtaining flip-chip responses at the experimental design points as detailed above, the next stage is to construct explicit analytical functions that can serve as an approximation to the objective and constraint function respectively. This study adopts second order polynomials as response surfaces (RS) for Von Mises stress in silicon die of flip-chip and

accumulated effective creep strain in solder joint. The coefficients in both RS approximations are determined by fitting the response data in Table 6.3 using the least square fitting approach. If these RS approximations represent the flip-chip responses accurately, then they can be used to evaluate approximately the objective function (N_f) and the constraint ($\overline{\sigma}$) values for any flip-chip design. In particular, the RS models can be used within the optimisation procedure for design point evaluation.

The usual optimisation procedure can now be used to find the optimal flip-chip design variables that maximise solder joint lifetime. The major advantage of coupled DoE and Response Surface Methodology here is that we can use these approximations for investigating response variables (i.e. solder creep strain and die stress) instead of expensive repeated calls to the FEA package. Also, as it will be emphasised in the following, the RS models provide a more robust way in locating the global optimal solution of an optimisation task and offer options of performing sensitivity analysis.

The follow quadratic polynomial approximation for maximum accumulated effective creep strain in solder joint, $\varepsilon_{sum}^{creep}$, as function of the design variables can be used:

$$\varepsilon_{sum}^{creep} \ (E(GPa), \ CTE(ppm), \ SOH(\mu m), \ ST(mm)) =$$

$$3.2673 -1.012628*E +0.0634970*CTE -0.027310*SOH +0.923348*ST$$
$$+0.0000827*E*CTE +0.00322507*E*SOH -0.1078052*E*ST \qquad (6.2)$$
$$-0.00002987*CTE*SOH -0.000204881*CTE*ST -0.004198105*SOH*ST$$
$$+0.063671*E^2 +0.000041*CTE^2 +0.000055*SOH^2 +0.185029*ST^2$$

In the above approximation, the accumulated effective creep strain $\varepsilon_{sum}^{creep}$ is in % and uses un-scaled values of the design variables.

Similarly, an explicit full quadratic polynomial that approximates the maximum Von Mises stress in the die, $\overline{\sigma}$, is obtained and used in the optimisation procedure instead of calling the FEA package:

$\overline{\sigma}$ $(E(GPa), CTE(ppm), SOH(\mu m), ST(mm)) =$

    $191.236 + 3.450*E - 0.17946*CTE - 0.34391*SOH - 53.13357*ST + 0.094647*E*CTE$

    $+0.04797*E*SOH - 0.27162*E*ST + 0.00079*CTE*SOH - 0.075936*CTE*ST$ (6.3)

    $-0.029470*SOH*ST - 1.12095*E^2 + 0.00016*CTE^2 + 0.000223*SOH^2 + 12.6837*ST^2$

In the above RS model (6.3) stress is in MPa and the values of design variables in this polynomial approximation are also un-scaled.

The statistical tools available through the *VisualDOC* software framework are used to judge the quality of the above generated RS approximations. These tools are based on the Residual Analysis, Analysis of Variance and statistical efficiency measures outlined in Chapter 3. The techniques have shown that the coefficients predicted for each RS polynomial provide a good fit and are accurate for both die stress and creep strain responses.
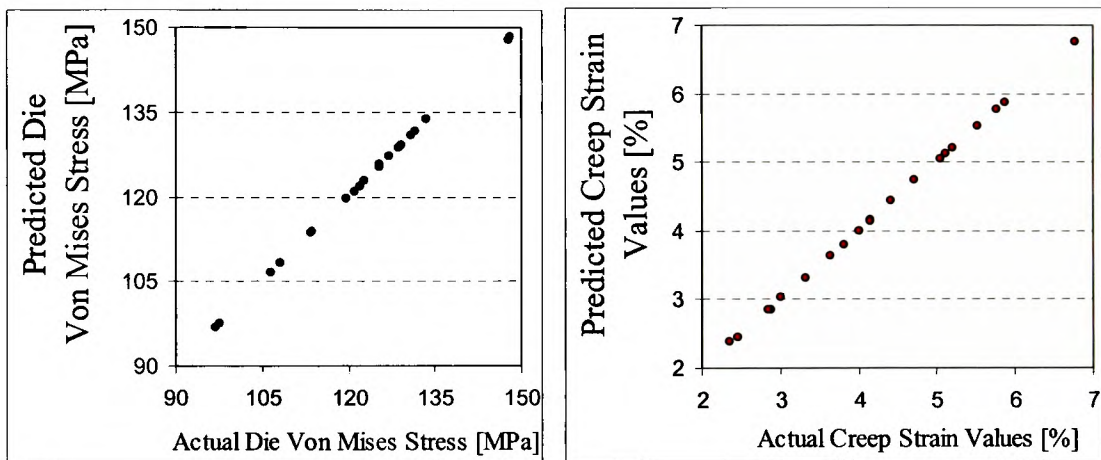
Figure 6.11: Actual vs. Predicted Responses

In Figure 6.11 the values of **actual** design responses at experimental points that are estimated using finite element analysis are compared with the **predicted** values from the RS polynomials that approximate these responses. Clearly we can see that the polynomials provide very good accuracy to the response values at the experimental design points. The coefficient of variation (see Chapter 3, Equation 3.34) which defines how large is the error compared with the response value and the adjusted coefficient of multiple determination

R_{adj}^2 (see Chapter 3, Equation 3.32) which is a measure of RS degree of fitting the experimental data are:

1. For calculated Von Mises stress in silicon die, the coefficient of variation is 0.25% and the $R_{adj}^2 = 99.949\%$;

2. For maximum accumulated creep strain in solder joint, the coefficient of variation is 0.63% and the $R_{adj}^2 = 99.953\%$.

6.5.3.3. Sensitivity Analysis of Flip-Chip Design

Apart from the benefits of avoiding actual FE analysis runs in the optimisation process, Response Surface Modelling has the additional advantage to perform sensitivity analysis (something which is not explicitly available in the direct optimisation approach). Sensitivity analysis helps identify interactions between the design variables and their importance and level of influence on design responses. These sensitivities are obtained from the coefficients in the RS approximation.

To avoid the influence of different design variable units and the different lengths of design boundaries (i.e. Young's modulus E from 3 to 5 GPa and CTE from 50 to 90 ppm/°C), the sensitivity data is based on the RS approximation where design space is normalised. In the normalised design space the lower design variable bounds correspond to -1, and the upper bounds to +1. Normalization procedure for design variables is described in Chapter 3, Section 3.2.1. As a result of design space scaling, one unit of change in normalised design space for E, CTE, SOH and ST corresponds respectively to changes of 1.5 GPa, 20 ppm/°C, 26 µm and 0.3 mm in their values.

Note that the RS model for sensitivity purposes has different polynomial coefficients than those in the Equation 6.2 which corresponds to the un-scaled problem definition. The following quadratic polynomial approximation for accumulated effective creep strain in solder joint, $\varepsilon_{sum}^{creep}$, as function of the scaled over [-1, 1] design variables E_{sc}, CTE_{sc}, SOH_{sc} and ST_{sc} can be used now to perform sensitivity analysis:

$$\varepsilon_{sum}^{creep}\ (E_{sc},\ CTE_{sc},\ SOH_{sc},\ ST_{sc}) =$$

$$4.13360 - 0.45539*E_{sc} + 1.30600*CTE_{sc} - 0.26454*SOH_{sc} + 0.16231*ST_{sc}$$
$$+0.00248*E_{sc}*CTE_{sc} + 0.12578*E_{sc}*SOH_{sc} - 0.04851*E_{sc}*ST_{sc} \tag{6.4}$$
$$-0.01553*CTE_{sc}*SOH_{sc} - 0.00123*CTE_{sc}*ST_{sc} - 0.03275*SOH_{sc}*ST_{sc} + 0.14326*E_{sc}^{2}$$
$$+0.01630*CTE_{sc}^{2} + 0.03701*SOH_{sc}^{2} + 0.01665*ST_{sc}^{2}$$

In the above approximation, E, CTE, SOH and ST stand for underfill Young's Modulus, underfill co-efficient of thermal expansion, solder joint stand-off height and substrate thickness respectively as these design variables were previously defined. The subscript "sc" denotes that now these variables are *scaled* over [-1, 1] interval.

Figure 6.12 illustrates the sensitivity of solder creep strain response to terms in the RS approximation (6.4). These are represented by the coefficients in the RS approximation, generated in the described above normalised design space. For example, a positive unit change in the Young's modulus (i.e. 1.5 GPa) of the underfill will result in a decrease of 0.455 % in the creep strain through the RS approximation linear term (E) for the underfill Young's modulus. But at the same time the quadratic term in RS model for underfill E (denoted $E**2$) will increase the creep strain with 0.143 %.
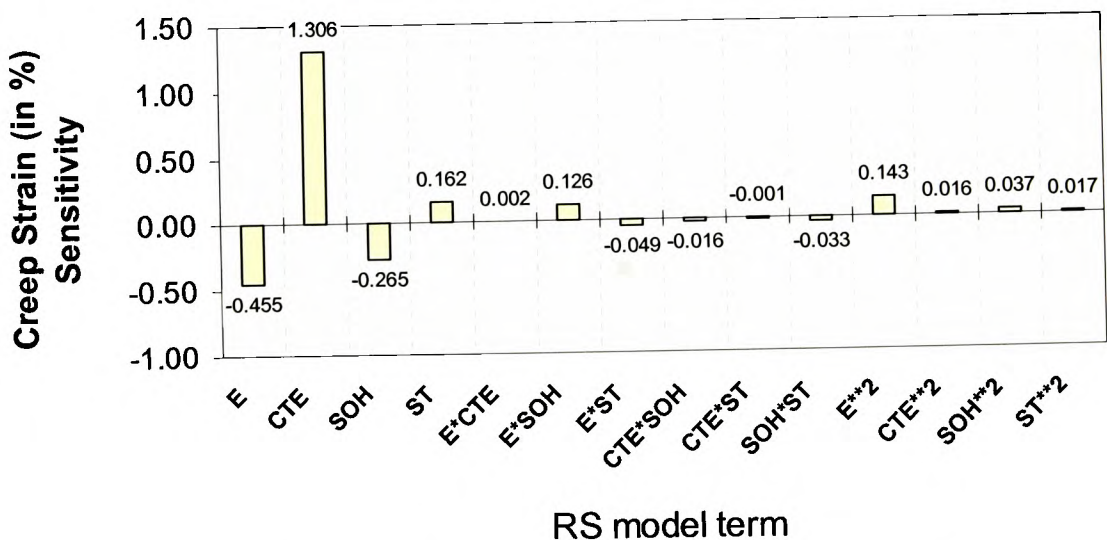
Figure 6.12: Sensitivity of Accumulated Effective Creep Strain in Solder Joints to Design Variables and Their Interactions

This sensitivity analysis also shows that the underfill *CTE* has significant impact on creep strain in solder joint and is the most influential design variable in this flip-chip package. This is followed by the underfill Young's modulus E, then solder stand-off height *SOH*, then substrate thickness *ST* and then the other terms in the response surface approximation.

The sensitivity analysis also identifies that creep is influenced mainly through the linear terms in the model. The most influential quadratic term in this polynomial approximation is the Young's Modulus quadratic term ($E**2$).

Design variable interaction for this design problem is also less influential, with most significant contribution to creep strain from *Underfill Young's Modulus   Stand-of-Height* interaction ($E*SOH$). Positive unit increase in value of both variables will contribute to an increment in solder creep strain of 0.126%.

We have to emphasise that for a polynomial RS model the most important term is the constant term. For the above RS approximation using scaled design variables this constant term is 4.1336. All sensitivities illustrated in Figure 6.12 relate to changes in creep strain values from the value of this constant term.

6.5.3.4. RS Models used to Perform Approximate Optimisation

DoE and RS based optimisation uses the constructed RS approximations to evaluate explicitly the constraint and objective function values without running finite element analysis to provide certain design evaluation. Direct gradient-based optimisation techniques such as Modified Method of Feasible Directions (MMFD) or Sequential Linear/Quadratic Programming (see Chapter 4) can then be used to obtain the optimal solution of design task (6.1). MMFD carries out the optimisation calculation for this DoE +RS based design process.

Optimisation results provide optimal value for underfill Young's Modulus and CTE of 5GPa and 50ppm/°C respectively. The best stand-of-height is 156 microns and the best substrate thickness for the flip-chip assembly is 1.2 mm. At this optimal flip-chip design, the approximated values from RS models for the Von Mises stress in die and accumulated solder creep strain are 133.70MPa and 2.38% respectively. The approximate prediction for the optimised lifetime of this package is 579 cycles to failure.

Figure 6.13 illustrates the lifetime in terms of cycles to failure based on the solder creep strain RS model (Equation 6.2). The first graph relates lifetime to both underfill E and CTE and is generated by keeping the other two design variables, SOH and ST, at their optimal values. Clearly we can see that taking underfill CTE towards its lower bound and Young's modulus (E) towards its upper bound results in an increase in lifetime.

In the second graph (Figure 6.13) underfill CTE and E are kept at their optimal values and SOH and ST are allowed to vary. In this case we can see that decreasing substrate thickness (ST) and increasing stand-off-height (SOH) results in an increase in lifetime.
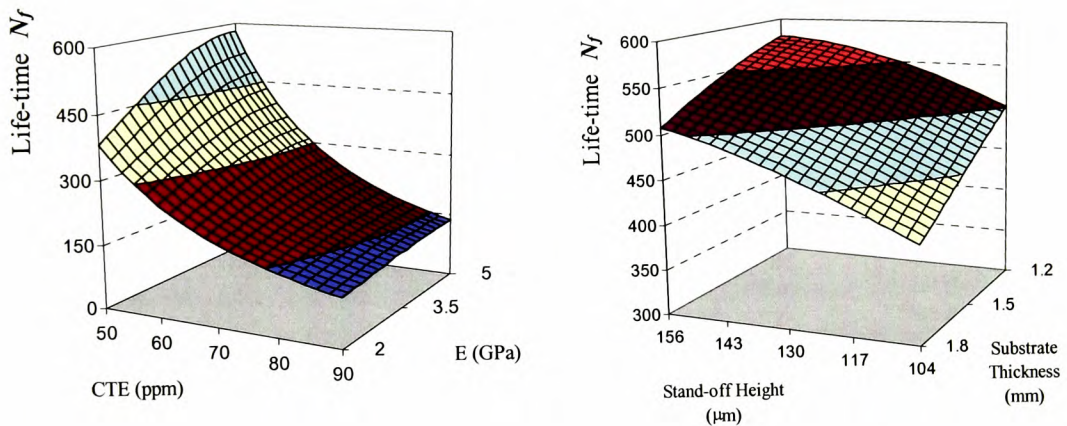
Figure 6.13: Solder Fatigue Lifetime as Function of Design Variables

Figure 6.14 shows similar graphs for Von-Mises stress in the silicon die. In this case, the generated RS approximation (Equation (6.3)) for this response is used. It is evident from the left graph that lower underfill Young's Modulus will reduce stresses in the silicon die. Geometric parameters have also influence – higher substrate thickness and solder joint stand-of height will reduce the stress in die.
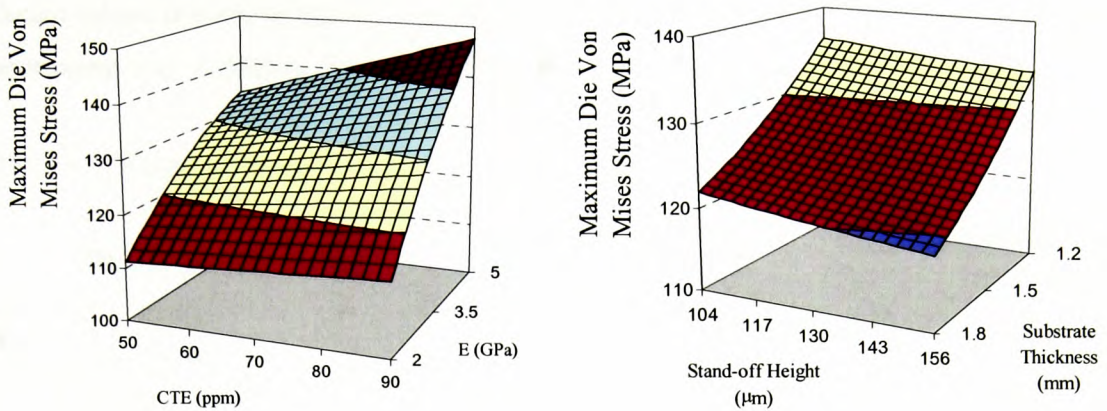
Figure 6.14: Maximum Die Von Mises Stress as Function of Design Variables

6.6. Comparison between Direct Optimisation & DoE/RS based Optimisation

The summarised data of optimisation studies described in this chapter is given in Table 6.4. Maximum Von Misses stress in the die at the end of cooling down phase is noted as *constraint* in this table and the number of cycles to failure (lifetime) as *objective*. For the observed flip-chip design variables and design space limits, both tested optimisation strategies (Direct and DoE/RS) locate the same optimal design:

- Best material properties: underfill Young's modulus E at its upper imposed bound (5GPa) and underfill *CTE* at its lower bound (50 ppm);

- Best geometric parameters: stand-off height *SOH* at its upper bound (156 µm) and substrate thickness *ST* at its lower bound (1.2 mm).

Finite Element Analysis based evaluations for both the initial and optimal designs show that a significant improvement in solder joint reliability is obtained. The calculated optimal design suggests thermal fatigue reliability in terms of cycles to failure is 3.33 times higher than that estimated for the initial design. This optimal design satisfies all imposed constraints. An acceptable increment of 11.53 % in maximum Von Mises stress in the die is observed when the optimal and initial designs are compared. In the case of contradicting design quantities (e.g. this study results: increasing solder fatigue life causes increasing of die stress) simulation-based optimisation approach identifies efficiently and automatically the optimal

design values. It is the design engineer's responsibility to represent in the optimisation task all restrictions and requirements that he or she feels are important.

Table 6.4: Summarised Optimisation Results / Optimisation Task (6.1) /

Design parameters		Lower bound	Upper bound	INITIAL design	OPTIMAL design	
					Direct	DOE / RS
1.	E (GPa)	2.0	5.0	3.0	5.0	5.0
2.	CTE (ppm)	50.0	90.0	80.0	50.0	50.0
3.	Stand-off height (μm)	104.0	156.0	130.0	156.0	156.0
4.	Substrate thickness (mm)	1.2	1.8	1.5	1.2	1.2
FEM analysis EXACT evaluation at design : Constraint Objective Accumulated effective creep solder strain (%)				119.617 136 4.96993	133.414 589 2.35652	133.414 589 2.35652
Approach (Direct / DOE RS) PREDICTION at the optimum Constraint Objective Accumulated effective creep solder strain (%)					133.414 589 2.35652	133.709 579 2.37745
Difference between EXACT and PREDICTED optimum: Constraint Objective Accumulated effective creep solder strain (%)					0 % 0 % 0 %	0.22 % 1.70 % 0.89 %
Number of FEM analyses for obtaining the solution					15	20
Total CPU time to perform one FE analysis (1GHz Compaq Alpha ES45 processor)					5 hours 28 min. 30 sec.	

A very good agreement in the results given by Direct Optimisation approach and DoE based approximations technique is reached for this particular design problem. The variation between exact and approximate optimal values for the objective function is only 1.7 %. Since the RS approximations for both responses $\varepsilon_{sum}^{creep}$ and $\overline{\sigma}$ have been generated accurately, at the optimal design the differences in their exact and approximate values are only 0.89% and 0.22 % respectively.

It should be noted that for this particular design problem, Direct Optimisation was capable to find the optimal solution with relatively low number analysis calls (15 *PHYSICA* simulations compared with 20 calls for DoE). Although in this case the DoE analysis was specified with number of experimental points higher than the number of analysis calls occurred in direct optimisation, generally this will not be the case for most design problems.

If optimum design point lies inside the design space (which is not the case in this problem) rather than to be located at the design space boundaries a gradient search technique will require certain amount of analysis calls to ensure satisfaction of convergence criteria.

Usually a significant number of simulations will be required to perform the search process using gradient-based optimisation and therefore the preferred option is the DoE method with predictions for design response values from RS approximations. In addition, the DoE approach offers the opportunity to perform sensitivity analysis and related evaluation of importance and influence of the design variables.

It is known that the search in direct gradient-based optimisation is dependent on the initial design point and will always locate the nearest to that point optimum. This optimal point could be only a local optimum and for a general optimisation problem with multiple optima the only way to verify this is to run the optimisation search starting with different initial points. This is impractical strategy for design optimisation with time consuming FEA evaluations of design points.

Optimisation based on RS models is more robust and cannot fail due to poor estimation of gradients. It also has the advantage to be more capable of locating the global optimal solution of design problem. The reason is that a RS model is based on response data at experimental points distributed over entire design space and such an approximation represents the response everywhere in the design space. A RS model, for example a second order polynomial, is likely to capture accurately the location of the true global optimum and to ignore the potentially existing local optima. Depending on the RS model (polynomial, multi-quadratic function, etc.) a gradient-based optimisation that utilises this approximation still may require a number of runs from different initial points to verify the global optimum. This will not be a problem, because there will be no calls to FEA, and the optimisation process will be time efficient. As it will be shown in the next chapter, with RS models a design engineer can use also non-gradient optimisation techniques that are very efficient in locating the global optimum.

6.7. Flip-Chip Design Optimisation: Conclusions

The thermo-mechanical analysis to predict solder joint reliability has been carried out as integrated part in an optimisation modelling of no-flow underfilled flip-chip assembly. Design process that identifies the optimal material selection and package geometry has been demonstrated. This integrated methodology can be used to identify optimal conditions for solder joint reliability.

Different optimisation strategies combined with response modelling and approximation techniques have been tested as part of the integrated design framework. The solder joint lifetime has been improved dramatically in a very time efficient, intelligent and automated manner. The total CPU time required to perform the optimisation/sensitivity analysis using the Direct and the DoE/RS based optimisation strategies was 82 hours 7 min. 30 sec. and 109 hours 30 min. 00 sec. respectively.

Optimisation modelling indicates that for analysed flip-chip package higher underfill Young's modulus E and stand-off height combined with lower underfill CTE and substrate thickness have a positive effect on solder joint thermo-mechanical reliability. In this analysis the variations of both geometric and material parameters are taken as continuous. In the manufacturing process, a suitable underfill can be chosen with the properties that are closest to the optimal estimated (taking into account all aspect from the sensitivity analysis).

The results from this analysis have investigated both underfill material properties and geometric parameters. In this chapter we have demonstrated that the influence on solder joint reliability is dominated by underfill material properties and the impact of flip-chip geometric design is less important. As a result of this, there is a base for comparison of presented optimisation results and the experimental studies performed by Wong [169, 171, 184] that investigate the influence of underfill material properties on solder joint reliability for no-flow flip-chip packages. The trends in solder joint reliability from this optimisation modelling work and the detailed above experimental analyses agree very well. The results are comparable in terms of underfill property influence and identified trends in cycles to failure.

Some of the design trends towards optimal reliability identified in this study were also observed for flip-chip packages using the traditional capillary flow underfill process. A

number of other optimisation modelling studies, which are not detailed here, have focussed and provided results on

- Underfill material optimisation [179];

- Underfill properties, die thickness, PCB thickness and stand-off height optimal design [178, 180];

- A study on the performance of optimisation strategies [177] available in *VisualDOC-PHYSICA* design tool using flip-chip assembly and the traditional underfill process. This analysis has provided further details on the methodology for performing effective design process. The results were used to state the advantages and disadvantages of each of these optimisation strategies;

- Optimisation task with six design variables, including parameters such as dimensions of pads on integrated circuit board [185].

The above design optimisation procedure has illustrated two approaches:

1. Direct gradient-based optimisation with calls to FEA package;

2. DoE and RS modelling plus RS-based optimisation and sensitivity analysis.

Both approaches need results from FEA thermo-mechanical simulations. In terms of design trends and predictions both of these methods give the same results as provided in Chapter 5.

The benefit of using optimisation is that it provides an automated approach to identify trends and optimal designs when compared with the parametric study approach.

The DoE and RS method has some advantages over the Direct Optimisation approach. In particular, for most problems it will require fewer FEA simulations and hence it will be faster. Also, it provides the ability to undertake sensitivity analysis on each of the design variables. RS models can be used over and over again to evaluate different design options without carrying out any further FE analysis.

CHAPTER 7

ECONOMICAL NITROGEN SUPPLY FOR LEAD-FREE WAVE SOLDERING PROCESS USING OPTIMISATION MODELLING

7.1. Wave Soldering Process

7.1.1. *General Overview on the Wave Soldering Concept*

Wave soldering is a manufacturing process used in the electronics industry to join components to printed circuit boards. It was first introduced at the time when through-holes interconnect technology was the commonly used method to assemble electronic products. In a wave soldering process the printed circuit board is carried by a conveyer over one or a multiple of solder waves. The solder wave is a continuously flowing molten solder fountain formed by a pump inside a solder pot. When the bottom side of the travelling printed circuit board is in contact with the solder wave the metallized parts of the board provide a wettable surface for solder deposition. Once the board has left the soldering bath and due to the cooling conditions the solder starts to solidify and forms the solder joints.

Inside the wave soldering machine there are three major steps to the soldering process. These stages are illustrated in Figure 7.1 and include:

- Preparation of the assembly surface for soldering (flux application);

- Pre-heating;

- Application of the molten solder to the bottom side of the assembly.
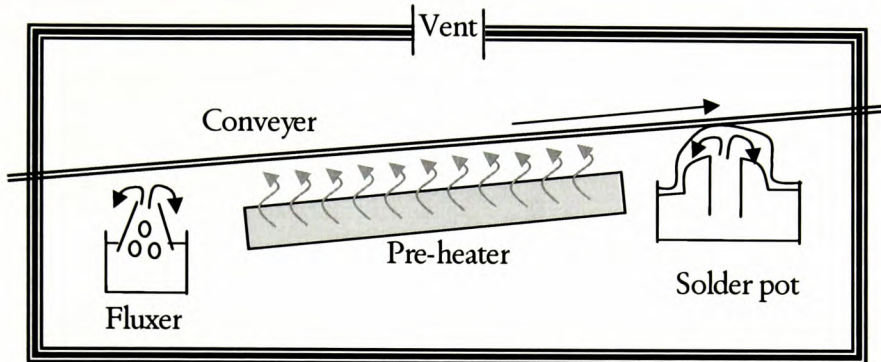
Figure 7.1: Schematic View of a General Wave Soldering Machine

Two major requirements to wave soldering process exist. First, all metallic parts at the bottom side of the board where solder joints will be formed have to be wetted by the molten solder wave. In the same time it is essential to avoid any formation of bridges or poor joints due to the excessive solder amounts captured from the wave. These opposing requirements can be satisfied if solder wave is adjusted in a way that gives turbulent flow at the entrance side of the wave and laminar at the exit. In the manufacturing process this strategy can be implemented by using multiple waves (typically double wave soldering). In this case each wave performs differently. In double solder waves the first wave is more turbulent, and the second solder flow is characterized with laminar flow to allow all unnecessary solder to drain down leaving proper solder amounts to form the joints in the subsequent cooling stage of the process.

7.1.2. *Flux Application*

The first stage of wave soldering process includes cleaning of all metallic surfaces designed to be soldered. The cleaning uses fluxes based on acidic compounds that treat chemically the surfaces. As result any oxidation of the board metallic parts and collected with the time dirt are removed. This action is performed just before the wave soldering operation. Three different types of fluxes – foam, wave and spray – are most commonly used [137, 186, 187].

7.1.3. *Preheating*

The wave soldering process requires the freshly fluxed board to be heated before it passes over the solder wave. This preheating temperature range may vary based on the assembled process but usually is of order 80-100 °C [186]. Many factors determine the requirements for preheating with emphasis on the physical rather than the chemical reasons. Heating the board causes flux solvent to evaporate and flux activation. It also minimises the thermal shock between the board and the solder wave by bringing the board temperature closer to the one of solder.

Avoiding a rapid absorption of heat in a very short period of time (thermal shock) reduces potential dangers caused by the components' different expansion coefficients and subsequent component failures at board level.

7.1.4. *Wave Soldering*

During the wave soldering phase of the process the board assembly passes over the solder wave and captures molten solder from the wave. A solder wave is typically produced by forcing molten solder upwards through a vertical conduit. It ends with a wave nozzle design to give particular shape characteristics of the wave itself. Figure 7.2 illustrates a typical wave soldering bath.
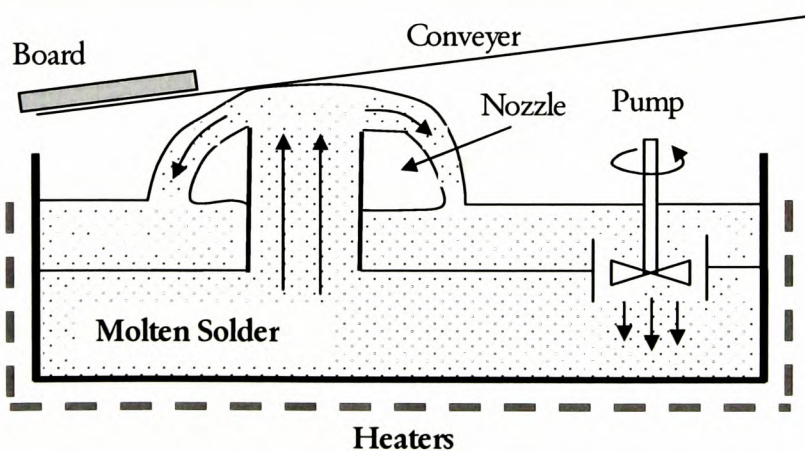
Figure 7.2: Wave Soldering Bath Components

As seen in Figure 7.2 the solder can be pumped using an impeller pump that pushes the molten solder into a pressure chamber. From there the solder moves upwards through the conduit and the wave nozzle. The formed solder wave is intended to contact the pre-heated board, to permit the solder to wet completely each metal pad on the board, and finally to allow the solder to drain away from all places where it must not remain. In a typical wave soldering application, discrete through-hole components are placed on top of the PCBs and glued underneath the boards. Because wetting occurs over a short period of time (order of seconds), during which the board is on top of the molten solder the components are not overheated.

As discussed earlier, wave soldering is normally performed using a turbulent wave prior the smooth solder wave to ensure proper application of the molten solder around the leads of components beneath the board. Figure 7.3 illustrates a real wave soldering bath and the concept of double wave soldering process.
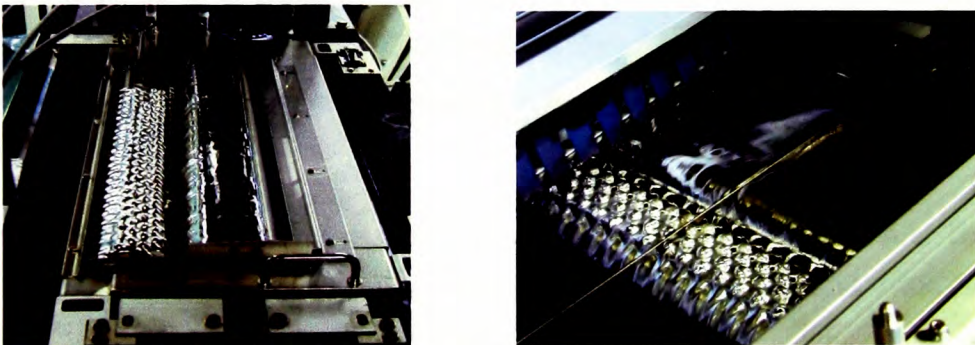
Figure 7.3: Double Wave (Turbulent/Smooth) Wave Soldering

7.2. Oxygen-Free Atmosphere Soldering

One of the most important key factors for successful wave soldering is removing the oxides from the metal parts. Normal air contains approximately 78 % nitrogen, 21 % oxygen and 1% taken from other gasses such as argon, carbon dioxide and etc. Thin layer of oxide forms when the metal joint parts and the molten solder itself are in contact with the air. As a result of this oxidation process at the top of molten solder wave, amount of dross is accumulated and trapped into the solder bath. The dross phenomenon has negative impact on soldering quality and may cause defects. It also leads to solder loss and requires maintenance procedure

for dross removal. Without effective removing of the oxide the soldering would not be possible.

To solve these problems a variety of fluxes and dross reducing technologies have been invented, developed and used in the manufacturing process. However, the fluxes have negative effects such as leaving undesirable flux residues on boards [186].

Numerous methods have been implemented and tested to reduce the amount of oxygen in the wave soldering machine. These are:

- Wave Soldering under a cover of hot glycerine flux [186];
- Wave Soldering in a vapour-phase atmosphere [186];
- Wave Soldering in nitrogen.

The most successful solution for oxygen-free soldering environment, for both lead and lead-free wave soldering, is soldering in nitrogen. In this process, nitrogen, either pure or with a small percentage of other gasses added, is injected in the soldering furnace with rates that will ensure the normal air replacement with a nitrogen rich environment. Unfortunately, nitrogen is expensive.

Successful oxygen-free soldering processes were implemented first in late 1980's. The first wave soldering machine processing in nitrogen atmosphere appeared in 1988 (SEHO, Germany) [186]. Since then a number of different wave soldering machines based on this principle have been proposed.

7.3. Wave Soldering in Nitrogen: Process Characteristics

Normally, the wave soldering in nitrogen must ensure a pure nitrogen environment with oxygen levels less than 20 ppm. However, if the major concern is dross, oxygen levels of 500 or even 1000 ppm are acceptable and will not cause failures in the soldering process for many applications [188].

One approach for nitrogen bared wave soldering is to place a nitrogen gas flow enclosure over the whole operation, normally a tunnel along the machine that contains the complete soldering line. Nitrogen is injected into the machine and pushes out the air that is trying to

enter the enclosure through the machine's openings (board entrance and exit). Some soldering machines are designed to have a nitrogen enclosure only around and on top of the solder waves, thus drastically reducing the amount of nitrogen required.

The second approach to obtain an inert atmosphere in wave soldering process uses shrouds - designed near the wave nozzles   that provide inerting only around the solder waves and on top of the solder pot. In this process the nitrogen injectors are located beneath the shrouds and near the solder waves. When the board passes on top of the waves, it starts to act, although only temporary, as part of the enclosure system for solder waves (top side). This helps oxygen reduction to reach low levels below 50 ppm and dross reduction over 80% [188].

The flow rate of nitrogen is one of the key control parameters for wave soldering in an inert atmosphere. Its consumption is critical for the cost of the process, dross reduction, solder loss, process maintenance and the overall efficiency of the oxygen free wave soldering.

The rest of this chapter presents optimisation modelling of the double wave soldering process. The aim is to identify optimal nitrogen flow rates for a particular machine specification. Nitrogen consumption is minimized (cost effective process) subject to restrictions that require low levels of oxygen around the molten solder waves (quality effective process). The study is performed within *VisualDOC-PHYSICA* virtual design framework and utilises a coupled **thermal-flow** analysis with integrated calculations for oxygen concentration.

This study was performed as a part of an ongoing project with an industrial partner developing a retrofit design for nitrogen injection in wave soldering machines. The modelling was intended to provide knowledge on flow of the injected into the soldering machine nitrogen and how it affects the oxygen levels near the solder bath. The ultimate goal was to investigate different what-if scenarios for a number of process parameters related to the wave soldering process and to correlate the predicted oxygen concentration levels to dross/ defect rate production. In this chapter, the investigated problem is used to demonstrate the process optimisation for this application.

7.4. Furnace Atmosphere Modelling: Test Case

7.4.1. *Test Case Specification*

The test case addresses the problem of modelling and predicting the flow profile and oxygen (O_2) concentration inside a horizontal oven (furnace) subject to nitrogen intervention. The furnace is 8.23 m long (L), and the height (H) and width (W) are respectively 0.152 m and 0.203 m. The furnace entrance and exit have actual openings of 0.076 x 0.203 m (height x width) managed by using adjustable shutter doors suspended above the furnace inlet/outlet doors.

The nitrogen (N_2) gas injectors are horizontally laid pipes along the furnace ceiling and along its width, with several holes pointed downward onto the solids/boards. If the furnace length is along the X axis of the Cartesian coordinate system and the oven entry (inlet) is at X=0, then the location of the three injectors is specified as X = 1.8288 m, X = 3.5052 m and X = 6.7056 m from the oven inlet (X = 0) respectively. Figure 7.4 details schematically the furnace domain and the injectors.
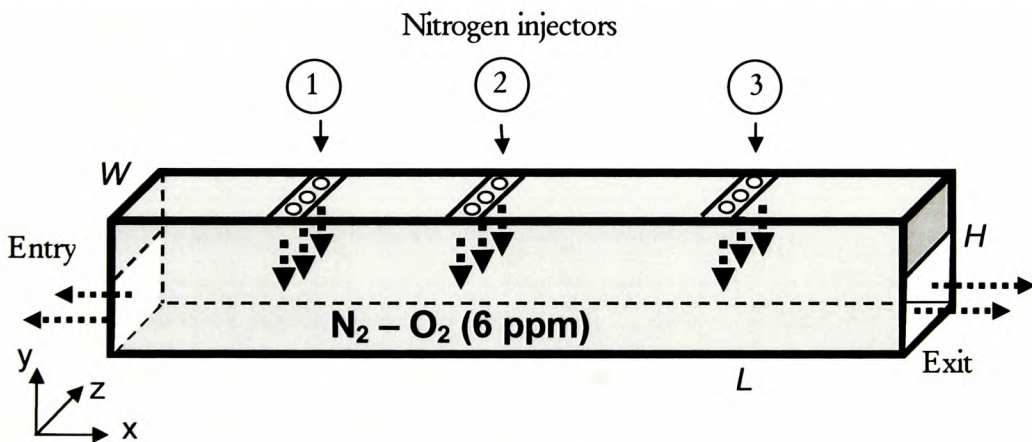
Figure 7.4: Schematic Drawing of Furnace

The injected gas is almost pure nitrogen containing approximately 6 ppm oxygen. Two different gas flow rates are tested and compared:

- **Scenario 1** (LOW flow rate): 2.12 m³/hour : 7.08 m³/hour : 2.12 m³/hour (total 11.32 m³/hour) through injectors 1:2:3;

- **Scenario 2** (HIGH flow rate): 3.54 m³/hour : 12.75 m³/hour : 3.54 m³/hour (total 19.83 m³/hour) through injectors 1:2:3.
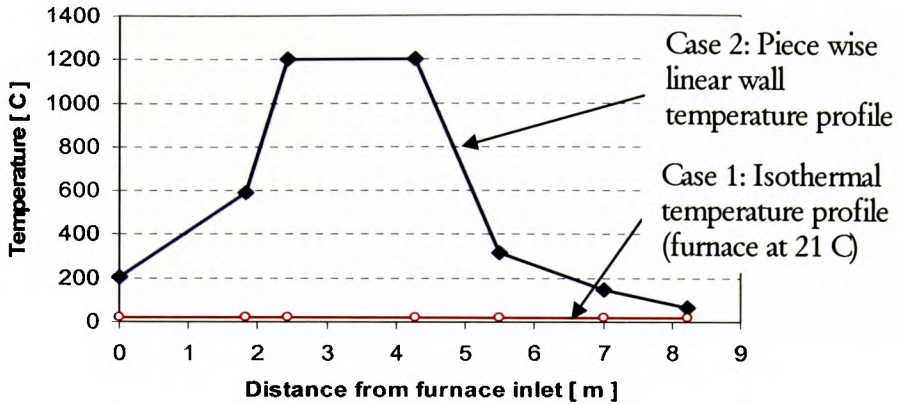
Figure 7.5: Furnace Temperature Profiles

The nitrogen flow rates are investigated for two different furnace temperature profiles:

- **Case 1** : Cold furnace with temperature 21 °C everywhere;

- **Case 2** : Furnace with a piece wise linear wall temperature profile (Figure 7.5). The temperature in the non-isothermal profile is a function only of the length along the furnace.

7.4.2. *Finite Volume (FV) Furnace Modelling and Analysis*

A 2D finite element representation of the furnace is used to analyse the temperature, flow and oxygen concentration characteristics across the oven subject to specified above nitrogen flow rates and temperature profiles. Each of the injectors is represented in the model as a nitrogen flow inlet at the corresponding location with length 0.0254 m. The nitrogen velocity at these inlets (required for analysis inputs) is calculated based on the gas flow rate and representation of injected pipe openings equivalent to 0.0254 x 0.203 m slots (length x width, 3D furnace) at furnace ceiling. The rest of the dimensions are taken as specified in the previous section.

Finite volume analysis of coupled heat, flow and oxygen concentration calculation using *PHYSICA* package is based on the following settings:

- Steady State analysis;

- Gas physical properties are reported in Table 7.1 [191]:

Table 7.1: Gas Physical Properties

Gas Property	Value	Unit
Kinematic Viscosity	Linear: 1.5E-05 at 20 °C 15.6E-05 at 1200 °C	$[m^2 s^{-1}]$
Specific heat	1.0057E+03	$[J\ kg^{-1}\ K]$
Ideal Gas Density	$$\rho = \frac{P\omega}{RT}$$ ω = 28.9 [kg kmol -1] molar mass P = 101325.0 [Pa] – pressure R = 8314 [$J\ kmol^{-1}\ K^{-1}$] - gas constant T [Kelvin, K] - temperature	$[kg/m^3]$
Thermal Conductivity	0.02624	$[W\ m^{-1}\ K^{-1}]$

- Gas velocity through injectors 1/2/3 as specified in Scenarios 1 and 2;

- Temperature profiles according to Case 1 and Case 2;

- Pressure 0 at the inlet/outlet of the furnace;

- Gas temperature of injected nitrogen of 20 °C;

- Buoyancy force modelled ([118], also Chapter 2, Equation 2.21);

- No-Slip momentum boundary conditions on the wall of the furnace;

- Turbulence calculations ON ($k - \varepsilon$ model) [118];

- *PHYSICA* convergence tolerance of solved variables 1.0E-03 (default) and O_2 concentration calculations with convergence tolerance 1.0E-05.

The analysis assumes that initially the whole furnace is filled with non-industrial air (21 % oxygen, i.e. 210 000 ppm) and the injected gas is almost pure nitrogen with only 6 ppm oxygen.

The ultimate response of interest in this analysis is the oxygen concentration in the furnace. Flow profiles in the furnace and phenomena such as air ingress in the oven are also observed.

7.4.3. *Test Case Modelling Results*

For the isothermal furnace (Case 1) the analysis shows that for both low (Scenario 1) and high (Scenario 2) nitrogen flow rates the whole furnace reaches a state of inert atmosphere. The oxygen concentration everywhere is 6 ppm and the gas flows out from both entrance and exit openings. No air ingress into the machine is detected in this operating environment.

The focus was on the study with non-isothermal furnace. The following results detail only this case (Case 2) for both low and high flow rates.

1. Temperature Profile Predictions

Figure 7.6 shows graph of temperature along the furnace at height of 1 inch. It clearly indicates the drop of the temperature at the central injector location caused by the low temperature injected nitrogen comparing with the furnace temperature at that location. Moving both sides of this central injector indicates rise of the temperature due to increasing impact of the furnace wall temperature and the buoyancy force. The injected gas from the first and third injector moves towards the entrance and exit respectively as seen from the flow profile in the furnace (Figure 7.7). This phenomenon causes a smooth decrease in the temperature as the entrance and exit are approached and temperature of the furnace walls drops down further.
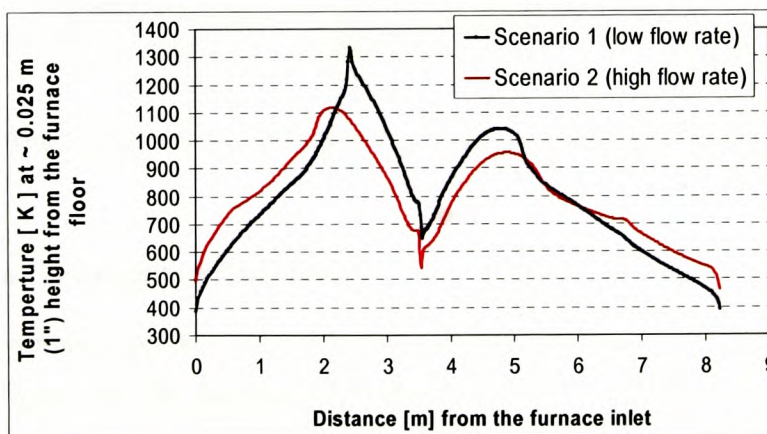
Figure 7.6: Temperature Profile along Furnace at 1 inch Height, Non-Isothermal Furnace

2. Flow Profile Predictions

For non-isothermal furnace one can expect an impact of the buoyancy force as the temperature inside the oven varies. Indeed, this modelling study shows that for both low and high nitrogen flow rates there is an amount of air ingress through the furnace entrance and exit. The air ingress decreases with increasing flow rate of injected gas, but the overall flow profile is similar. Figure 7.7 illustrates the flow profile at the furnace openings (entrance and exit) and three injector locations for Scenario 2 (high flow rate) and Case 2 (non-isothermal furnace) analysis.
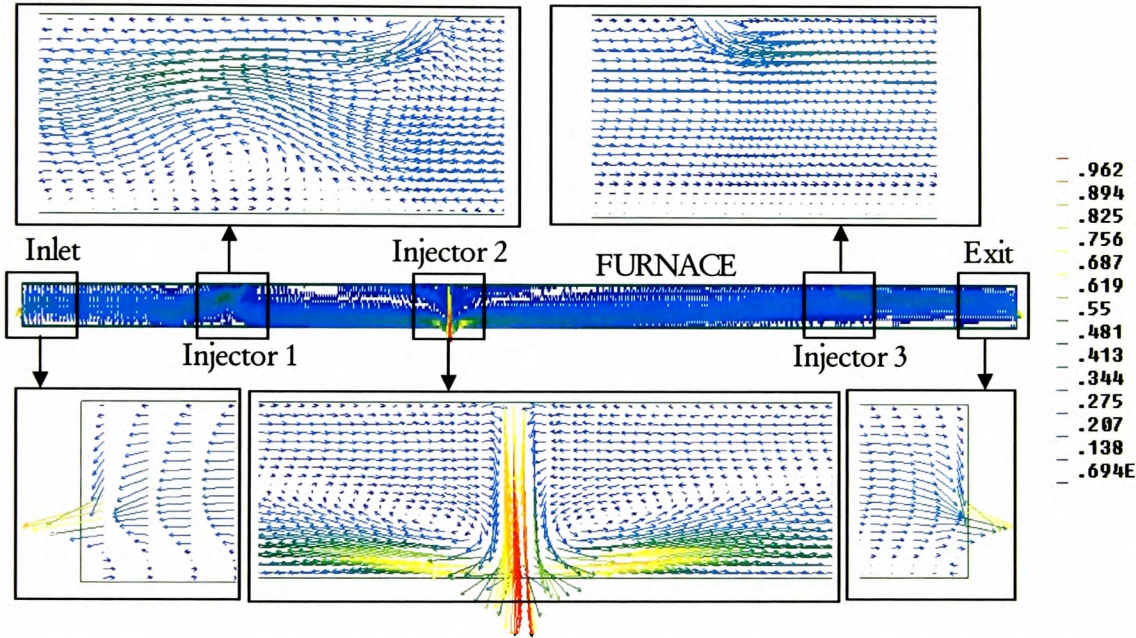
Figure 7.7: Flow Profile: Resultant Velocity [m/s] (High Flow Rate, Non-Isothermal Case)

3. Oxygen Concentration Predictions

Oxygen concentration profiles for both low and high nitrogen flow rates scenarios are provided in Figure 7.8. The ingress of air, which contains 21% oxygen, keeps the oxygen level high near the furnace openings, but elsewhere the oxygen concentration is 6 ppm due to the injected nitrogen. Obviously, higher flow rates of injected gas will decrease the amount of air ingress. The graph in Figure 7.9 illustrates the modelling results for oxygen concentration along the furnace from left to right at a height of one inch.

The results also indicate that there is a nitrogen flow rate that can ensure no air ingress. Of course, a higher flow rate of injected nitrogen would result in extra cost with no added benefit.
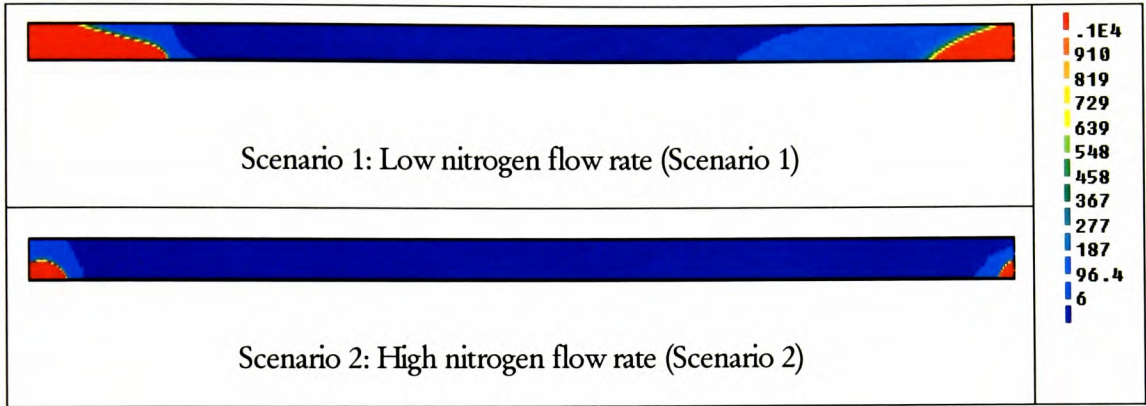
Figure 7.8: Oxygen Concentration [ppm] Profile
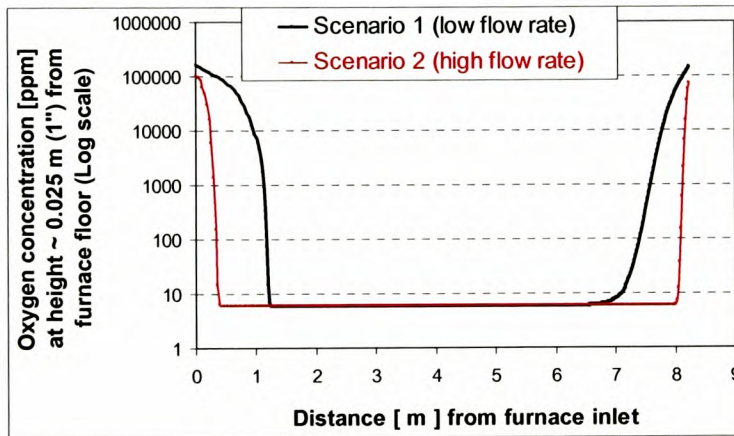
Figure 7.9: Oxygen Concentration along Furnace at 1 inch Height, Non-Isothermal Furnace

7.4.4. *Test Case Conclusions*

The above test case indicates that the model for gas flow and temperature is capable to capture important phenomena associated with the process such as the air ingress in the machine due to buoyancy forces. Air ingress inside the furnace is stronger for low flow rates of injected nitrogen. The existence of the air ingress phenomenon was confirmed by the

industrial partner. It is an important feature of the observed process that brings oxygen within the soldering bath even if the nitrogen injection is implemented.

Non uniform O_2 concentration across the furnace with lowest level of 6 ppm is observed in the region between the first and the third injector while at the furnace regions near the inlet/outlet oxygen concentration starts to increase dramatically (up to 19 % for Scenario 1 and up to 13 % in Scenario 2). The modelling results presented in Figure 7.9 were confirmed to be close to the experimental oxygen measurements available at the industrial partner side.

7.5. Wave Soldering Machine

The preliminary modelling work on the furnace test case described above is followed by an investigation and optimisation of the wave soldering process using a realistically defined machine. Schematic view of the wave soldering machine used in this study is given in Figure 7.10. Some major dimensions are also reported. The outlined specifications of the machine illustrate the inert atmosphere here is obtained based on the approach of using *shrouds*. The design of these shrouds intends to direct nitrogen around the waves. The soldering bath section is preceded by pre-heating tunnel area. The conveyer line throughout the entire machine is declined by angle of 5 degrees. The entrance and exit have height of 80 mm.
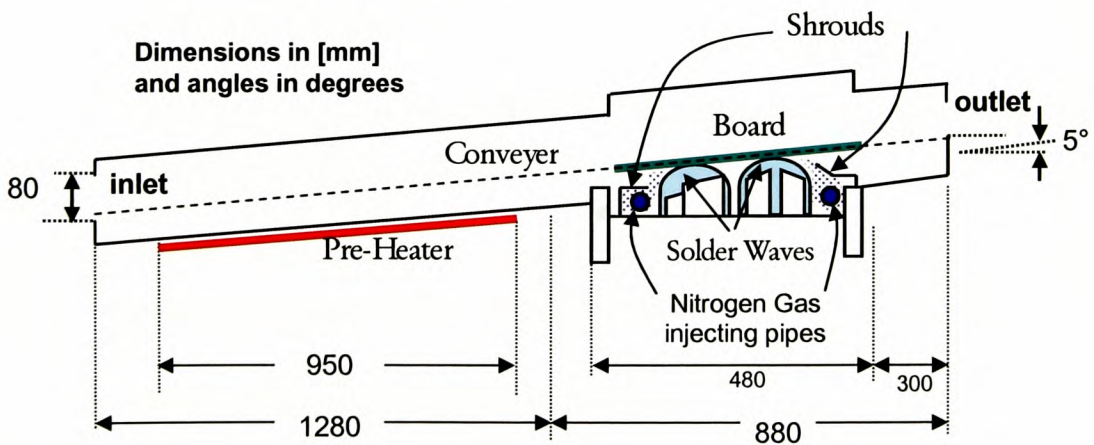
Figure 7.10: Schematic of Wave Soldering Machine and Some Important Dimensions

Two nitrogen injecting pipes are placed beneath the left (front) wave and right wave shroud with diameters 6 and 12.5 mm respectively. When the board passes over the waves it acts as an enclosure around the waves as shown in Figure 7.11. The figure also illustrates the soldering bath prototype used in the model.
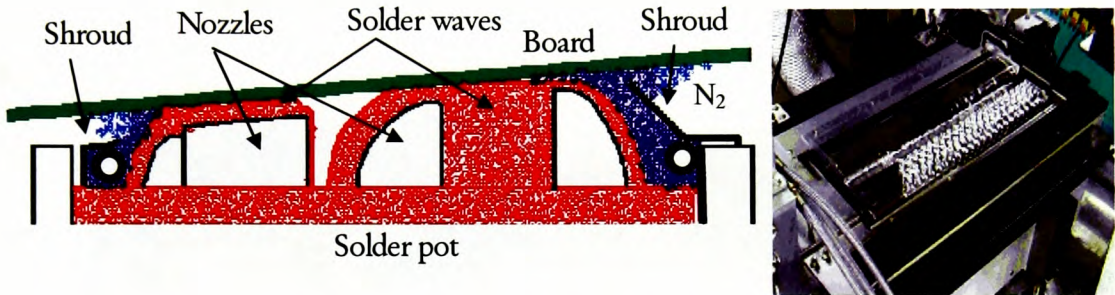
Figure 7.11: Schematic and Real View of Wave Soldering Bath Section

7.5.1. *FV Modelling, Analysis Settings and Design Tasks*

The wave soldering machine study is performed in the following stages:

- Build the CAD and finite element mesh model of the wave soldering machine;

- Set up of the FV analysis that predicts heat transfer, nitrogen gas flow, air ingress and oxygen concentration;

- Define the design optimisation task for economical nitrogen consumption (cost and quality effective process);

- Specify simulation-based optimisation strategy and perform the design optimisation task using the *VisualDOC-PHYSICA* software framework;

- Analysis of the optimised soldering process.

A two-dimensional model is chosen as a representation of soldering machine to reduce the computational cost of the simulations. The model of whole machine is given in Figure 7.12 with colours representing gas flow enclosure and solder waves. Detailed view of soldering bath section and corresponding finite element mesh is provided in Figure 7.13.
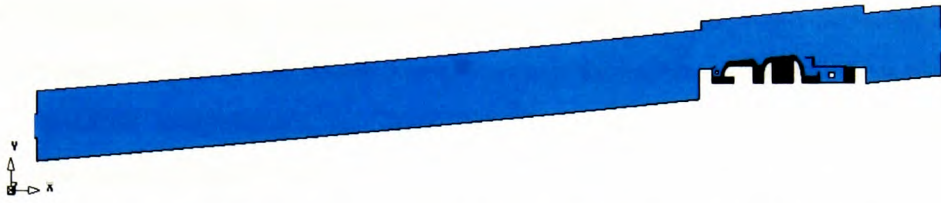
Figure 7.12: Wave Soldering Machine Finite Element Model
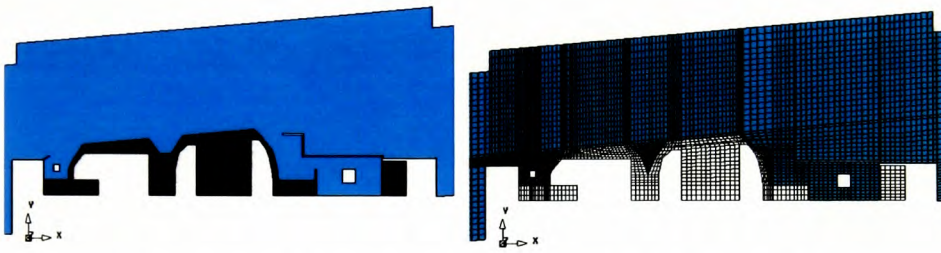
Figure 7.13: Soldering Bath – Geometry, Materials and Mesh

The summarised mathematical model for coupled heat-flow-O_2 concentration analysis is:

1. Governing Equations

- Continuity (Chapter 2, Section 2.3.1, Equation 2.18)

- Momentum Equations with buoyancy source (Chapter 2 Section 2.3.2, Equation 2.19 and Equation 2.21)

- Temperature (Chapter 2, Section 2.3.3, Equation 2.23)

- Scalar Variable (O_2 concentration) (Chapter 2, Section 2.3.4, Equation 2.24)

2. Boundary Conditions

- *Flow*: Nitrogen consumption is handled as volumetric flow rate (cubic meters per hour) injected through a pipe with specific diameter and length of 0.5 m in the third dimension of machine. This data is manipulated and transformed into velocity/mass flow rate boundary condition at the injector boundary (flow inlets, see Figure 7.14). The openings of the machine (entrance and exit) are modelled using outlet boundary conditions, i.e. pressure zero.

- *Temperature*: The boundary temperature profile of the wave soldering process assumes a piece wise linear function along the machine boundary, based on the specified temperatures in Table 7.2 for a number of boundary locations as illustrated in Figure 7.14.

Table 7.2: Temperature at Boundary Locations of Machine in Wave Soldering Process

Location at the Boundary (Figure 7.14)	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11
Temperature [°C]	20	120	120	260	260	127	20	20	52	52	20
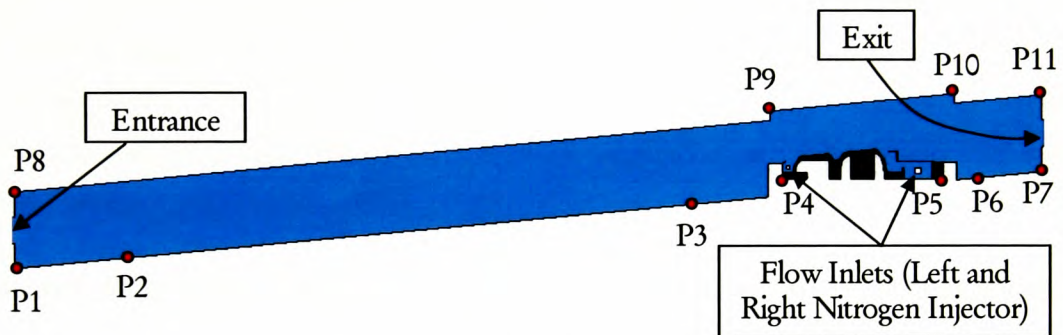
Figure 7.14: Points at Boundary of Machine used to Specify Boundary Conditions

The pre-heater is at temperature of 120 °C (P2 to P3) and the machine entrance and exit boundary is characterised with ambient temperature (P1 to P8, and P7 to P11). As the injected nitrogen is with ambient temperature, at the boundary of flow inlets (nitrogen injectors) temperature of 20 °C is specified. Lead free solder waves are included and modelled as heat source with constant temperature of 260 °C.

- *O₂ concentration*: Because injected nitrogen has 6 ppm oxygen, the boundary condition for O_2 at the flow inlet (i.e. boundary of nitrogen injectors) is 6 ppm and at the machine openings (entrance and exit) the value of 21% (non-industrial air) is given as a boundary condition.

3. **Initial Conditions**

In this steady state analysis the initial conditions are: zero velocity components for flow, temperature 20C everywhere in machine and oxygen concentration of 21%.

The rest of the analysis settings and the material properties used in the analysis are as outlined in the test case problem in Section 7.4.2. In addition, solder physical properties for thermal conductivity ($50\ \mathrm{W\ m^{-1}\ K^{-1}}$) and specific heat ($200\ \mathrm{J\ kg^{-1}\ K}$) are used.

7.5.2. Nitrogen Consumption Optimisation Task

Benefits of soldering in a nitrogen rich atmosphere have been already discussed and the advantages of this process outlined. However, the question for process optimality and overall cost requires careful observation of process parameters. The focus is on the amount (or the rate) of injected nitrogen. There is a clear trade-off for this specification. Higher nitrogen flow rates are good for preventing oxide formation processes and reduce dross and defects in assemblies but it involves higher cost due to the amount of nitrogen used. Lower nitrogen flow rates will reduce the financial cost, but may compromise the major concern that relates to the oxygen levels near the solder waves.

The rest of this chapter illustrates simulation-based optimisation of wave soldering process.

The ultimate goal is to minimise the total nitrogen flow rates through both injectors of the machine while still obtaining low oxygen concentration around the solder waves. To define this design problem, nitrogen flow rates through first (f_1) and the second injector (f_2) are specified as design variables. The objective function is the sum of these two design variables and represents the total nitrogen gas flow rate required in the process. The goal is to minimise the objective (i.e. to minimise the total nitrogen gas flow rate).

Based on preliminary analysis, four different locations (see Figure 7.15, points A, B, C and D) near the top of solder waves are specified as monitoring points for oxygen concentration. The aim of this analysis is to satisfy the requirement for oxygen concentration near solder waves at each of these points to be lower than a specified value. These have the form of inequalities of form "oxygen concentration at point $i \leq 1000$ ppm".

This study relies on lower limit of 8 m³/hour and upper limit of 20 m³/hour flow rate for both design variables f_1 and f_2 (the nitrogen flow rates through left and right injector respectively).
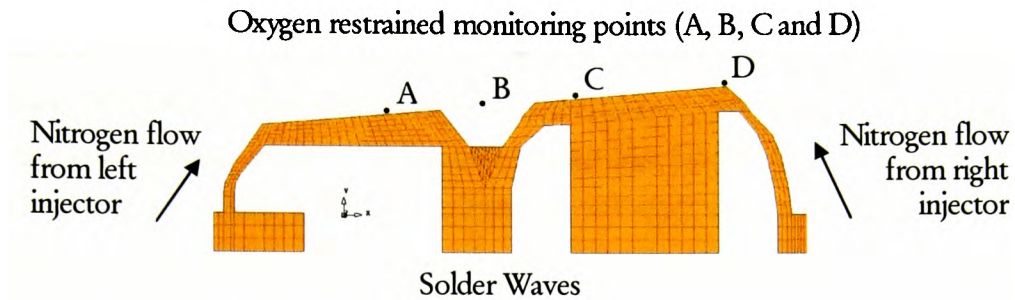
Figure 7.15: Four Monitoring Points for Oxygen Concentration Constraints

The final optimisation task is formulated as

Find the design point $X(f_1, f_2)$, i.e. the values of nitrogen flow rate through left injector f_1 and through right injector f_2 of wave soldering machine, that will

Minimize $f(f_1, f_2) = f_1 + f_2$ [m³/hour]

Subject to:

(1) $g_i(f_1, f_2) \leq 1000$ ppm, $i \in \{A, B, C, D\}$ (see Figure 7.15)

where $g_i(f_1, f_2)$ is oxygen concentration [ppm] at point i

(2) $8 \leq f_1 \leq 20$ (m³/hour)

(3) $8 \leq f_2 \leq 20$ (m³/hour)

(7.1)

As discussed throughout this thesis, in simulation-based design the optimisation functions are evaluated using data from numerical analysis. In this task the objective function is calculated directly for a specific flow rate as the sum of the two design variables. The predictions from finite volume analysis are used then only in constraint evaluations, i.e. oxygen concentration value at each of the monitoring points.

7.5.3. *Review of Optimisation Strategy and Optimisation Method*

The design task (7.1) for calculating the optimal nitrogen flow rates in the wave soldering process is solved by using the Incrementally Refined Response Surface (IRRS) approach as outlined in Chapter 4, Section 4.8.4.3. Optimisation calculations within an iteration of this

strategy are carried out by a non-gradient search technique called Particle Swarm Optimisation (PSO) [32, 156, 190]. Also, instead of using a polynomial form of the Response Surface (RS) model, this study integrates multi-quadratic function as approximation to oxygen concentration response. The multi-quadratic function RS model is defined in Section 7.5.3.2. The RS model, numerical optimisation technique and the ovetall design strategy differs from the design route demonstrated in the previous chapter. It is intended to demonstrate another approach for simulation-based design optimisation in microelectronics process design.

7.5.3.1 Starting Set Experimental Points for IRRS

IRRS requires finite element responses for the initial set of design points to construct the initial Response Surface model. The choice of these points can be random or may rely on experimental design (DoE). Generally, the latter approach will provide better quality for this initial RS approximation. The experimental points here are identified in the 2-dimensional design space of flow rate design variables $X=(f_1, f_2)$, where each of the design variables varies in the range 8 to 20 (m³/hour). The experimental design chosen here is the full factorial DoE, as outlined in Chapter 3, Section 3.3.7, with four levels for each design variable. In this case the experimental design points will be given by all possible combinations of four discrete values - 8, 12, 16 and 20 (m³/hour) - of the design variables f_1 and f_2 and will result in 16 design points. For each design point, i.e. particular nitrogen flow rate through the first and second injector, the finite element analysis predicts the oxygen concentration, heat transfer characteristics and flow profile for the soldering process.

The results of interest in the design task are oxygen levels at the monitoring points near the solder waves (Points A-D, Figure 7.15). Table 7.3 provides details on analysis results in terms of design experimental points (flow rates scenarios) and the corresponding levels of oxygen concentration at the four locations observed in the optimisation task.

Table 7.3 also shows that there are flow rate combinations through injectors that are feasible (concentration of oxygen below 1000 ppm) and injection rates that do not satisfy these design requirements. One can expect that there is optimal solution – it will come as output from the optimisation task solution.

The results reported in Table 7.3 clearly indicate that design points 12 and 16 provide minimum O_2 concentration, but unfortunately very high flow rates of nitrogen N_2 are required which is costly.

Table 7.3: Initial DoE and CFD Results for Oxygen Concentration at the Monitoring Points

Design Point No. (i)	Design Point $\overline{X}^i$	Design Variables		Optimisation Constraints			
		f_1 [m³/hour]	f_2 [m³/hour]	$g_A(f_1, f_2)$ O₂_point_A [ppm]	$g_B(f_1, f_2)$ O₂_point_B [ppm]	$g_C(f_1, f_2)$ O₂_point_C [ppm]	$g_D(f_1, f_2)$ O₂_point_D [ppm]
1	$\overline{X}^1$	8	8	1543.6	3467.7	6156.8	2654.3
2	$\overline{X}^2$	8	12	1113.2	2616.8	4775.4	1201.9
3	$\overline{X}^3$	8	16	673.7	1763.6	3357.9	666.1
4	$\overline{X}^4$	8	20	555.1	948.9	1368.1	218.4
5	$\overline{X}^5$	12	8	2476.5	4447.2	6289.9	2228.0
6	$\overline{X}^6$	12	12	1932.7	3414.7	4805.5	1205.4
7	$\overline{X}^7$	12	16	1264.9	2192.9	3041.2	719.7
8	$\overline{X}^8$	12	20	251.0	385.9	490.8	99.6
9	$\overline{X}^9$	16	8	2543.0	4635.2	6045.6	1414.9
10	$\overline{X}^{10}$	16	12	2159.3	3251.0	4318.3	709.0
11	$\overline{X}^{11}$	16	16	1105.1	1744.2	2116.0	402.8
12	$\overline{X}^{12}$	16	20	6.0	6.0	6.0	6.0
13	$\overline{X}^{13}$	20	8	2578.7	3759.4	3818.9	162.9
14	$\overline{X}^{14}$	20	12	2594.3	4906.3	5666.8	1743.4
15	$\overline{X}^{15}$	20	16	1718.4	3462.4	3880.0	881.9
16	$\overline{X}^{16}$	20	20	6.0	6.0	6.0	6.0

7.5.3.2. Multi-Quadratic Response Surface Model

Response multi-quadratic approximation [189] uses the following function:

$$g(X) = \sum_{i-1}^{m} a_i \sqrt{\left| X - \overline{X}^i \right|^2 + h} \qquad (7.2)$$

where $X \in R^n$ is the design point, i.e. the vector of n design variables, for evaluation, $\overline{X}^i \in R^n$ are the experimental design points ($i=1,...,m$) with known response values and h is so a called smooth parameter. The coefficients a_i in the multi-quadratic function are

computed by requiring the function in Equation 7.2 to fit exactly the given set of finite element analysis response data ($\overline{X}^i$, g ($\overline{X}^i$)) for experimental points ($i=1,...,$ m). This requirement results in solving a linear system of m equations with the coefficients in the multi-quadratic response model a_i ($i=1,...,$ m) as unknowns. The value of the smooth parameter h used in this study is set to one.

7.5.3.3. Particle Swarm Optimisation (PSO): Overview of Algorithm

The optimisation problem (7.1) is solved using the Particle Swarm Optimisation – a recently developed non-gradient based probabilistic search algorithm [32, 156]. The search for the optimal point in this method uses only the values of optimisation functions, and for our problem the required evaluations are based on the predictions from the constructed RS models. This eliminates the major disadvantage of the method – compared with gradient-based algorithms - associated with the huge number of function evaluations typically required.

Appendix A (see page 239) provides overview of the PSO algorithm. More details about the PSO can be found also in references [32, 156, 190]. Discussion on the rules for parameter settings of this algorithm and the numerical implementation strategies are available in reference [156].

7.5.4. Optimisation Results

The IRRS approach is adopted to perform the optimisation design task. At iteration of this process the set of experimental design points defined at the previous iteration is updated with an additional design point found as the optimal design point for the last iteration RS model. It is a refinement procedure for the RS model that starts to build approximation to each oxygen concentration constraint using more points near the true optimum thus making the RS model more and more accurate in this particular design space region. At each design iteration the optimisation problem is solved using the current RS models for oxygen constraints and the previously described PSO evolutionary algorithm. Although such an optimisation method requires a massive number of function evaluations it does not impose any extra computational cost because RS is used for these function evaluations and no FE analysis is required. The PSO does not require any gradient calculations and uses only

optimisation function values. The benefit of applying the PSO method is in its ability to locate the global optimum as it searches the whole design space. When the RS model uses multi-quadratic functions for a response approximation, the above capability becomes important because of the possible multi local optima.

Convergence criterion for the IRRS method is the coincidence within specified tolerance between RS and FE prediction for the optimal solution at the current iteration. This study uses tolerance of 2 % difference between FE and RS predicted values for oxygen concentration at the monitoring points of soldering machine to terminate the optimisation process at that specific design iteration. Starting with RS models, based on 16 experimental points (see Table 7.3) and after three optimisation iterations, i.e. updating RS models with three more design points corresponding to the optimal points at the last three iterations (see Table 7.4), the convergence criteria are met. Optimal solution at the convergence iteration is observed as the final solution of our optimisation problem.

Table 7.4: Design Points used for RS Refinement and Corresponding Oxygen Concentration

Design Point No. (i)	Design Point $\bar{X}^i$	Design Variables		Optimisation Constraints			
		f_1 [m³/hour]	f_2 [m³/hour]	$g_A(f_1, f_2)$ O₂_point_A [ppm]	$g_B(f_1, f_2)$ O₂_point_B [ppm]	$g_C(f_1, f_2)$ O₂_point_C [ppm]	$g_D(f_1, f_2)$ O₂_point_D [ppm]
17	$\bar{X}^{17}$	9.760000	19.99000	414.8441	648.0366	816.2125	129.4117
18	$\bar{X}^{18}$	9.078570	20.00000	530.2921	842.8163	1104.193	178.0011
19	$\bar{X}^{19}$	9.419838	19.88858	548.4274	877.4758	1158.174	198.4633

Table 7.5 provides details on the last refined multi-quadratic Response Surface model,

$g_C(X) = \sum_{i=1}^{19} a_i \sqrt{\left| X - \bar{X}^i \right|^2 + 1}$, for oxygen concentration at the location of monitoring point

C. This approximation uses 19 design points (Tables 7.3 and 7.4) – those from the initial DoE set plus the refining three design points associated with the IRRS procedure at last three design iterations. Similar multi-quadratic RS models are obtained for the rest of the oxygen constraints.

Table 7.5: Coefficients a_i ($i=1,\ldots,19$) of the refined at last optimisation iteration RS Model $\left(g_C(X) = \sum_{i=1}^{19} a_i \sqrt{|X - \bar{X}^i|^2 + 1}\right)$ for Oxygen Concentration Constraint at Monitoring Point C

RS Coefficient a_i	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}
Value	-28	59	77	-388	-63	-76	-189	-241	-428	280

RS Coefficient a_i	a_{11}	a_{12}	a_{13}	a_{14}	a_{15}	a_{16}	a_{17}	a_{18}	a_{19}	
Value	341	78	761	-680	-492	483	6217	5555	-11071	

Figure 7.16 is plot of the objective function (total gas flow rate) as function of nitrogen flow rates for the first and second injector. It is a simple linear combination of the two design variables. The converged optimal solution determines flow rates of f_1=9.42 m³/hour and f_2=20.0 m³/hour through first and second injector respectively. While the optimal nitrogen rate flow through second injectors goes to the upper imposed limit, the rate through the first injector has a value within the specified rate limits.
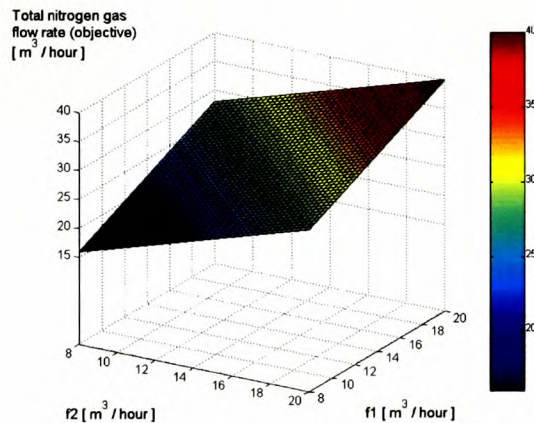
Figure 7.16: Surface Plot of Total Nitrogen Flow Rate (Objective)

All constraints specified in the design task (7.1) are met. There is only one active (i.e. the constraint is satisfied as equality within the tolerance) constraint at the optimal solution for nitrogen flow rates. This is the constraint for oxygen concentration at location Point_C $g_C(f_1, f_2)$ - with value of 1000 ppm (within numerical tolerance, actual RS value 1002.9 ppm).

Figures 7.17 and 7.18 illustrate response surface models obtained at the last iteration of IRRS procedure. These plots clearly outline oxygen concentration dependences by nitrogen flow

rates at the observed four locations. The contour line of oxygen concentration of 1000 ppm (the limit) in the flow rate design space is also given in these plots for each of the responses.

As emphasised, the most critical location is the one at Point_C, i.e. if oxygen concentration at point C is lower than 1000 ppm it will be lower than this limit for the other three points A, B and D. The reason for Point_C location on top of the solder waves to be critical with respect to the oxygen concentration level is discussed later, under Section 7.5.5., and can be explained with the characteristics of the flow profile (illustrated in Figure 7.19). Obviously this information is not available at the beginning and this optimisation approach is an effective way to find this important process characteristic. In fact, in automated optimisation these relationships are taken into account through the optimisation process itself.
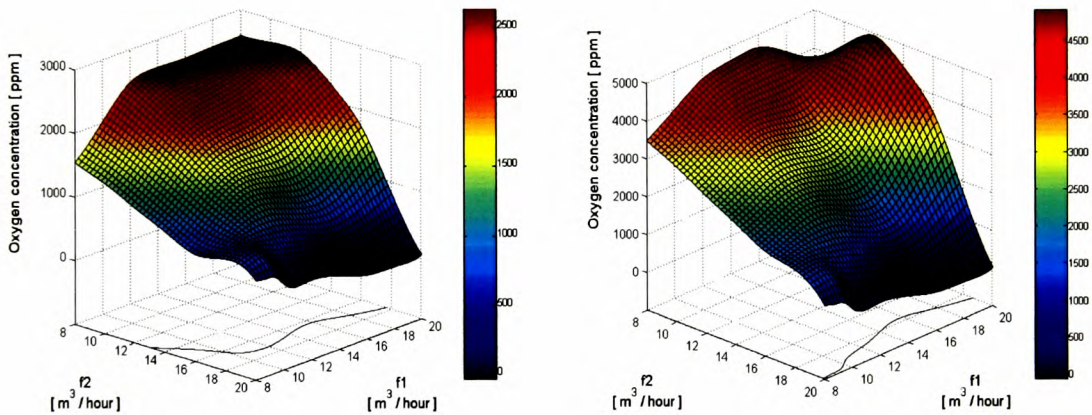
Figure 7.17: RS Models of Oxygen Concentration at the monitoring Point_A (left) and Point_B (right) (Contour Level for 1000 ppm O_2 Concentration Given)
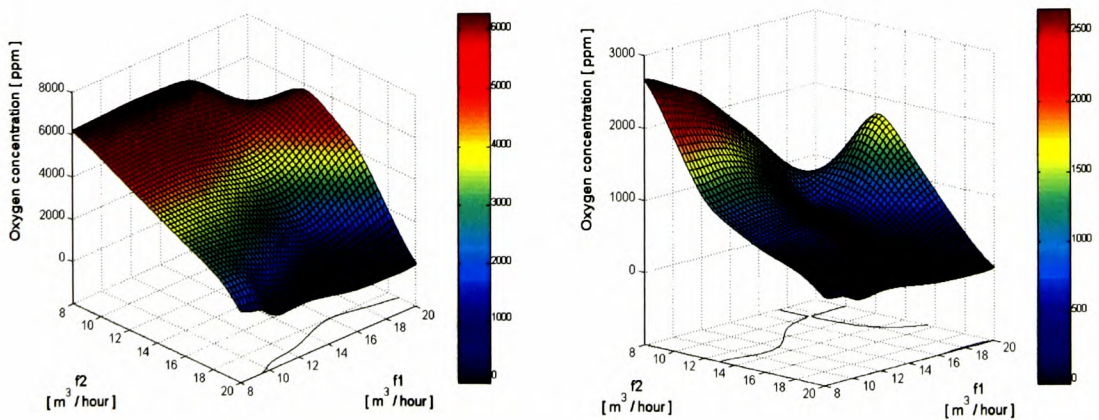
Figure 7.18: RS Models of Oxygen Concentration at the monitoring Point_C (left) and Point_D (right) (Contour Level for 1000 ppm O_2 Concentration Given)

Optimisation results for nitrogen flow rate in the soldering machine are summarized in Table 7.6.

Table 7.6: Nitrogen Flow Rate Optimisation Results

Design Variables	Lower Limit	Optimal Value	Upper Limit
Nitrogen flow rate : first injector f_1 [m³/hour]	8.0	9.42	20.0
Nitrogen flow rate : first injector f_2 [m³/hour]	8.0	20.0	20.0

Optimisation Task Function	Optimal Solution			Percentage difference in RS-FVA predictions
	Last Refined RS Model	Finite Volume Analysis (FVA)	Upper Limit	
Objective Function: total nitrogen flow rate, $f_1 + f_2$ [m³/hour]	29.42	29.42	N/A	N/A
Oxygen constraint, Point_A [ppm]	495.6	486.0	1000.0	1.98 %
Oxygen constraint, Point_B [ppm]	779.5	770.7	1000.0	1.14 %
Oxygen constraint, Point_C [ppm]	1002.9 (active)	996.8	1000.0	0.61 %
Oxygen constraint, Point_D [ppm]	164.4	163.2	1000.0	0.7 3%
Total FE analysis				19
Total CPU time to perform one FE analysis (800 MHz processor)				3 hours 19 min 48 sec.
Total design points evaluated using PSO in optimisation process using last RS model				3800
Total CPU time to perform PSO using last RS models (800 MHz processor)				36.8 sec.

7.5.5. *Finite Volume Analysis: Optimal Process*

In general, the injected nitrogen cannot feed the entire machine enclosure with nitrogen. The soldering machine observed in this study is designed to achieve an inert atmosphere only near the waves. The idea is to use nitrogen but at minimum cost.

The CFD results of coupled *thermal-flow-O_2 concentration* finite volume calculations for optimal process defined in the previous section of the chapter are discussed here. These results provide an understanding of the gas flow profiles in the machine, air ingress caused by

the temperature and how the underlying interaction affects the oxygen levels in the soldering machine, particularly near the solder waves. In the following plots, the CFD results show the optimised process where the nitrogen flow rate through the first injector is f_1 =9.42 m³/hour and the nitrogen flow rate through the second injector is f_2=20.0 m³/hour.

Because of leaned machine's tunnel and the buoyancy force there is air ingress through the machine entrance. It starts to heat inside and to move upwards, into the soldering bath and further to the exit. This air flow mixes with injected nitrogen at the bath section and increases the oxygen level of 6 ppm that characterises the injected nitrogen. Figure 7.19 shows the resultant velocity contour plot of injected nitrogen and air ingress near the soldering bath. Despite the highest flow rate through the right injector it is not sufficient to withstand the combined flow of air ingress and nitrogen from the first injector. The overall flow is towards the exit of machine, mainly due to the buoyancy phenomenon.
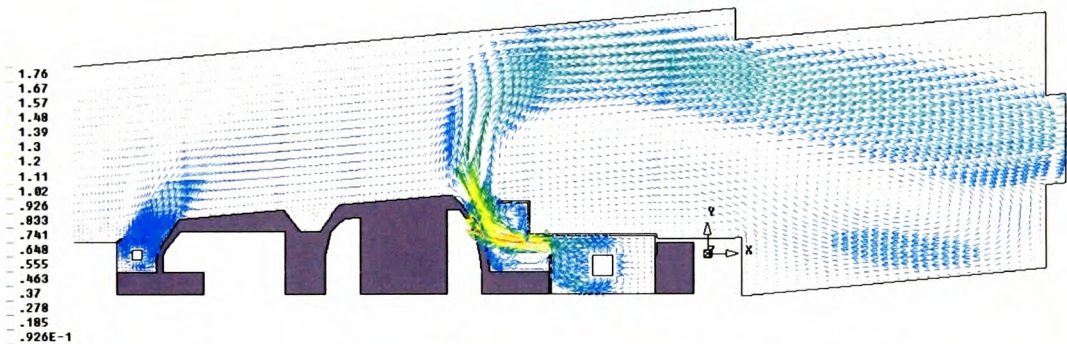
Figure 7.19: Flow Profile (Resultant Velocity [m/s]) at Soldering Bath Section
for Optimal Process (f_1=9.42 and f_2=20.0 m³/hour)

Details of flow phenomena at soldering machines entrance and exit are illustrated in Figure 7.20. As already discussed the buoyancy force causes air ingress through the entrance at the bottom side while at the same time at the top of the entrance some amount of gas exits. The exit of the machine acts as an outlet for the flow with no air ingress taking place. The outgoing flow is strong enough to cause a flow vortex at the bottom side of the machine near the exit.
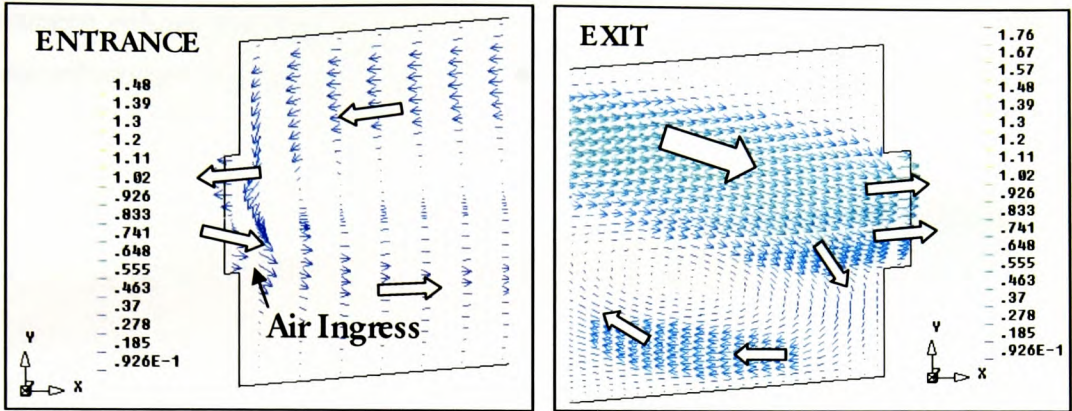
Figure 7.20: Machine Entrance/Exit Flow Profile (Resultant Velocity [m/s]) for Optimal Process (f_1=9.42 and f_2=20.0 m³/hour)

Temperature profiles are illustrated in Figure 7.21. Near the pre-heater and the solder waves the gas heats and moves upwards along the machine. However, both the air ingress and nitrogen enter the machine at ambient temperature. As a result of this the gas domain is heated at relatively low degree. Clearly, the temperature boundary conditions are important and affect the flow and subsequently the oxygen levels. Lower temperatures in the machine will reduce the buoyancy effect and the amount of air ingress at the entrance. In this case, the effect of injected nitrogen will be more prominent.
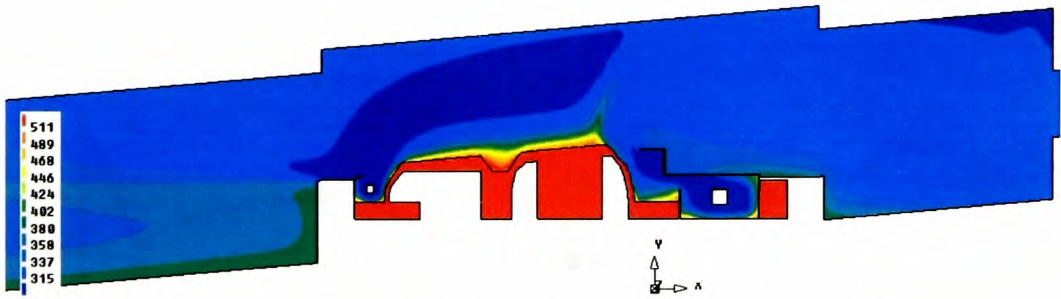
Figure 7.21: Temperature [Kelvin] Profile for Optimal Process (f_1=9.42 and f_2=20.0 m³/hour)

Finally, important results on oxygen concentration are observed. Contour levels of oxygen along the soldering machine (Figure 7.22) vary from 21 % down to 6 ppm. The machine enclosure from entrance to soldering bath is dominated by air ingress and characterized with high oxygen concentration (above 50 000 ppm). Around the solder waves, the injected

nitrogen reduces the oxygen to required levels. Clearly, beneath the shrouds where the injected nitrogen tubes are located, almost pure inert atmosphere is obtained (6 ppm).
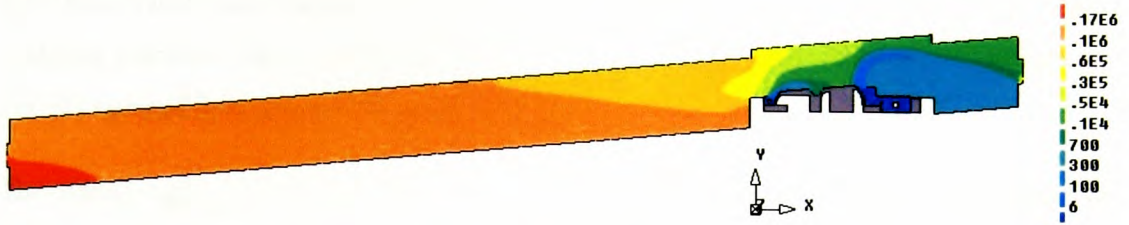
Figure 7.22: Oxygen Concentration [ppm] Profile for Optimal Process
(f_1=9.42 and f_2=20.0 m³/hour)

Based on the constraints used in optimisation of nitrogen flow rates, oxygen concentration below 1000 ppm along top of the waves is also ensured. Oxygen and temperature variations along the line from the left to the right shroud and just above the solder waves are given in Figure 7.23. The red line in these plots corresponds to the domain line at the top of the waves along which the results are presented. By solving the optimisation problem defined in this section, along the entire surface of solder waves oxygen concentration below imposed limit of 1000 ppm is obtained. The same figure outlines the critical area for concentration – near monitoring point C, at which the oxygen is at the level of the limit of 1000 ppm.
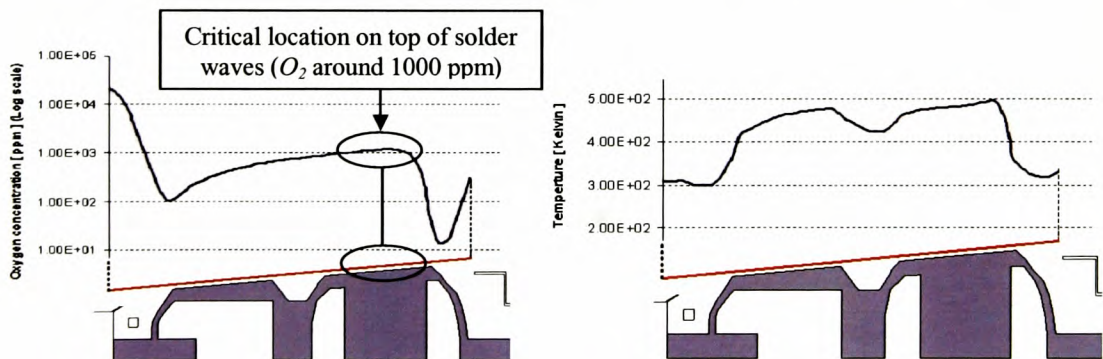
Figure 7.23: Oxygen Concentration [ppm] and Temperature [K] Variations
above Solder Waves for Optimal Process (f_1=9.42 and f_2=20.0 m³/hour)

7.6. Conclusions

This chapter has presented optimisation technology combined with predictions for fluid flow, heat transfer and oxygen concentration. The application area demonstrated is the wave soldering machine. The modelling approach has demonstrated how optimisation can quickly identify optimal flow rates for nitrogen injection

The virtual design of optimal nitrogen rates for wave soldering machine are obtained in efficient and effective manner. The potential savings from implementing inert atmosphere soldering with optimal nitrogen consumption could be significant. The presented study is focussed on worst case process with no board presented. In such a scenario to obtain low oxygen levels around the solder waves will require relatively high nitrogen consumption. In reality, the nitrogen consumption should be such to ensure required oxygen concentration at discrete time intervals when the printed circuit board passes over solder waves. In this case significantly lower nitrogen injection is required because the board and the shrouds act locally as part of the nitrogen enclosure.

From modelling prospective this study has demonstrated the Incrementally Refined Response Surface approach that utilises efficient non-gradient optimisation techniques. In this case, the Particle Swarm Optimisation method was applied to solve the problem for most economical nitrogen consumption in the lead-free wave soldering process.

CHAPTER 8

MULTI-PHYSICS MODELLING AND OPTIMISATION FOR THERMAL MANAGEMENT SOLUTIONS IN ELECTRONICS DESIGN

8.1. Introduction in Thermal Management of Electronic Systems

In a general aspect the term *thermal management* refers to all aspects of electronic systems cooling and in some cases heating. This includes design of heat generating components, placement and layout of devices, selection and layout of fans, heat sinks and other cooling devices, effective airflow through the overall electronic system and in particular around the active components with the highest heat dissipation. Systems using boxed electronic components such as microchips all require thermal management [97,111, 194-196]. The ultimate goal of thermal management is to keep the electronic components at or below their maximum operating temperatures.

Proper thermal management is achieved when heat is transferred from the devices to the system air, which is then vented out of the system.

Thermal analysis and different aspects of the thermal design in general are also included in the scope of the thermal management.

8.1.1. *Importance of Thermal Management*

Thermal management plays a major role in the electronic and packaging design of power electronic systems. A thermal design solution takes a combination of heat-reducing devices, factors and conditions to ensure the reliability of electronic systems.

Thermal management becomes an enormously important issue in electronic equipment design due to the need to pack more components into smaller spaces and then operate it reliably in extreme conditions and environments. The thermal management of implementing the best possible thermal solutions in electronic products can benefit enormously by using computational mechanics and fluid dynamics virtual prototyping.

8.1.2. *Thermal Management Issues*

A major part of in-service failures of the electronic systems is temperature related and due to the heat generated by components. The thermal fluctuations which the products experience during their operational life become even more severe with each new generation of components generating more heat than the last.

In modern electronic systems the heat-dissipating components are typically placed within highly confined spaces such as sealed metal cases. If such enclosure is not subject to some form of cooling, very high temperature gradients have to be expected throughout all components, air and the casing itself. The danger comes from the different rates of expansion and contraction of the different materials, e.g. silicon, solder, FR-4, etc., used in the package as they heat up and cool down.

Since all materials used in the components are joined in some way, very high thermal stresses are built up and these lead to different type of failure mechanisms such as cracking or de-lamination. Common defects are the die cracking and the cracking of the solder joints.

It is vital to implement thermal management solutions for each particular electronic product that will minimise the heat dissipation of components, and respectively the thermal mismatch and impact of the temperature.

8.1.3. *Thermal Management Challenges*

The primary concern is maintaining the junction temperature of power semiconductors and operating temperature of passive components within the manufacturer's limits. To keep these characteristics within the prescribed limits forced-air cooling and intelligent design and component placement are required. The reliability of power electronic systems depends on the success of preventing excessive heating of the components within an electronic product.

At package level the challenge is the trend of the constantly increasing temperatures at which the packages operate and hence the temperature range of a power-on/ power-off thermal cycles. These factors result in very large temperature gradients in the active components and this may lead to performance problems and critical thermally induced stresses within the die and die attach.

At board level the thermal problems relate to the induced in-plane strains and relative displacements of the board-package interconnection.

At system level thermal management requires solutions for effective air cooling and system design including board and component placement and distribution inside the casing.

8.1.4. *Heat Reducing Devices and Conditions*

Fans are one of the common solutions to insure forced-air convection cooling of the hotter power electronic systems. In the last decade the reliability and performance of the fan has improved, e.g. introduction of a motor drive ICs that control fan speed in accordance with the ambient temperature.

Other heat-reducing devices are the heat sinks. They are stationary, reliable devices that eliminate heat via conduction and radiation. Fin structures can be pins, flat plates, or other variations of the flat plates. Heat sinks combined with fans provide a more efficient cooling system.

Thermal interface materials used in mounting power semiconductors to heat sinks also help heat removal from the system and provide electrical isolation and good thermal conductivity.

The computer modelling and analysis can predict the system or process behaviour only if correct material data is specified. For some material properties there is a lack of data, or the data may vary from source to source. Many factors have impact when material properties are investigated through experimental tests and the data is collected, i.e. storage conditions of the materials may change with time some of the properties.

Thermal conductivity data used in electronic industry [192] is given in Table 8.1.

Table 8.1: Thermal Conductivity

MATERIAL	Thermal Conductivity [W/mK]
CONDUCTORS	
Aluminium	121-208
Aluminium Nitride	170-230
Copper	395
Silicon	84-118
Solder	22-52
INSULATORS	
Epoxy	0.35-0.87
Epoxy fibreglass	0.24
Nylon	0.24
PTFE	0.24
Still air	0.026
Thermal pastes	1.1

8.2. Heat Dissipation

The heat transfer occurs in the following dissipation mechanisms:

8.2.1. Heat Dissipation by Conduction

Heat conduction is a mechanism of heat transfer from one solid body to another by direct contact (Figure 8.1). The heat flux Q through a surface with area A observed through this phenomenon can be represented by the Fourier's low for heat conduction (k is the thermal conductivity). If T denotes the temperature and the vector normal to the surface area A is along the X direction, then the heat flux Q is defined as

$$Q = -kA\frac{dT}{dX}$$ (8.1)
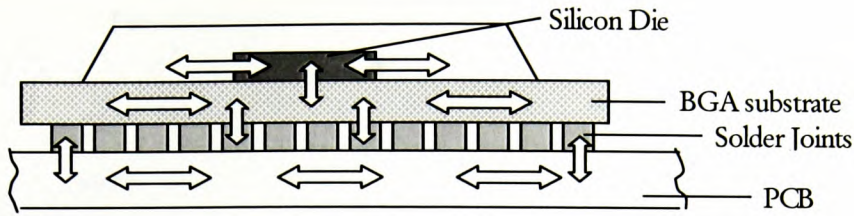
Figure 8.1: Heat Conduction Mechanism (BGA Example)

In electronic systems normally the heat generated by the components is transferred by conduction throughout all solid parts in the package, e.g. from die to the solder joints, into bond wires, circuit board and the other electronic items, and further outward into the casing. Finally, if the casing is placed or attached on a solid surface it will conduct the heat to this ambient surrounding, but also will dissipate heat by natural convection and radiation from the exposed surfaces.

The thermal contact resistance between connected materials can be reduced by using flat and smooth contact surfaces and thermal interface materials (e.g. pastes) [192].

8.2.2. *Heat Dissipation by Convection*

The components with heated surfaces transfer heat to the ambient air by convection. The fluid flow (a cooler fluid, e.g. air) carries the heat away. It is an important mechanism of cooling and may occur due to natural convection, but more often forced convection is implemented in electronic systems (Figure 8.2). In this figure, U_∞ stands for the velocity of the cooling fluid flow.

Newton's law of cooling states:

$$Q = h\,A\,(T_W - T_\infty) \tag{8.2}$$

where A is the surface area, Q is the heat dissipation, T_W is the surface temperature and T_∞ is the temperature of cooling fluid. The convective heat transfer coefficient (h) is function of fluid properties, geometry and convection type.
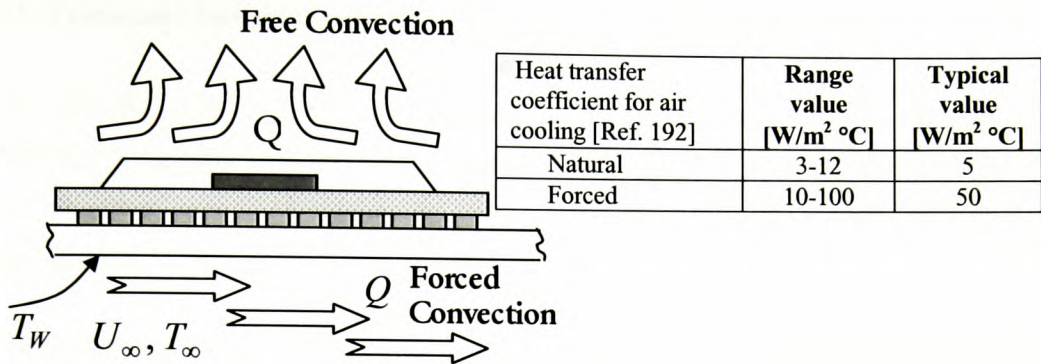
Figure 8.2: Cooling by Convection

In the analysis tool *PHYSICA* computational fluid dynamics solution procedure determines the convective heat transfer through the energy governing equation. The local heat transfer coefficients are determined implicitly for each heated object.

8.2.3. *Heat Dissipation by Radiation*

Radiation is the transfer of heat directly from a hot surface to a cooler surface or to the ambient surroundings by line-of-sight radiation, without direct contact. In electronic systems, heat is transferred by radiation from the hotter surfaces of IC's and the attached components directly to the inside of the casing.

Radiation, striking a surface, has the conservation property (Figure 8.3)

Incident =
Reflected + Transmitted + Absorbed
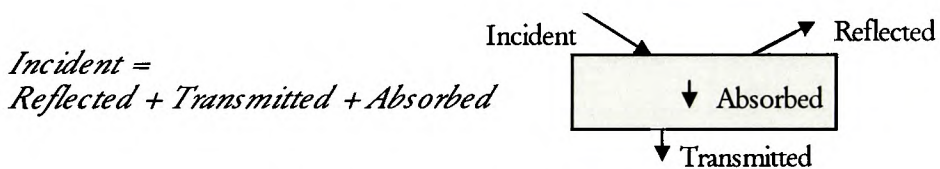
Figure 8.3: Radiation Phenomenon

In general solids do not transmit radiation and hence they have zero transmissivity property. Emissivity, ε by definition, compares the heat emitted by a surface with that emitted by an ideal radiating body (a "blackbody"). For most solids the emissivity is equal to the absorbtivity. Details on numerical modelling of radiation can be found in reference [103].

8.3. Thermal Characterisation: Modelling and Simulation Tools

The electronic systems virtual prototyping and the coupled thermal-flow-stress analysis of a thermal management problem take place in the early stages of the design process. This allows the industry to save much time, money and to go to the markets early with reliable products that meet the customers' demands for greater performance.

The software tools for multiphysics analysis allow to test in a virtual manner the electronic products subject to any operational and environmental conditions before building their real physical prototypes [97, 111]. While a software package for thermal simulations can only predict the temperature across the electronic system, the more advanced and sophisticated tools for multiphysics analysis take also into account other physical phenomena and the existing complex interactions. The temperature and velocity profiles of the cooling airflow and heat, displacement, stress and strain results associated with the solid parts of the electronic product due to the thermal load are predicted in this case within the same computational analysis.

The modelling of thermal management problems often requires simplifications of some parts of the electronic system. The *compact* models - which model the system at a particular level - can be built and used to represent the whole thermal model accurately, without modelling in details the components from the sub-level.

The *compact* models [195] are faster to compute than the detailed models. For complex electronic systems, using simplified entities in the model is often the only option. If the intention is to predict the overall airflow temperature, velocity, heat transfer and stresses at system level then the details with respect the individual components, from modelling point of view, can be avoided. Computational analyses for thermal management problems in electronics require advanced numerical techniques for coupled calculations of the physical phenomena. The analysis tool *PHYSICA* is capable to analyze such complex problems where the thermal-flow-stress interactions in the electronic systems are observed using integrated calculation procedures for computational mechanics and fluid dynamics.

A number of other commercial software tools for thermal analysis are available and used to aid the thermal design of electronic systems:

- FLOTHERM [www.flotherm.com] (from Flomerics Ltd. [201]) is a software package developed specifically for thermal analysis and simulation of electronic equipment. FLOTHERM uses Computational Fluid Dynamics (CFD) techniques to predict air flow and heat transfer in and around electronic components, heat sinks, boards, and complete systems. The tool FLO/STRESS is a new package for stress calculations developed by Flomerics and integrated within their thermal analysis tool FLOTHERM;

- The major EDA software vendors (Mentor Graphics, Cadence, Zuken Redac) supply limited board-level thermal analysis software tools such as the AutoTherm (www.mentor.com/autotherm) and the Thermax software packages;

- ICEPAK [www.icepack.com] - a joint development of Fluent Inc. and ICEM-CFD Engineering - is a fully interactive software tool, used also for thermal management analysis by design engineers in electronics industry. By predicting the air flow and the heat transfer at the component, board or higher-level of the system, ICEPACK can be used to improve the design performance and to reduce the need for initial physical prototyping;

- ANSYS (from Ansys Inc. [www.ansys.com]) is widely used for detailed thermal analysis of complex geometries and thermo-mechanical stress prediction;

- Other software tools for thermal analysis are developed by: C&R Technologies [www.crtech.com], Dynamic Soft Analysis Inc. [www.betasoft-thermal.com], Harvard Thermal Inc. [www.harvardthermal.com], Innovative Research Inc. [www.inres.com], Maya [www.mayahtt.com], MicRed [www.micred.com].

8.4. Thermal Management Illustrative Example

The rest of Chapter 8 outlines the optimisation modelling study where complex multi-physics analysis approach of integrated thermal-flow-stress calculations are used for optimal thermal management solution in electronic system design. Some results on the problem are published [185,193] and presented as keynote presentations at two different international forums on electronic packaging and design.

8.4.1. *Description of the Electronic System*

The illustrative example is on an electronic system that is represented by a number of heat dissipating components assembled on the printed circuit board. The board assembly is placed within a box (casing) which is subject to air cooling from the fan. The electronic system and components distribution are given in Figure 8.4.

The effect of the fan air cooling is represented as airflow through an opening at one side of the casing and the other two openings - at the opposite side of the box - act as outlets for the flow. Apart from the PCB, there are two other solid parts (Obstacle 1 and Obstacle 2) inside the box. They act as obstacles to the air flow and change in a highly non-linear way its profile. Three different electronic components that dissipate heat are assembled within this casing: one (HS1) on solid part near the flow inlet (Obstacle 1), and the other two, HS2 and HS3, are chips mounted onto the circuit board. The components HS1 and HS2 are at fixed locations. The component of interest is the chip HS3 represented by a Plastic Ball Grid Array (PBGA) package on the board.
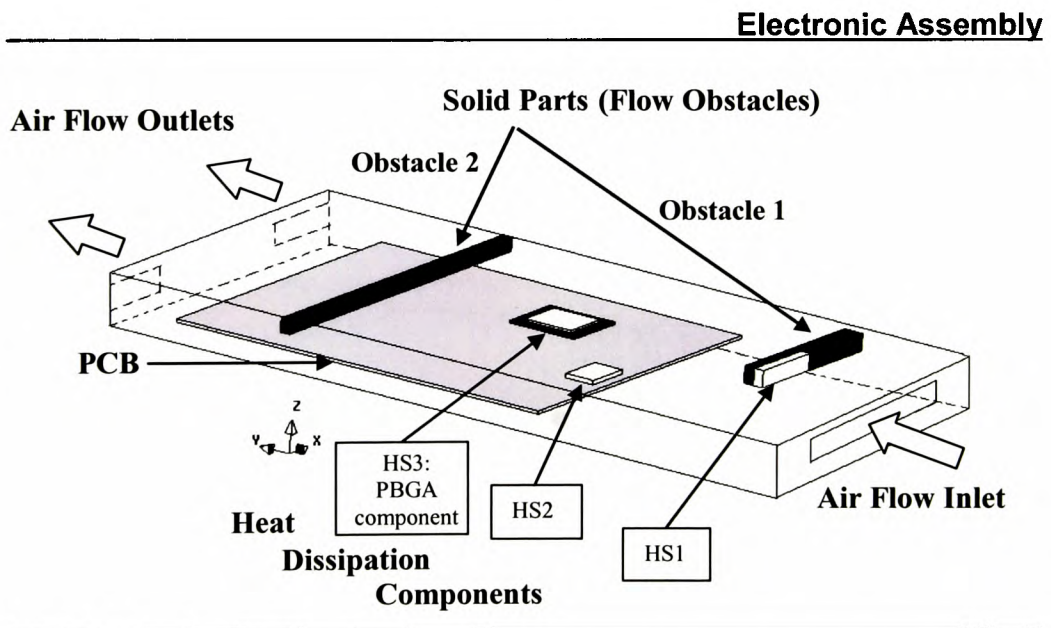
Figure 8.4: Schematic View of Designed Electronic System

8.4.2. *Finite Element Model of Electronic System Geometry*

The package under detailed observation has a typical PBGA structure with substrate and attached silicon die. The die is covered by a plastic overmold compound. This example assumes that the package is underfilled. The heat dissipated from the die conducts across the PBGA substrate and solder balls/underfill layer, and further across the printed circuit board.

Only the PBGA component (HS3) is modelled with an increased degree of details in the geometric finite element representation. The two other heat sources, HS1 and HS2, are modelled without any specific details. The PBGA component is implemented as a compact model to the level *substrate-die-mold compound* without representing explicitly the solder interconnects. A uniform layer that corresponds to the underfill encapsulation of solder joints is modelled between the package and the board. The bond wires as well as the chip bonding glue are neglected.

This example for thermal design requires modelling at the system level – an electronic systems that consist of a casing with placed mother board and other solid parts, plus number of assembled components subject to air cooling. As a result of this the above assumptions and simplifications in the modelling of the PBGA and the other two components are necessary.

Figure 8.5 shows a schematic drawing of the PBGA package.
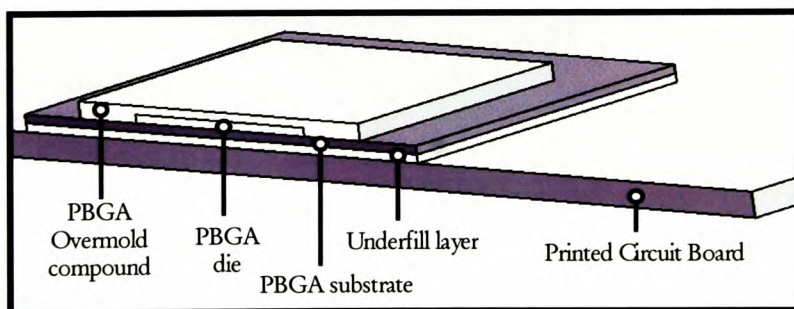
Figure 8.5: Modelling Details of PBGA Component

The modelling of the printed circuit board and the rest of the solid parts inside the box follows all details illustrated in Figure 8.4. Table 8.2 details the important dimensions of this electronic system.

Table 8.2: Geometric Specifications of Electronic System

Electronic System	Dimensions [mm]
PCB dimensions	175.0 x 195.0 x 1.5
HS3 (PBGA component)	
PBGA substrate	35.0 x 35.0 x 0.5
- PBGA die	15.0 x 15.0 x 0.5
Mold compound on die: height	0.5
Underfill layer height	0.5
Casing	190.0 x 375.0 x 19.0
Fan Air Flow Inlet Hole	110.0 x 10.0
Fan Air Flow Outlet Holes (each)	40.0 x 10.0

8.4.3. *Computational Grid*

The computational mesh of finite elements consists of 64728 nodes and 54188 elements throughout the whole domain (air flow part of the sealing, the circuit board and the rest of the solid parts and components within the box) (Figure 8.6).
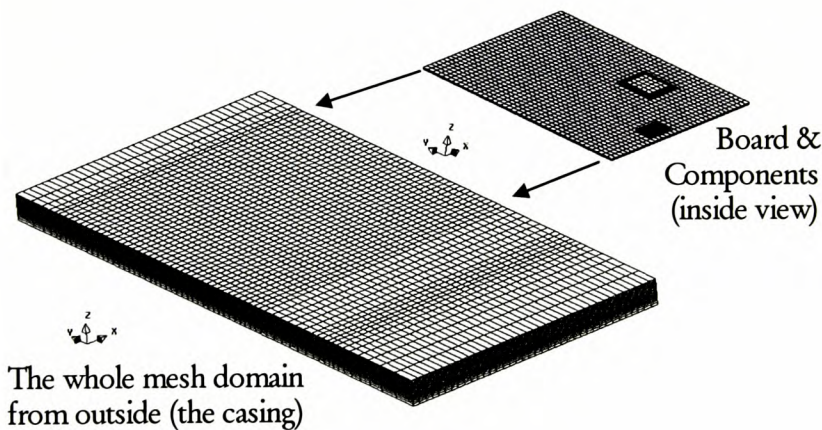
Figure 8.6: Computational Mesh

The model and the mesh grid are generated in such a way to ensure that the uniform mesh element size and density across the whole PCB board (PCB x-y in-plane) are exactly the same as the mesh grid throughout the PBGA component (in the same x-y plane) .

The generated grid is specified in this manner to allow movements (placements) of the PBGA component everywhere onto the board with discrete steps equal to the mesh element

size in the plane of the board. Based on such mesh implementation in this finite element model, we can move (and place) the PBGA component almost continuously on the board and still have, regardless the position of the component, an unchanged mesh domain throughout the entire domain. The reason is that the PBGA package itself, as being modelled, can be "fitted" with its own mesh into the mesh elements placed on top of the board.

This strategy is implemented because: (1) it simplifies the overall solution procedure of the investigated thermal management problem for identifying the optimal placement of the PBGA on the board and (2) it offers mesh independent design process of the PBGA placement by eliminating any mesh effects due to the different locations of the component that will evaluated during the optimisation process.

8.4.4. *Heat Dissipation in the Die*

Since the silicon die has high thermal conductivity, the heat dissipation can be observed in the model either as generated by the whole die or by a thin layer inside the die. In both cases the die temperature is very similar. In this study, the heat generated in the die is represented in the model using a uniform heat dissipation of the whole die. The power of the PBGA die is assumed to be 5 W.

8.4.5. *Material Data and Analysis Settings*

A major part of the heat dissipated in the component is transferred to the printed circuit board but also a part flows through the component surface to the ambient air. The dominant phenomenon here is heat transfer by convection and conduction. In the case of forced convection (cooling airflow) the heat dissipated through the convection mechanism increases significantly. Therefore the airflow specification and profile plays a key role in the component cooling.

In this study all materials are modelled as elastic, with properties detailed in Table 6.2 (Chapter 6).

Thermal conductivity of the board is calculated using the *rule of mixture* - copper (12%) vs. FR4 (88%) - and the related property values for these materials. The thermal conductivity of

the bulk PCB is 48.0 W/mK and this value accounts for the higher conductivity due to the presence of copper traces and thermal vias in the FR4 material.

The heat dissipation from components in power-on mode is modelled using fixed heat flux condition. The initial temperature for all parts within the sealing is 20°C. Temperature of 20°C is also used to specify the boundary condition at the external boundary of the sealing.

The flow calculations in the multiphysics analysis assume that the air behaves as an incompressible fluid subject to buoyancy forces. Calculations of the buoyancy forces use the Boussinesq approximation. The modelling requires a buoyancy source per unit volume into the i-th momentum equation as described in Chapter 2:

$$S_i = -\rho_{ref}\, \beta\, g_i\, (T - T_{ref}) \tag{8.3}$$

where ρ_{ref} is the reference density, which is equal to the constant density of the air, β is the thermal expansion coefficient (volumetric definition), g_i is the gravity component in the i-th direction, and T_{ref} is the reference temperature corresponding to the reference density constant.

The air velocity at the inlet of the box is specified to 1m/s (flow boundary condition).

For the multiphysics simulations in this study, the properties of the air are taken as approximately those at 25 °C. These air properties are reported in Table 8.3.

Table 8.3: Physical Properties of Air

Air Properties at 298 K (25 °C)	
Specific Heat	1.007E +03 [m² sec⁻² K⁻¹]
Thermal Conductivity	2.630E-02 [kg m sec⁻³ K⁻¹]
Density	1.161E +00 [kg m⁻³]
Thermal Expansion	3.500E-03 – 8.0E-06 * (T-273) [K⁻¹]
Laminar Kinematic Viscosity	1.590E-05 [m² sec⁻¹]
Gravity (acting in the negative z-direction)	9.810E +00 [m sec⁻²]

The CPU time required to perform a single multiphysics analysis of the investigated electronic system is 210 minutes on Alpha DEC computer with 450 MHz processor.

8.4.6. *Thermal Management Problem and Optimisation Task*

The thermal management problem is to identify the optimal location of PBGA (HS3) placement on the feasible PCB area. This optimal placement must ensure the lowest possible levels of stress in the silicon die of PBGA component. Since the die stress is developed as a result of the thermal load due to the heat dissipation, i.e. die temperature increases from the power OFF state temperature up to the operating temperature of the die, this thermal design problem requires solution which will lower the temperature in the component. Subsequently, this will improve the reliability of the whole assembly and the component in particular.

The other heat sources in the domain are with fixed positions. A change in the PBGA placement involves changes in the temperature field across the entire box domain, cooling air flow profile and magnitude, and the overall stress distribution across the solid parts of the electronic assembly. Different chip placements will be characterized with different junction temperature of die and magnitude of thermal stresses.

The following study illustrates *the multiphysics modelling* and *optimisation* design approach to idenrify the optimal placement of the PBGA component associated with the electronic assembly in Figure 8.4.

Two design variables, X-coordinate and Y-coordinate of the chip central point with respect PCB horizontal plane, are used to determine the PBGA component location. Based on the electronic system geometry, the feasible design space for both these variables in the co-ordinate system applied to the model is given by the lower bound of 17.5 mm and the upper bound of 137.5 mm (Figure 8.7). The initial PBGA placement is specified at location with co-ordinates $X=117.5$ mm and $Y=37.5$ mm (PCB in-plane). A location on the board is associated with the central point of the PBGA package (Figure 8.7).

The goal (objective) in this thermal design problem is to find those feasible values of X and Y (i.e. placement of the PBGA component) that will ensure the lowest possible stresses in the silicon die of PBGA package. In mathematical form the optimisation problem states:

Chip central point with (X, Y) coordinates in the X-Y plane (board)

(137.5, 137.5)

Design space Y

(17.5, 17.5)

X
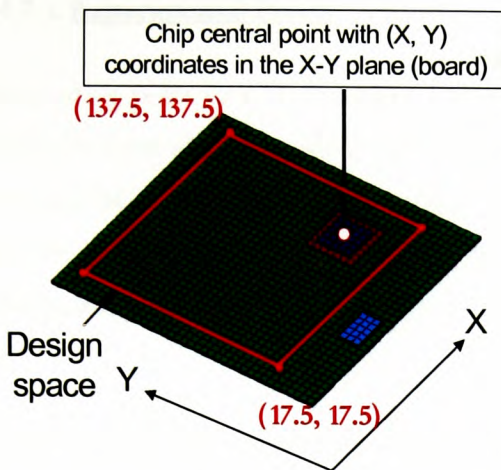
Figure 8.7: Design Space for PBGA Placement

Find these values of X and Y that will

Minimize Die Von Mises Stress $\qquad$ (8.4)
Subject to:
$17.5 \leq X \leq 137.5$ $\qquad$ (8.5)
$17.5 \leq Y \leq 137.5$ $\qquad$ (8.6)

Initial design $X^\circ = (117.5, 37.5)$

This design problem (8.4)-(8-6) states a general unconstrained optimisation problem where two design variables are optimized such as to minimize the objective function subject only to design variable limits only.

8.4.7. *Optimisation Procedure*

Design of Experiments approach (*DoE*) is exploited within the sensitivity and optimisation analysis by:

1. Identifying the experimental design points, i.e. locations on printed circuit board for PBGA placement;

2. Undertaking a multiphysics analysis at each design point, i.e. PBGA placement, to obtain the electronic system thermal-flow-stress responses at these design points. Distributed usage of the computer recourses is recommended;

3. Using the response data – the stress magnitude for silicon die of PBGA component - to construct response surface (RS) approximation; Analyze the fit characteristics of RS model and the accuracy of this approximation;

4. Perform sensitivity and optimisation analysis using the generated response surface;

5. Performs steps 1 to 4 in reduced design space to obtain more accurate results.

8.4.7.1. Experimental Points

The starting point in a DoE analysis is to establish the set of experimental design points at which the finite element analysis will be undertaken to provide system response data (i.e. the stress in PBGA die). The Von Mises definition of effective stress σ is used to quantify stress level in the silicon. The stress representative value used in the following calculations is volume weighted average of Von Mises stress associated with the mesh elements in silicon die.

In this study a full 3-level composite experimental design for two variables is used [143, Chapter 3]. This set contains 9 design points. Once the design points have been identified, the next step is to calculate the electronic system responses at these points, in particular the stress across the die of PBGA component. Each design point, being (X, Y)-coordinates that identify the PBGA position onto the PCB, represents different scenario for the PCB design and related characterisation of the sealing domain and cooling process.

At each of the design experimental points the multiphysics analysis is run using the *VisualDOC-PHYSICA* capabilities. The Von Mises effective stress in the die of PBGA component is obtained for all 9 locations. Table 8.4 shows the results from this DoE analysis denoted as a *DoE_1* for future references.

Table 8.4: *DoE_1* Points and PBGA Die Stress Responses

DoE_1 Points	PBGA X-coord. [mm]	PBGA Y-coord. [mm]	Average Stress in PBGA Die [MPa]
DoE_1 Point 1	137.5	137.5	50.72272
DoE_1 Point 2	137.5	17.5	56.37742
DoE_1 Point 3	17.5	137.5	24.68072
DoE_1 Point 4	17.5	17.5	30.40224
DoE_1 Point 5	137.5	77.5	47.72418
DoE_1 Point 6	17.5	77.5	24.28095
DoE_1 Point 7	77.5	137.5	40.98372
DoE_1 Point 8	77.5	17.5	53.23265
DoE_1 Point 9	77.5	77.5	39.27462

After the component responses have been obtained at the experimental design points, the next stage is to construct a second order polynomial approximation to the observed stress in

the PBGA die. The coefficients in RS approximation are determined by fitting the stress response data using the least squares approach, in a way similar to the presented in Chapter 6. The RS approximation for die stress σ provides directly the evaluation of objective function values at different design points in the optimisation procedure. The constructed RS model is used in the optimisation to avoid any further calls to the finite element analysis programme. As already discussed, this is the major difference between the DoE/RS optimisation strategy and the direct gradient based optimisation search.

The implemented strategy to have an unchanged mesh for the entire domain of this model, whatever the design configurations (i.e. any PBGA placement), suits the DoE approach in the above thermal management problem. The finite element models that correspond to the design experimental points can be constructed without major difficulties. This will not be the case if the direct gradient algorithms are used instead. They require special and relatively complex software routines to implement in an automated manner the different continuous placements of the PBGA component on the board. Additional problems - related to the subsequent automated mesh generation will also exist. The implementation of the direct gradient optimisation for this problem is a highly sophisticated task without offering any significant benefit when compared to the undertaken DoE strategy.

8.4.7.2. *DoE_1* Approximation and Residual Analysis

Based on *DoE_1* analyses reported in Table 8.4, the stress σ_1 in silicon die of component as function of component placement on the circuit board is approximated using the following full quadratic Response Surface polynomial (*RS_1*):

$$RS\_1 \ Model:$$

$$\sigma_1\ (X,\ Y) = A0 + A1*X + A2*Y + C12*X*Y + B1*X**2 + B2*Y**2$$

A0	= 29.8968099768
A1	= 0.4445781676
A2	= -0.30881813796
C12	= 4.64055555555e-006
B1	= -0.00151822935
B2	= 0.00156666509

(8.7)

The statistical tools of Residual Analysis, Analysis of Variance and statistical efficiency measures [Chapter 3] are used to observe the quality and accuracy of the generated RS

approximation. These techniques have shown the degree at which the coefficients predicted for RS polynomial provide a good fit and accuracy to the experimental data in Table 8.4. All residual analysis results and the ANOVA table for *DoE_1* response surface are reported in Appendix B (see page 241). The G-efficiency criterion, for example, estimates 91% prediction capability of the *DoE_1/RS_1* model (measure of the uncertainty of the approximated values produced by the RS model).

In Figure 8.8 (left) the predicted from the multiphysics analysis values for stress in the silicon die are compared with those obtained using the constructed RS polynomial approximation. The difference in the predicted and the actual value at each of the experimental points is the residual at that point (see Appendix B). The adjusted coefficient of multiple determination R^2_{adj} (see Chapter 3, Equation 3.32) a measure of the RS degree of fitting the experimental data is 95% and indicates a good fit of stress values data at the experimental points. The root mean square error of the stress in die - based on the response model (8.7) is estimated to 2.67 MPa, and the related coefficient of variation (see Chapter 3, Equation 3.34) is 6.5 %. This coefficient of variation defines how large is the error compared with the average die stress value.
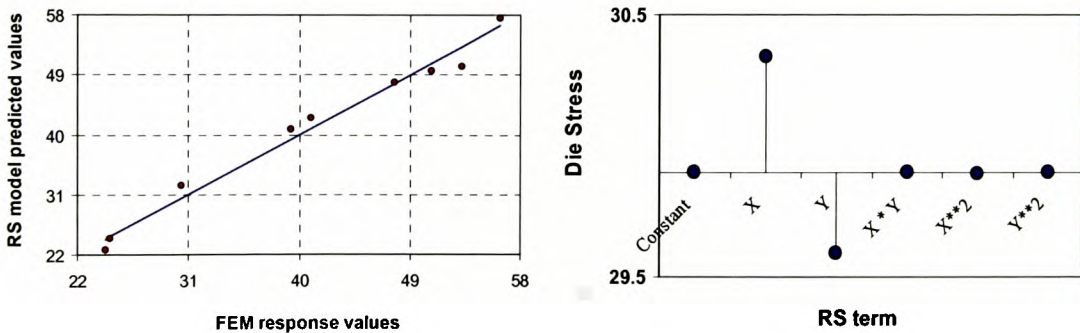
Figure 8.8: *DoE_1* Actual (FEM) Response vs. Predicted (*RS_1*) σ_1 (left) and Plot of Design Variable Sensitivities (right)

Apart from the benefit of avoiding actual runs of FE analyses in the optimisation process, the Response Surface Modelling has the additional advantage to perform *sensitivity analysis*. This sensitivity analysis helps to identify which design variables and their interactions are important and influence the electronic system responses, in particular what is the impact of

the PBGA location on stress level in the silicon die (Figure 8.8, right). These sensitivities are obtained from the coefficients in the RS approximation.

As normally observed, the constant term here is the one with the highest influence in the RS model. The RS terms that correspond to the linear coefficients have also significant impact. The sensitivity graph in Figure 8.8 (right) shows that in this design space the PBGA die stress is more sensitive to the first design variable, i.e. the component position in the X-direction of PCB plane, than to the second design variable (the Y co-ordinate). The linear interactions and the second order terms of the polynomial approximation have an insignificant impact on the stress response. The die stress sensitivity is almost entirely dominated by the RS linear terms. This sensitivity analysis shows that the stress in the die is approximately 44 % more sensitive to the X-direction of chip placement on the board that to its Y-direction placement.

Figure 8.9 shows the *DoE_1* RS model (8.7) in terms of an approximate relation between the chip placement on the PCB and the corresponding average Von Mises stress in the PBGA die. The bottom part of the same figure illustrates the PCB and the initial placement of the chip (*X*-coordinate equals 117.5 mm and *Y*-coordinates equals 37.5 mm).
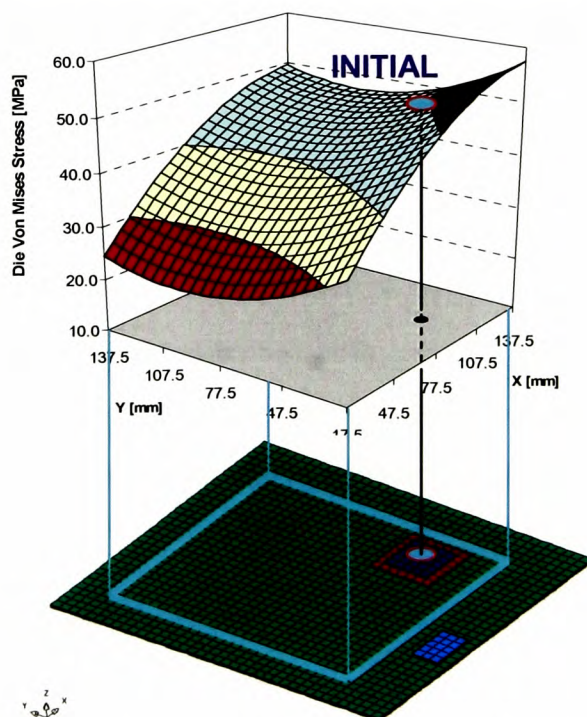
Figure 8.9: PCB Design Space with Mesh and the Initial Design: *RS_1* Model

8.4.7.3. Optimisation Based on *DoE_1* RS Model

The optimisation process for calculating the optimal placement of the PBGA component is performed using the Sequential Quadratic Programming method [26]. In this optimisation process, every time when the technique requires evaluation of the die stress at a certain location of the chip on the circuit board, the objective function is approximated using the *DoE_1* RS model (8.7). The numerical optimisation procedure identifies the optimal design point – the solution of the design task (8.4)-(8-6) - in terms of a continuous optimal solution as reported in Table 8.5.

Table 8.5: Optimisation Results based on *DoE_1* and *RS_1* model

DESIGN VARIABLES

Name	Lower Bound	Optimal Value	Upper Bound
X	17.500	17.50000	137.500
Y	17.500	98.53316	137.500

OBJECTIVE FUNCTION

Name	Type	Target Value	Optimal Objective Value
Die Stress	Response	minimise	22.00156

The required solution of the optimisation task in this study is in fact a discrete solution as the chip movements have to match the PCB mesh density. As a result, the extra step of obtaining that discrete solution, based on the continuous one, is undertaken. At this discrete optimal design point the X-coordinate of the chip placement is equal to 17.5 mm and the Y-coordinate is 97.5 mm. The approximate value of the stress in the die obtained using the Response Surface model is $\sigma_{max}^{RS} = 22.0032$ MPa. With the same optimal design point, Finite Element analysis is run to estimate what is the exact response value for the thermal stress. The exact simulation results predict value for PBGA die stress $\sigma_{max}^{FEM} = 23.1097$ MPa. The absolute error between the above approximate and exact predictions at the discrete optimal point is $err_{ABS} = 1.1065$ MPa. The corresponding relative error

$$err_{REL} = \frac{err_{ABS}}{\sigma_{max}^{FEM}} = 4.79\%, \text{ is also calculated.}$$

8.4.8. *Subsequent DoE analysis*

The final optimisation results based on a DoE/RS analysis can be based on the constructed RS model if the user is satisfied with the accuracy of the approximation and if the error at the optimal solution is acceptable. Alternatively, if this is not the case, a subsequent DoE analysis can be performed to tune the response approximation. This procedure will provide a more accurate optimal solution and can be performed in the following manner.

The found optimal point - as a result of the last undertaken procedure for DoE analysis, RS modelling and optimisation - is used as a criterion to specify of a new reduced design region. In fact, this last optimal point, without being very accurate, provides important information where in design space the precise optimum is. The new reduced design region is a part of the previously considered design space and must contain the associated with that design space optimal point.

Once the new reduced design space is determined, the typical *DoE* analysis, RS modelling procedure, and finally optimisation follow – this time applied using the reduced design space. It is clear, that this strategy will provide a construction of a more accurate response approximation and will guarantee more accurate prediction about the true optimum.

A subsequent analysis is undertaken in this study. The strategy to specify the new reduced design space is to reduce by a quarter the initial one - the one that contains the *DoE_1* optimum (see Figure 8.10).

Design space and the 9 points composite *DoE_1*

New reduced design space and the 9 points for composite *DoE_2*
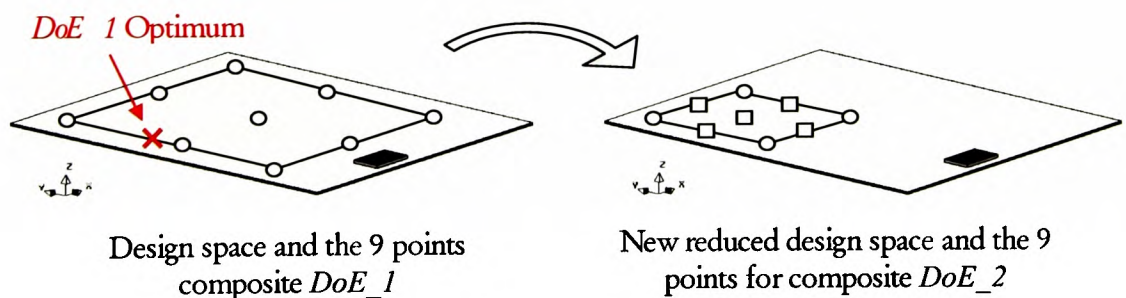
Figure 8.10: Reduction of Initial Design Space for New Subsequent *DoE_2* Procedure and Corresponding Experimental Points

The subsequent *DoE_2* study is also based on a 3-level full composite design. We can take advantage now of the choice of the reduced design space since four of the *DoE_2* points (see Table 8.6, Points 1-4) have been already evaluated by FEM simulations in the previous *DoE_1* analysis. As a result of this, the FEM simulations are undertaken only at the new five (unevaluated) design points (Table 8.6, Points 5-9) to predict the die stress responses of the chip component.

Table 8.6: *DoE_2* Points and PBGA Die Stress Responses

DoE_2 Points	PBGA X-coord. [mm]	PBGA Y-coord. [mm]	Average Stress in PBGA Die [MPa]
DoE_2 Point 1	77.5	137.5	40.98372
DoE_2 Point 2	77.5	77.5	39.27462
DoE_2 Point 3	17.5	137.5	24.68072
DoE_2 Point 4	17.5	77.5	24.28095
DoE_2 Point 5	77.5	107.5	35.47948
DoE_2 Point 6	17.5	107.5	22.23486
DoE_2 Point 7	47.5	137.5	36.68016
DoE_2 Point 8	47.5	77.5	36.42482
DoE_2 Point 9	47.5	107.5	32.37817
DoE_2 Point 10	17.5	97.5	23.10970

In addition to these experimental design points, the optimal design point obtained in the *DoE_1* analysis is also used (Table 8.6, Point 10). The usage of this optimal point in the new DoE analysis helps to improve further the accurate of constructed new RS model, especially in the region of the best design point.

These ten experimental points are used then to obtain the new RS model for the die stress response in the reduced design space.

Similarly to the previous *DoE_1* analysis, a full second order polynomial is used to fit the *DoE_2* response data (Table 8.6). This new approximation (*RS_2*) represents the component die stress only in the region of the reduced design space:

RS_2 Model:

$$\sigma_2\ (X,\ Y) = a0 + a1*X + a2*Y$$
$$+ c12*X*Y + b1*X**2 + b2*Y**2$$

a0	=	56.81505982
a1	=	0.59985292
a2	=	-0.82883109
c12	=	0.00045121
b1	=	-0.00427534
b2	=	0.00380829

(8.8)

The residual analysis for the above RS model shows improved accuracy with all terms in the polynomial being well defined. The results from this residual analysis are reported in Appendix C (see page 243). The G-efficiency for *DoE_2/RS_2* has value similar to the one observed for the *RS_1* model (86.5%). The adjusted R^2-statistic indicates better fit of the second approximation *RS_2* (8.8) to the response data at the experimental points - over 98 % against 95% for the *RS_1* model. The root mean square error for this response model is 0.97 MPa. This value is over 2.5 times smaller than the same error observed for the *RS_1* model. The related coefficient of variation is 3.07 % - over 2 times smaller than the estimated value for the *RS_1* model. The overall sum of the residuals is reduced almost three times. The Analysis of Variance indicates that the *total sum of squares SST* drops by 60 %.

A detailed analysis of the whole residual data provided in Appendix B and Appendix C and shows that the RS model obtained in the subsequent *DoE_2* analysis offers a significantly improved accuracy. Figure 8.11 (left) presents FEM versus *DoE_2* RS approximate stress responses at *DoE_2* experimental points.
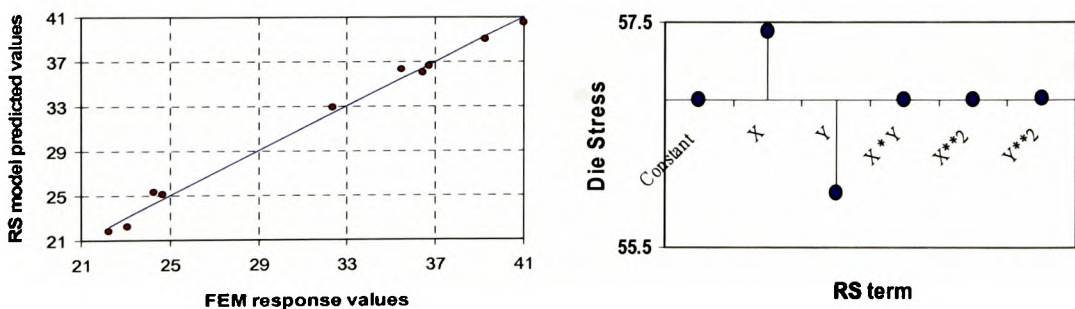
Figure 8.11: *DoE_ 2* Actual (FEM) Response vs. Predicted (*RS_2*) σ_2 (left) and Plot of Design Variable Sensitivities (right)

The sensitivity analysis for *DoE_2* response model (8.8) indicates - in opposite to the *DoE_1* sensitivity analysis - that now the die stress is approximately 38 % more sensitive to the Y-direction placement of the PBGA on board than to its position with respect the X-direction (Figure 8.11, right). The different sensitivity of the chip die stress to the *X* and *Y* location of this component on board, observed for the *DoE_1* and *DoE_2* RS models, can be explained by analysing the whole domain of electronic system sealing, including all components, the board and the rest of the solid parts. In the *DoE_1* analysis, with design space for chip placement equivalent to almost the entire board, it is important to move the chip away from the "dead" flow region behind the front solid obstacle (Obstacle 1, Figure 8.4) for the cooling air, i.e. to improve its *X*-placement. In the *DoE_2* case, the board area is reduced and localized already away from the "dead" area for air flow associated with the whole enclosure. The impact on cooling is now dominated by the solid obstacle specified on top of the board and along the X-direction of the domain (Obstacle 2, Figure 8.4). As a result, in the localized design space (*DoE_2* board area) the Y-direction placement of the chip on board becomes more important, i.e. what is the chip location with respect that solid part above the board

Similarly to the *DoE_1/RS_1* based optimisation process, the optimal procedure in the localized PCB area for the best chip placement is performed using the (8.8) *RS_2* approximation. The optimisation task here has the same objective (8.4) but the design limits (8.5) and (8.6) are modified to $(17.5 \leq X \leq 77.5)$ and $(77.5 \leq Y \leq 137.5)$ respectively to account for the new reduced design space. The continuous and the subsequent discrete optimal design point are identified as a result of the optimisation calculations.

The continuous optimal solution is given in Table 8.7. The *RS_2* model and the optimal design point in terms of chip placement are given in Figure 8.12.

Table 8.7: Optimisation Results based on *DoE_2* and *RS_2* model

DESIGN VARIABLES

Name	Lower Bound	Optimal Value	Upper Bound
X	17.500	17.50000	77.500
Y	77.500	107.7826	137.500

OBJECTIVE FUNCTION

Name	Type	Target Value	Optimal Objective Value
Die Stress	Response	minimise	21.76194
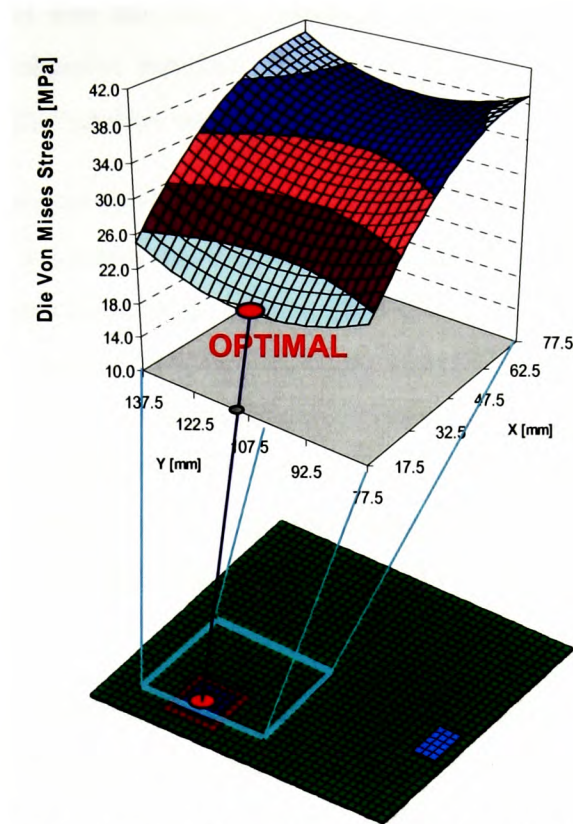
Figure 8.12: PCB Design Space with Mesh and the Final Optimal Design: *RS_2* Model

At the optimal discrete design point (*X* co-ordinate equal to 17.5 mm and *Y* co-ordinates equal to 107.5 mm) the approximate value of the stress in the die from *RS_2* is $\sigma_{die}^{RS} = 21.7622$ MPa. With the same optimal design point, the Finite Element Analysis predicts the exact (based on the multiphysics simulation) response for stress in the silicon die $\sigma_{die}^{FEM} = 22.2349$ MPa. The absolute error at that optimal design point between the exact and the predicted value is $err_{ABS} = 0.4727$ MPa. The corresponding relative error is also

estimated - $err_{REL} = \dfrac{err_{ABS}}{\sigma_{die}^{FEM}} = 2.13\ \%$.

The comparisons between the absolute and the relative errors of the actual and predicted values of the stress in the component die, at the optimal points for *DoE_1* and *DoE_2*, show that these errors for the latter analysis decrease more than twice. This is due to the improved accuracy of the response model. The optimal placement is shifted slightly for the *DoE_2*

optimum if compared with the *DoE_1* optimal design point, which affects and improves further the optimal objective function value. Since the accuracy of the last RS model is accepted, no further *DoE* analysis need to be performed.

To find the best placement for this component is not a straightforward task. The PBGA can not be placed on the board area that is closer to the air flow inlet because, although normally this must be the preferable choice, its temperature will be affected by the heat dissipation of the other component on the board, the HS2 component (Figure 8.4). There are also regions on the board which are not subject to effective cooling due to the geometry features of the electronic assembly (e.g. behind the Obstacle 1, see Figure 8.4). It is difficult to observe the impact of these factors separately. They act simultaneously and the multiphysics modelling is important as it can offer predictions on the complex physics observed in such thermal management problem.

For the analysed above thermal problem, the best cooling of the component (i.e. maximum reduced junction temperature and die stress) can be achieved if its placement is specified: (1) beneath the solid part above the board (Obstacle 2, Figure 8.4) where the air flow speeds up (2) away from the dead flow region behind Obstacle 1) and (3) far from the heat dissipated HS2 component.

8.4.9. *Multiphysics Analysis Results: Optimal Chip Placement*

The multiphysics finite element analysis predicts the temperature distribution across the entire domain and in particular across the circuit board and the associated components. Figure 8.13 (left) presents the temperature contours across the circuit board, the component with a fix placement HS2 and the investigated PBGA (HS3) component. The results illustrate the PCB assembly with the optimal placement of the PBGA component. The average temperature in the PBGA die is 71.47°C, if the component is placed at its optimal location on the board, comparing with 92.96°C for the initial PCB design (i.e. the initial placement of the PBGA component). The magnitude of die temperature is reduced by 23% as a result of the thermal optimisation of the electronic system design.
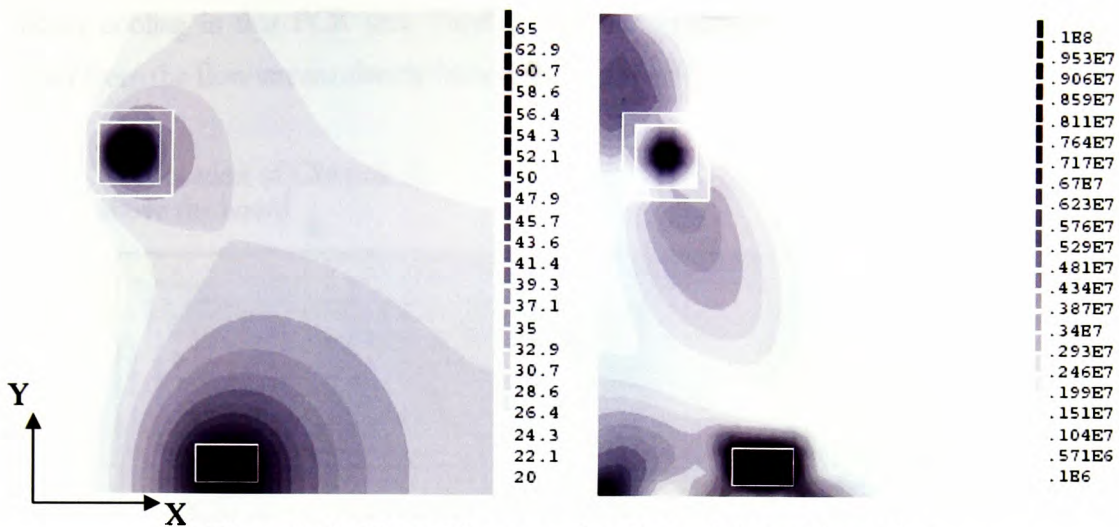
Figure 8.13: PBGA Placement for Optimal Assembly Design:
PCB Top View of Temperature [C] (left) and Von Mises Stress [Pa] (right)
Contours across Board and Components

The temperature profile has impact on the thermally induced stresses inside the PBGA die. The contour plot of the Von Mises stress across the board and components for the optimal design assembly is presented in Figure 8.13 (right). The average Von Mises stress in the PBGA die drops from 47.4420 MPa for the initial PBGA placement to 22.2359 MPa for the optimal PCB design. This is a significant decrease in the stress level - more than a half of the one observed for the initial placement. The decrease of this stress, and respectively the junction temperature of die, was the objective in this optimisation study. The minimised stress in the PBGA die – the objective in this thermal optimisation task - has a positive impact on the reliability of the component and associated electronic assembly.

Figure 8.14 illustrates the air flow profile in the X-Y plane across the entire sealing of the electronic assembly and 2.5 mm above the circuit board. The solid parts within the box change the direction and the magnitude of the cooling air flow. The solid part near the flow inlet (Obstacle 1, see Figure 8.4) affects the flow direction and causes the formation of small air vortices at that location. Also, the magnitude of the flow is significantly small behind this flow restriction, and this has an influence on the degree of cooling in that side of the PCB. The other solid flow restriction (Obstacle 2, Figure 8.4) is near the flow outlets and, because it is located above the PCB, a small gap exists between the board and this solid part. As a result, the air velocity in the slot becomes significantly high with a magnitude which suggests

better cooling in that PCB area. Finally, optimisation results indicate the optimal location away from the flow stream already heated due to the HS2 component heat dissipation.
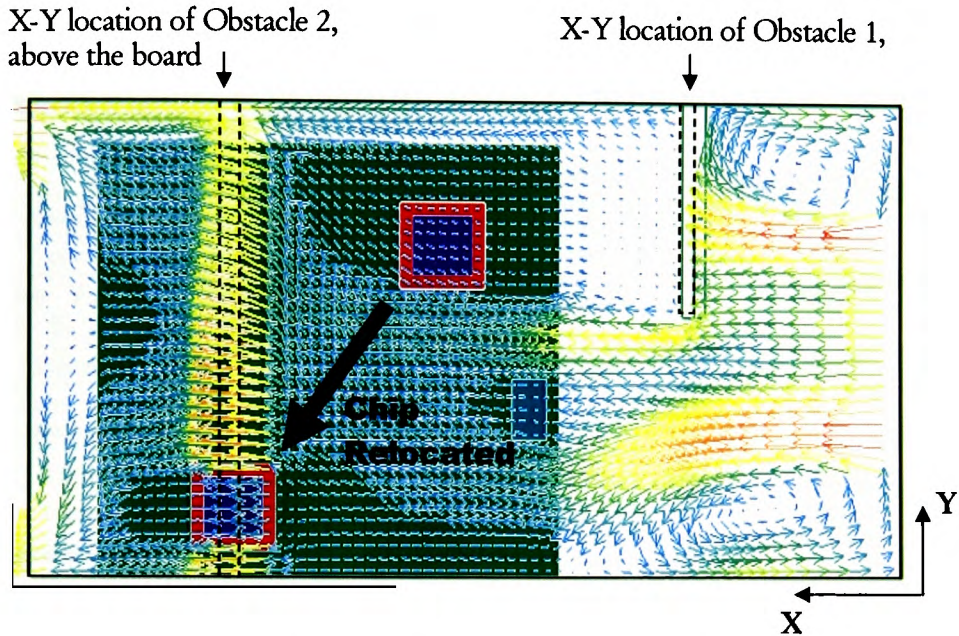
Figure 8.14: PBGA Placement for the Optimal Assembly Design: Resultant Velocity Profile in X-Y Plane across the box and at Height 2.5 mm above PCB

8.4.10. *Thermal Management Problem: Summary*

The results of optimisation study of illustrated thermal management problem are summarised in Table 8.8.

The effect of the air cooling – the profile and the magnitude – should be considered in conjunction with the temperature profile of heat sources in the electronic system and their heat dissipation. With respect to the cooling air the best placement for the PBGA component on board should be right in front of the cooling air inlet. Unfortunately, in this case the heat sources with the fixed placement, HS1 and HS2, will be located near the PBGA component, and the heat - especially from HS2 component - will conduct through both the board and the air to the PBGA. This will reduce the effect of the air cooling, or can make the temperature operating conditions for the PBGA even worse.

The advantage of applying optimisation based on multiphysics analysis of the electronic system is in (1) identifying efficiently and systematically the optimal solution of the design task and (2) obtaining realistic representation of the complex physics associated with the cooling process.

Table 8.8: Optimisation Results Summary

	Initial Design	Optimal Design
HS3 X-coordinate [mm] on PCB :	117.5	17.5
HS3 Y-coordinate [mm] on PCB :	37.5	107.5
Average Von Mises Stress [MPa] in HS3 (BGA) die : OBJECTIVE	47.4420	22.2359
Die Temperature [°C]	92.96	71.47
Air flow temperature at a fixed point above HS3 (PBGA) [°C] • 0.5 mm above PBGA • 1.5 mm above PBGA	 85.0 70.7	 53.5 30.2
Air flow resultant velocity at fixed point above HS3 (PBGA) [ms⁻¹] • 0.5 mm above PBGA • 1.5 mm above PBGA	 0.021 0.059	 0.35 0.65
Improvement in the objective (decrement)		53.13 %
Overall number of multiphysics finite element analyses		16
Total CPU time to perform one multiphysics analysis (Alpha DEC computer with 450 MHz Processor)		3 hours 30 min.

This optimisation study was based on the subsequent application of DoE and Response Surfaces methodology as a strategy to identify the best placement for a component on the printed circuit board. The total number of the multiphysics analyses in this study was 16.

This example shows that the thermal management in electronics requires the consideration of all aspects of the problem and analysis of the complex heating and cooling processes. The multiphysics modelling can provide details on this complex thermal-flow-stress behaviour of the electronic system and can be used to find solutions for effective cooling. The optimal solution cannot be obtained in a straightforward manner – based only on the intuition or by observing only some aspects of the operating conditions. The optimal thermal solution for an electronic system can be identified effectively if the advantage of both the multiphysics modelling and optimisation are combined.

CHAPTER 9

CONCLUSIONS AND FUTURE WORK

9.1. Conclusions

The research presented in this thesis has focused on the integration of

- Design of Experiments statistical methods;

- Response Surface Methodology (RSM) approximation techniques;

- Numerical Optimisation techniques (gradient and non-gradient)

with multiphysics finite element analysis. The application areas have focussed on design optimisation in electronic packaging and microelectronics design. As with most software packages for finite volume or finite element analysis, the simulation tool exploited in this research (*PHYSICA*) did not originally incorporate capabilities for optimisation.

For a robust design analysis both analysis and optimisation are important and play a key role in efficient design.

The design framework uses computational modelling based on Finite Element Methods to predict the physical behaviour of electronic packaging and associated manufacturing processes. Computer analyses can shorten dramatically the required design time at the

development phase of a product and may provide invaluable knowledge and understanding of the impact of design parameters on system response and performance.

Obviously, from a design engineer's point of view a computer analysis should rely on reliable methods and models to provide predictions that have an acceptable degree of accuracy and also represent the interactions between the physical phenomena associated with the process. At present, many engineering applications with a high degree of complexity, such as those observed in electronic packaging, require multi-physics modelling. This requirement was clearly recognized in this research programme and emphasized throughout. The application problems have demonstrated an increasing level of physics, from single-physics towards complex multi-physics analysis. This research has used the *PHYSICA* software as the analysis tool. When it comes to the analysis tool integration with external software modules such as those for numerical optimisation, the access to some parts of the source code and related computational routines (e.g. the analysis output results) are very important.

The other component to multiphysics design optimisation is the optimiser. Optimisation routines combined with statistical methods (Design of Experiments) and approximation techniques (Response Surface Modelling) provide a very robust and efficient design approach. In optimisation modelling, it is very important that the optimisation module also has the capabilities for easy integration to an analysis tool. For this research it was also important to identify, implement and test the performance of optimisation strategies that require as few computational analyses (i.e. different design evaluations) as possible. The optimisation routines in the *VisualDOC* software package were used in this research.

This research has outlined the benefits of using the integrated optimisation-simulation modelling for applications in microelectronics design. This methodology offers advantages that go beyond the scope of performing effective and inexpensive virtual computer analyses - it allows the engineer to run virtually the whole design process where the analysis is just a part in this process.

The thesis has presented the solution procedures for the following design problems:

- In Chapters 5 and 6 a flip-chip package has been designed virtually. The ultimate goal was to improve the reliability of the solder joints and the performance of the whole

package. The design was performed using both the traditional parametric study approach and the proposed optimisation modelling methodology. The benefits of the second approach were demonstrated. Design parameters where calculated that maximised the lifetime of the solder joints.

The optimal properties of no-flow underfills have been obtained simultaneously with the best geometric dimensions of the package. The approximated (DoE/RS) and exact (gradient-based optimisation) numerical approaches have both provided identical results for the observed design parameters. The sensitivity analysis has illustrated the significant impact of underfill properties on reliability of the flip-chip solder interconnects. With respect the geometric design of the flip-chip package, solder stand-off height is the most influential parameter. The optimal design process has proceeded in a fully automated computational procedure with 15 and 20 finite element simulations for DoE/RS and gradient optimisation respectively. The stress in the silicon die was also monitored and kept below a certain limit.

- In Chapter 7 the manufacturing process for lead-free wave soldering in an inert atmosphere has been studied. The design process has identified the optimal flow rates for the injected nitrogen. This design study has provided process specifications that potentially can lead to enormous savings in this soldering process. The solder quality is not compromised because all requirements that guarantee its high standards are observed and taken into account in the optimisation process.

The optimisation of the wave soldering process has relied on analysis results from coupled heat-flow calculations and oxygen concentration predictions. Chapter 7 has demonstrated the increasing complexity of the analysis with existing interactions between the solved quantities. The optimisation process itself utilised the Incremental Refinement approach for Response Surfaces based on multi-quadratic approximations instead of the traditional polynomial fits. The numerical optimisation has adopted a novel and non-gradient based technique (Particle Swarm Optimisation). The advantage of this approach is in the highly increased probability to locate the global optimum solution of the optimisation task. At the same time the computational cost to perform the overall design process is still acceptable as it is

dependent only on the number of analysis used to build and refine the Response Surface.

Optimal parameters of the wave soldering process are important for both the machine manufacturers and the nitrogen suppliers. Any improvement of the process conditions leads to enormous benefits and savings for these companies. As far as the author is aware, there are no published results on optimisation modelling for wave soldering process.

- The last application (Chapter 8) has extended the analysis complexity to a typical multi-physics simulation for coupled heat-flow-stress predictions. The optimal thermal management design – one of the most important problems in microelectronics system assembly has been discussed. The common task for effective cooling and placement of components on printed circuit boards has been observed through an illustrative example. It has provided a good demonstration for the capabilities of the design framework to identify efficiently, quickly and in an entirely automated way the optimal thermal management solution. For problems in microelectronics that require thermal management the multi-physics analysis is very critical and essential to account for the underlying relationships between the process of heat dissipation and transfer, cooling air flow and induced thermal stresses in the active components.

The microelectronics problems discussed in the application part of this thesis are very different and each of them has its own specific characteristics. In each of the analysed design problems, optimal solutions were provided using the integrated optimisation-modelling approach. The aspects of optimality (i.e. the objectives) in these studies have varied from reliability, to cost savings in a wave soldering process, to component placement on printed circuit board.

Numerical optimisation has been demonstrated throughout the thesis as a very powerful tool if combined with reliable finite element analysis for microelectronics design. The *optimisation-modelling* integration can benefit a virtual general design technology tool that, if adopted in the very early design stages of an electronic product development can shorten significantly the real time to market of a product and may provide enormous savings for the

industry by aiding solutions reliable, high performance and cheaper electronic components and products.

9.2. Future Work

From the optimisation point of view further comparisons between different optimisation methods should be undertaken. The performance of new techniques such as the sequential programming with movable limits and some novel non-gradient optimisation algorithms (e.g. differential evolution, simulated annealing, tabu search, etc.) should be investigated. The interest in optimisation performance should focus on aspects for their robustness, efficiency in terms of required function evaluations (respectively FE analysis calls) to obtain the optimum, quality of the optimum (local vs. global) and the degree of difficulty for software implementation and integration to a software analysis package.

Gradient-based optimisation for the applications outlined in this thesis has adopted the finite difference approach for gradient calculations of the optimisation functions. This is a general approach where the nature of the simulated phenomena and the analysed system itself are not taken into account explicitly. It is a suitable strategy for applications in multi-physics simulation-based optimisation. However, there are two major disadvantages of this method. First, finite difference gradient calculations require a significant number of design points to be evaluated and used in the calculations of function gradients at each specific design point. This is a major drawback for simulation-based optimisation where computationally expensive simulations are normally involved. Secondly, the gradients based on finite difference calculations could be poorly estimated because of the accumulated error due to numerical noise, the error of the method, and some mesh effects. As a result, the gradient based-optimisation may fail. As it was discussed in Chapter 1 for some simple physics phenomena there are developed techniques, so called explicit sensitivity analysis, where the function gradients are delivered by explicit calculations from the governing equations of the phenomena. As a part of the future research we intend to investigate further and to develop calculation procedures for explicit calculations of sensitivity of responses to the design parameters.

Response Surface models, different than the classical polynomial approximation, should be examined further (e.g. the Kriging models). Special emphasis should be given to artificial intelligence techniques such as Neural Network. In general, Neural Network Response Surface models tend to require relatively low number of design point evaluations for their construction and are very promising, but detailed investigation of this problem and comparisons with the existing models in future studies will provide better understanding for their performance.

Future work from the computer analysis point of view will address a number of important issues. Efforts will be spend on the development of reliable physics-based models that are capable to represent accurately the response behaviour of an electronic package or system. At the end, the optimisation provides accurate answers for the optimal design as long as the prediction results from the analysis tool are realistic and representative. Novel models for microelectronics analysis have to be implemented to offer more flexibility of the analysis tool with respect the physical behaviour of electronic packages. For example, fatigue models for lead free solder, reliable models for vibration or de-lamination predictions, etc. will widen the research scope and will offer the opportunity to investigate more packaging issues.

A very important problem that needs to be addressed in the future relates to the so called *mesh effect*. When the optimisation involves shape/size design parameters, these parameters are varied during optimisation process, they will change their values and will start to affect the mesh size of finite element model and respectively will have an impact on the analysis predictions. At present the best strategy, followed in this research, is to keep the mesh size ratio within some limits. This reduces the mesh size effect but more investigation, including the theoretical aspect, is required and should be addressed in the future work.

APPENDIX A

Particle Swarm Optimisation (PSO): Algorithm Overview

The Particle Swarm Optimisation [32, 156] is a non-gradient search algorithm based on a natural phenomenon – the social behaviour of a population of individuals in its adaptation to the environment. The knowledge and memory of each individual and the population as a whole are used to search for best regions associated with the environment in this stochastic process. In the PSO, the population is called *swarm*, and an individual is referred to as a *particle*.

A particle of the swarm corresponds to a design point in the design space. In the search process the position of each particle, denoted as X_k^i (particle i of the swarm at iteration k) is updated to a new position X_{k+1}^i using the corresponding for the particle velocity vector v_{k+1}^i.

The algorithm performed is according to the following steps:

1. Start with initial set of particles (design points), i.e. initial swarm. The particles together with the initial velocity vectors can be randomly determined.

$$X_0^i = X^{low} + r_1(X^{up} - X^{low}) \quad \text{and} \quad v_0^i = \frac{X^{low} + r_2(X^{up} - X^{low})}{\Delta t}$$

where X^{low} and X^{up} are the vectors of lower and upper bounds for the design variables, r_1 and r_2 are random numbers between 0 and 1, and Δt is a unit time step.

2. Calculate the velocity vector of each particle using the updating iteration formula

$$v_{k+1}^i = wv_k^i + c_1 r_1 \frac{(p^i - X_k^i)}{\Delta t} + c_2 r_2 \frac{(p^g - X_k^i)}{\Delta t}$$

where p^i denotes the best position so far found by the particle i and p^g is the best position in the swarm as a whole so far in the search [190]. The problem dependent parameters in the above formula are:

- the inertia of the particle, w, and it controls the exploration of the design space with higher values forcing the global search;

- c_1 and c_2 are "confidence" parameters and represent respectively the trust of the particle in itself and in the swarm.

The optimisation study using PSO, presented in Chapter 7, uses the following parameter settings as proposed by Venter [156]: c_1=1.5 ; c_2 =2.5 and w is dynamically updated according to $w_{new}=w_{old} f_w$ where the search starts with w=1.4 and is not allowed to drop bellow 0.35 and f_w=0.975

3. Update the position of each particle

$$X_{k+1}^i = X_k^i + v_{k+1}^i \Delta t$$

4. If converged – exit, otherwise Go to step 2.

APPENDIX B

Residual Analysis and ANOVA for *DoE_1* and Response Surface *RS_1* (Chapter 8)

DoE_1 Results: Efficiency Measures

D-Efficiency	46.2241
A-Efficiency	31.1688
G-Efficiency	90.9718
Scaled average predicted variance (all responses)	0.816497

Approximation Model *DoE_1/RS_1*

Independent response: Average Von Mises Stress in the PBGA die

Number of terms: 6

Term	Coefficient	Std. Error	t-Statistic	P-value
Constant	29.897	4.1043	7.2843	0.005341
X	0.44458	0.088044	5.0495	0.014982
Y	-0.30882	0.088044	-3.5076	0.039267
X*Y	4.64E-06	0.000371	0.012524	0.99079
X**2	-0.00152	0.000524	-2.8972	0.062644
Y**2	0.001567	0.000524	2.9896	0.058147

DoE_1 Summary: Residual Analysis

ANOVA Table

Source	Degrees of Freedom	Sum of square	Mean square	F0	P
Model	5	1165.437	233.0873	32.747	0.008062
Error	3	21.35356	7.117852		
Total	8	1186.79			

Sum of Residuals	2.84E-14
Prediction Error Sum of Squares (PRESS)	237.4602
Root Mean Square Error	2.66793
Mean Response	40.85325
Coefficient of Variation	0.065305
R**2	0.982007
Adjusted R**2	0.95202
R**2 from PRESS	0.799914

Design Point	Predicted Response	Residual	Standardized Residual	Studentized Residual
1	49.56729	1.155436	0.162329	0.982141
2	57.40892	-1.0315	-0.14492	-0.8768
3	24.3804	0.300316	0.042192	0.255273
4	32.28887	-1.88663	-0.26506	-1.60366
5	47.84811	-0.12393	-0.01741	-0.06968
6	22.69464	1.58631	0.222864	0.891876
7	42.43947	-1.45575	-0.20452	-0.81847
8	50.31452	2.91813	0.409973	1.64067
9	40.737	-1.46238	-0.20545	-0.8222

Design Point	R-Studentized Residual	PRESS Residual	H^Diagonal	Cook's D-Statistic
1	0.973563	5.942243	0.805556	0.666034
2	-0.83012	-5.30488	0.805556	0.530819
3	0.210731	1.54448	0.805556	0.044995
4	-3.46556	-9.70264	0.805556	1.775724
5	-0.05694	-0.27885	0.555556	0.001011
6	0.849492	3.569197	0.555556	0.165717
7	-0.75828	-3.27544	0.555556	0.139562
8	4.179454	6.565792	0.555556	0.560792
9	-0.76274	-3.29035	0.555556	0.140835

APPENDIX C

Residual Analysis and ANOVA for *DoE_2* and Response Surface *RS_2* (Chapter 8)

DoE_2 Results: Efficiency Measures

D-Efficiency	44.6765
A-Efficiency	29.6520
G-Efficiency	86.4505
Scaled average predicted variance (all responses)	0.774597

Approximation Model *DoE_2/RS_2*

Independent response: Average Von Mises Stress in the PBGA die

Number of terms: 6

Term	Coefficient	Std. Error	t-Statistic	P-value
Constant	56.815	8.8574	6.4145	0.003036
X	0.59985	0.089759	6.6829	0.002607
Y	-0.82883	0.15914	-5.2082	0.00648
X*Y	0.000451	0.000533	0.84725	0.44457
X**2	-0.00428	0.000746	-5.7296	0.004594
Y**2	0.003808	0.000721	5.2832	0.006157

DoE_2 Summary: Residual Analysis

ANOVA Table

Source	Degrees of Freedom	Sum of square	Mean square	F0	P
Model	5	469.1669	93.83339	100.26	0.000272
Error	4	3.743548	0.935887		
Total	9	472.9105			

Sum of Residuals	-1.07E-14
Prediction Error Sum of Squares (PRESS)	33.1983
Root Mean Square Error	0.967413
Mean Response	31.55272
Coefficient of Variation	0.03066
R**2	0.992084
Adjusted R**2	0.982189
R**2 from PRESS	0.9298

Design Point	Predicted Response	Residual	Standardized Residual	Studentized Residual
1	40.46938	0.514337	0.549572	1.197296
2	38.97414	0.300479	0.321064	0.687868
3	25.12514	-0.44442	-0.47486	-1.02414
4	25.25426	-0.97331	-1.03999	-1.87347
5	36.2943	-0.81482	-0.87064	-1.23277
6	21.76224	0.472623	0.505	0.620592
7	36.64507	0.035094	0.037498	0.053559
8	35.96201	0.462811	0.494515	0.71688
9	32.87607	-0.4979	-0.53201	-0.7533
10	22.16459	0.945111	1.009856	1.209976

Design Point	R-Studentized Residual	PRESS Residual	H^Diagonal	Cook's D-Statistic
1	1.294473	2.608426	0.802817	0.972744
2	0.634414	1.473732	0.79611	0.307918
3	-1.03258	-2.20876	0.798793	0.693998
4	-4.63506	-3.37491	0.711603	1.443404
5	-1.35579	-1.74553	0.533199	0.289315
6	0.565354	0.762642	0.380282	0.039389
7	0.0464	0.076499	0.541247	0.000564
8	0.665025	1.039233	0.554661	0.106679
9	-0.70424	-1.06663	0.533199	0.10803
10	1.316032	1.449753	0.348089	0.130288

PhD RESEARCH

List of Conference Papers and Journal Publications

1. **INTEGRATED COMPUTATIONAL MECHANICS AND OPTIMIZATION FOR DESIGN OF ELECTRONIC COMPONENTS**
   S. Stoyanov, C. Bailey, H. Lu and M. Cross
   Published in "*Optimization in Industry*", Ian C. Parmee and Prabhat Hajela (Eds.)
   Publisher: Springer-Verlag London Ltd. 2002, pp. 57-70 (ISBN: 1-85233-534-3)

2. **OPTIMIZATION TOOLS FOR FLIP-CHIP DESIGN**
   S. Stoyanov, C. Bailey, and H. Lu
   Proceedings of InterPACK 2001 (CD proceedings, paper No. IPACK2001-15741): The Pacific Rim/ASME International Electronic Packaging Technical Conference and Exhibition
   July 8-13, 2001, Kauai, Hawaii, USA

3. **RELIABILITY ANALYSIS OF NO-FLOW UNDERFILL MATERIALS**
   H. Lu, K. C. Hung, S. Stoyanov, C. Bailey and Y. C. Chan
   Proceedings of InterPACK 2001(CD proceedings, paper No. IPACK2001-15814): The Pacific Rim/ASME International Electronic Packaging Technical Conference and Exhibition
   July 8-13, 2001, Kauai, Hawaii, USA

4. **A MODELLING AND EXPERIMENTAL ANALYSIS OF THE NO-FLOW UNDERFILL PROCESS FOR FLIP-CHIP ASSEMBLY**
   H. Lu, S. Stoyanov, C. Bailey, K. C. Hung and Y. C. Chan
   Proceedings of the Fourth International Symposium on Electronic Packaging Technology
   August 8-11, 2001, Beijing, China, pp. 338-343. * Best Paper Award

5. **INTEGRATING COMPUTATIONAL MECHANICS AND NUMERICAL OPTIMIZATION FOR THE DESIGN OF MATERIAL PROPERTIES IN ELECTRONIC PACKAGES**
   S. Stoyanov, C. Bailey, and M. Cross
   The 2002 TMS Proceedings of Computational Modeling of Materials, Minerals and Metals Processing
   TMS Annual Meeting & Exhibition, February 17–21, 2002, Seattle, Washington, USA, pp. 551-561 (ISBN 0-87339-513-1)

6. **SOLDER JOINT RELIABILITY OPTIMIZATION**
   S. Stoyanov, C. Bailey, H. Lu and M. Cross
   APACK 2001 Proceedings of International Conference on Advances in Packaging, December 5-7, 2001, Singapore, pp. 178-186 (ISBN 981-04-4638-1)

7. OPTIMIZATION STRATEGIES IN RELIABILITY FOR ELECTRONIC PACKAGES
S. Stoyanov, C. Bailey and M. Cross
Emerald Journals, *Soldering & Surface Mount Technology*, Vol. 14, No. 1, 2002, pp. 49-58

8. RESPONSE SURFACE MODELING AND OPTIMIZATION FOR RELIABLE ELECTRONIC PACKAGING
S. Stoyanov and C. Bailey
The Fifth International IEEE Symposium on High Density Packaging and Component Failure Analysis in Electronics Manufacturing, *HDP 2002*
June 30 - July 3, 2002, Shanghai, China, pp. 49-57
(ISBN: 0-7803-9822-X; ISBN: 0-7803-9823-8) *Keynote Paper

9. NO-FLOW UNDERFILL FLIP CHIP ASSEMBLY – AN EXPERIMENTAL AND MODELING ANALYSIS
H. Lu, K. C. Hung, S. Stoyanov, C. Bailey and Y. C. Chan
Microelectronics Reliability, Vol. 42, 2002, pp. 1205-1212

10. OPTIMISATION AND FINITE ELEMENT ANALYSIS FOR RELIABLE ELECTRONIC PACKAGING
S. Stoyanov and C. Bailey
The 4th International Conference on Thermal & Mechanical Simulation and Experiments in Micro-Electronics and Micro-Systems, *EuroSIME 2003*
Aix-en-Provence, France, March 30 - April 2, 2003, pp.391-398
(ISBN (book): 0-7803-7054-6) (ISBN (CD-ROM): 0-7803-7056-2) *Keynote Paper

11. OPTIMISATION MODELING FOR ELECTRONIC PRODUCT DESIGN
S. Stoyanov and C. Bailey
Proceedings of PREP2003: Postgraduate Research Conference in Electronics, Photonics, Communications and Software
14 April - 16 April, 2003, Exeter University, Exeter, UK, pp. 131-132.

BIBLIOGRAPHY

[1] Dantzig, G. B. (1963) **Linear Programming and Extensions**, Princeton University Press.

[2] Kuhn, H. W. and Tucker, A. W. (1951) "Nonlinear Programming", *Proc. 2nd Berkeley Symp.* (Berkeley, Calif.: University of California Press).

[3] Bellman, R. (1957) **Dynamic Programming**, Princeton University Press, Princeton, New Jersey.

[4] Pontryagin, L. S., Boltyansky, V. G., Gamkrelidze, R. V. and Mishchenko, E. F. (1962) **The Mathematical Theory of Optimal Processes**, New York: Wiley.

[5] Hooke, R. and Jeeves, T. A., (1961) "Direct Search Solution of Numerical and Statistical Problems", J. Assoc. Computer Math, Vol. 8, pp. 212-229.

[6] Powell, M. J. D. (1964) "An Efficient Method for Finding the Minimum of a Function of Several Variables without Calculating Derivatives", Comp. J. Vol. 7, No. 4, pp. 303-307.

[7] Himmelblau, D. M. (1972) **Applied Nonlinear Programming**, McGraw-Hill, New York.

[8] Gomory, R. (1958) "Outline of an Algorithm for Integer Solutions to Linear Programs", Bulletin of the American Mathematical Society, No. 64, pp. 275-278.

[9] Balas, E. (1965) "An Additive Algorithm for Solving Linear Programs with Zero-One Variables", Operations Research 13, pp. 517-546.

[10] Zoutendijk, K. G., (1960) **Method of Feasible Directions**, Elsevier, Amsterdam, Netherlands.

[11] Fletcher, R. and Reeves, C. M. (1964) "Function Minimization by Conjugate Gradients", Computer Journal 7, 148–154.

[12] Davidon, W. C. (1959) "Variable Method for Minimization", Argon National Laboratory, ANL-5990 Rev., University of Chicago.

[13] Fletcher, R. and Powell, M. J. D. (1963) "A Rapidly Convergent Descent Method for Minimization", Comp. J. 6, pp. 163-168.

[14] Fletcher, R. (1970) "A New Approach to Variable Metric Algorithms", Comp. J., Vol. 13, pp. 317-322.

[15] Shanno, D. F. (1970) "Conditioning of Quasi-Newton Methods for Function Minimization", Math. Comput. , Vol. 24, pp. 647-656.

[16] Broyden, C. G. (1970) "The Convergence of a Class of Double Rank Minimization Algorithms", parts 1 and II, J. Inst. Math. Appl., Vol. 6, pp. 76-90, pp. 222-231.

[17] Broyden, C. G. (1971) "The Convergence of an Algorithm for Solving Sparse Nonlinear Systems", Math. Comp., Vol. 25, pp. 285-294.

[18] Goldfarb, D. (1970) "A Family of Variable-metric Methods Derived by Variational Means", Math. Comput, Vol. 24, pp. 23-26.

[19] Schmit, L. A. (1960) "Structural Design by Systematic Synthesis", *Proceedings of 2nd Conference on Electronic Computation, ASCE*, New York, pp. 105-122.

[20] Schmit, L.A. (1981) "Structural Synthesis Its Genesis and Development", AIAA J., Vol. 10, No. 10, pp. 1249 – 1263, October 1981.

[21] Dorn, W. S. and Greenberg, H. J. (1957) "Linear Programming and Plastic Linear Analysis of Structures", Q. Appl. Math., Vol. 15, No. 2, pp. 155-167.

[22] Livesley, R. K. (1959) "Optimal Design in Structural Frames for Alternative Systems of Loading", Civil Engineering (London), pp. 737-740.

[23] Moses, F. (1964) "Optimum Structural Design using Linear Programming", *Proceedings ASCE*, Vol. 90, ST6, pp. 89-104.

[24] Rosen, J. B. (1960) "The Gradient Projection Method for Non-Linear Programming, Part I, Linear Constraints", SIAM J. Appl. Math., Vol. 8, pp. 181-217.

[25] Vanderplaats, G. N. and Moes, F. (1973) "Structural Optimisation by Methods of Feasible Direction", J. Computers Struct., Vol. 3, pp. 739-755.

[26] Vanderplaats, G. N. (1999) **Numerical Optimisation Techniques for Engineering Design: with Applications**, VR&D, Colorado Springs.

[27] Haug, E. J., Choi, K. K. and Komkov, V. (1986) **Design Sensitivity Analysis of Structural Systems**, Academic Press, 1986.

[28] Haftka, R. (1981) "Techniques for Thermal Sensitivity Analysis", International Journal for Numerical Methods in Engineering, Vol. 17, pp. 71-80.

[29] Mukherjee, S. and Chandra, A. (1989) "A Boundary Element Formulation for Design Sensitivities in Materially Nonlinear Problems", Acta Mechanica , No.78, pp. 243-253

[30] Tortorelli, D. and Haber, R. (1989) "First-order Design Sensitivities for Transient Conduction Problems by an Adjoint Method", International Journal for Numerical Methods in Engineering, Vol. 28, pp. 733-752

[31] Hajela, P. (1990) "Genetic Search – An Approach to the Nonconvex Optimization Problem", AIAA Journal, Vol. 26, No. 7, pp. 1205-1210.

[32] Eberhart, R. and Kennedy, J. (1995) "Particle Swarm Optimisation", *Proceedings of the 1995 IEEE International Conference on Neural Networks*, Vol. 4, pp. 1942-1948.

[33] Price, K. and Storn, R. (1997) "Differential Evolution", Dr. Dobb's Journal, Vol. 22, No.4, pp. 18-24.

[34] Lau, H. J. (1996) **Flip Chip Technologies**, *McGraw-Hill*, New York.

[35] Dai, X. and Ho, P. S. (1997) "Thermo-Mechanical Deformation of Underfilled Flip-Chip Packaging", IEEE/CPMT International Electronics Manufacturing Technology Symposium, pp. 326-333.

[36] Peterson, D. W., Sweet, J. N., Burchett, S. N. and Hsia, A. (1997) "Stresses from Flip-Chip Assembly and Underfill; Measurements with the ATC4.1 Assembly Test Chip and Analysis by Finite Element Method", *IEEE Proceedings of 1997 Electronic Components and Technology Conference*, pp. 134-143.

[37] Ronald J. Ross Jr. and Liang-chi Wen, (1994) "Crack Propagation in Solder Joints during Thermal-Mechanical Cycling", Journal of Electronic Packaging, Vol. 116, pp. 69-75.

[38] Kitano, M. and Honda, M. (1997) "Shape Prediction of Solder Bump Joint by Surface Tension Analysis and Fatigue Strength Evaluation", *EEP Advances in Electronic Packaging, ASME* 1997, Vol. 19, No. 2, pp. 1407-1412.

[39] Darbha, K., Okura, J. H. and S., Dasgupta, A. (1999) "Thermo-Mechanical Durability Analysis of Flip Chip Solder Interconnects: Part 1 – without Underfill", Journal of Electronic Packaging, Vol. 121, pp. 231-236.

[40] Darbha, K., Okura, J. H., Shetty, S., Dasgupta, A. at al (1999) "Thermo-Mechanical Durability Analysis of Flip Chip Solder Interconnects: Part 2 – with Underfill", Journal of Electronic Packaging, Vol. 121, pp. 237-241.

[41] Dai, X. and Ho, P. S. (1997) "Thermo-Mechanical Deformation of Underfilled Flip-Chip Packaging", 1997 IEEE/CPMT Int'l Electronics Manufacturing Technology Symposium, pp. 326-333.

[42] Sarvar, F and Conway, P. P. "Effective Transient Processes Modelling of the Reflow Soldering", available from http://www.mayahtt.com/expertise/papres/tmg-con1.pdf

[43] Madenci, E., Shkarayev, S. and Mahajan, R. (1998) "Potential Failure Sites in a Flip-Chip Package with and without Underfill", Transaction of the ASME, Vol. 120, pp. 336-341.

[44] Syed, A. R. (1997) "ACES of Finite Element and Life Prediction Models for Solder Joint Reliability", in Design and Reliability of Solders and Solder Interconnects, The Minerals, Metals and Materials Society, 1997, pp. 347-355.

[45] Darveaux, R., Banerji, K., Mawer, A. and Dody, G. (1995) "Reliability of Plastic Ball Grid Array Assembly", in Ball Grid Array Technology, McGraw-Hill, New York, pp. 379-442.

[46] Lu, H., Bailey, C. and Cross, M. (2000) "Reliability Analysis of Flip Chip Designs via Computer Simulation", Journal of Electronic Packaging, Vol. 122, pp. 214-219.

[47] Lu, H. and Bailey, C., (2000) "Material Properties, Geometry and their Effect on the Fatigue Life of Two Flip-Chip Models", Int. Journal Electronic Packaging, pp. 12-19.

[48] Popelar, S. F., (1998) "A Parametric Study of Flip Chip Reliability Based on Solder Fatigue Modeling", 1998 International Electronics Manufacturing Technology Symposium, pp. 299-307.

[49] Michaelides, S. and Sitaraman, S. K. (1998), "Effect of Material and Geometry Parameters on the Thermo-Mechanical Reliability of Flip-Chip Assemblies", Proceedings of InterSociety Conference on Thermal Phenomena, pp. 193-200.

[50] Li, R. S. (1998) "Optimization of Thermal Via Design Parameters based on an Analytical Thermal Resistance Model", 1998 InterSociety Conference on Thermal Phenomena, pp. 475-480.

[51] Le Gall, C. A., Qu, J. and McDowell, D. L. (1997) "Influence of Die Size on the Magnitude of Thermo-mechanical Stresses in Flip Chips Directly Attached to Printed Wiring Board", ASME Advances in Electronic Packaging, EEP-Vol. 19, No. 2, pp. 1663-1670.

[52] Yegnasubramananian, S. et al "Flip-Chip-on-Board (FCOB) Assembly and Reliability", Lucent Technologies Document.

[53] Chambers, B., Tien-Yu, T. L. and Blood, W. (1998) "Steady State and Transient Thermal Analysis of Chip Scale Packages", Proceedings of 1998 InterSociety Conference on Thermal Phenomena, pp. 68-75.

[54] Primavera, A. "Influence of PCB Parameters on Chip Scale Package Assembly and Reliability", Surface Mount Technology Laboratory, Universal Instruments Corp., Binghampton, New York.

[55] Gektin, V., Bar-Cohren, A. and Ames, J. (1997) "Coffin-Manson Fatigue Model of Underfilled Flip-Chips", IEEE Transactions on Components, Packaging and Manufacturing Technology, Part A, Vol. 20, No. 3, pp. 317-326.

[56] Gektin, V., Bar-Cohren, A. and Witzman, S. (1997) "Coffin-Manson Fatigue Analysis of Underfilled DCA", ASME 1997, Advances in Electronic Packaging, Vol. 19-2, pp. 1655-1661.

[57] Rinne, G. A. (1997) "Solder Bumping Methods for Flip-Chip Packaging", *1997 IEEE Electronic Components and Technology Conference*, pp. 240-247.

[58] Huang, Y., Teo, K.H., Chua, K.L., Yang, M. and Ferng, W. (1999) "The Effect of Underfill on the Pressure Cooker Test Performance of Flip Chip on Board Assembly", EEP-Vol. 26-2, ASME 1999 Advances in Electronic Packaging, Volume 2, pp. 1121-1127.

[59] Miller, V. and Kanciak, H. (1997) "Impact of the Manufacturing Process on Component Package Design", EEP-Vol. 19-1, ASME 1997 Advances in Electronic Packaging, Volume 1, pp. 139-145.

[60] Boswell, D. and Wickman, M. (1992) **Surface Mount Guidelines for Process Control, Quality and Reliability**, McGraw-Hill Book Company Europe, Berkshire, England.

[61] Rahn, A. (1993) **The Basics of Soldering**, John Wiley & Sons, Inc., New York.

[62] Pang, J., Low, T. and Chong, D. (2001) "Thermal Shock versus Thermal Cycling Simulations for Flip Chip Solder Joint Reliability Assessment", *APACK 2001 Conference on Advances in Packaging, Singapore*, pp. 187-193.

[63] Akay, H. U., Bilgic, A., Zhang, H. and Paydar, N. (1997) "Combined Heat Transfer and Thermal Stress Analysis for Fatigue Life Prediction of Solder Joints of Resistors", in *Design and Reliability of Solders and Solder Interconnects, The Minerals, Metals and Materials Society*, 1997, pp. 187-195.

[64] Aoki, H. *et al* (1997) "Eutectic Solder Flip Chip Technology – Bumping and Assembly Process Development for CSP/BGA", 1997 *Electronic Components and Technology Conference*, pp. 325-331.

[65] Qian, Z. and Liu, S. (1999) "On the Life Prediction and Accelerated Testing of Solder Joints", The International Journal of Microcircuits and Electronic Packaging, Vol. 22, No. 4, Fourth Quarter 1999, pp. 288-304.

[66] Yuen, M. and Fan, H. (2003) "Delamination of Electronic Packages", *4-th International Conference on Thermal and Mechanical Simulation and Experiments in Micro-Electronics and Micro-Systems*, EuroSIME 2003, pp. 15-21.

[67] Chung, P., Chan, S., Yuen, M. and Lam, D. (2000) "An Energy Based Failure Criterion for Interfacial Delamination in IC Packaging", Inter. Symposium Elect. Material Packaging, pp. 201-205.

[68] Liu, C. J. *et al* (2003) "Prediction of Interfacial Delamination in Stacked IC Structures using Combined Experimental and Simulation Methods", *4-th International Conference on Thermal and Mechanical Simulation and Experiments in Micro-Electronics and Micro-Systems*, EuroSIME 2003, pp. 337-343.

[69] Cimtalay, S., Peak, R. S. and Fulton, R. E (1996) "Optimization of Solder Joint Fatigue Life using Product Model-based Analysis Models", ASME Intl. Mech. Engr. Congress and Exposition, EEP-Vol. 18, pp. 47-53.

[70] Mahajan, R. L. (1997) "Process Modeling, Optimization and Control in Electronics Manufacturing", ASME Advances in Electronic Packaging, EEP-Vol 19, No. 1, pp.127-135.

[71] Zahn, B. A. (1998) "Steady State Thermal Characterization and Junction Temperature Estimation of Multi-Chip Module Packages using the Response Surface Method", *1998 InterSociety Conference on Thermal Phenomena*, pp. 76-81.

[72] Zhang, L., Subbarayan, G., Hunter, B. C. and Rose, C. (2001) "Response Surface Models for Efficient, Modular Estimation of Solder Joint Reliability in Area Array Packages", *APACK 2001 Conference on Advances in Packaging, Singapore*, pp. 234-244.

[73] Deshpande, A. M., Subbarayan, G., and Mahajan, R. L. (1997) "Maximizing Solder Joint Reliability through Optimal Shape Design", Journal of Electronic Packaging, Vol. 119, pp. 149-155.

[74] Li, Y, Mahajan, R. L. and Nikmanesh, N. (1996) "Fine Pitch Stencil Printing Process Modeling and Optimization", Journal of Electronic Packaging, March 1996, Vol. 118, pp. 1-6.

[75] Variyam M. N. and Sitaraman, S. K. (1999) "Thermo-mechanical Design of Microelectronic Packages through Design-of-simulations Methodology", The International Journal of Microcircuits and Electronic Packaging, Vol. 22, No 4, pp. 353-362

[76] Simsek, A. and Reichl, H. (1998) "Evaluation and Optimization of MCM-BGA Packages", IEEE Topical Meet. Electrical Perf. Electronic Packaging, pp. 132-135

[77] Craig, K. J., J. de Kock, D. and Gauche, P. (1999) "Minimization of Heat Sink Mass using CFD and Mathematical Optimization", Journal of Electronic Packaging, Vol. 121, pp.143-147.

[78] Snyman, J.A. and Stander N. (1994) "A New Successive Approximation Method for Optimum Structural Design", AIAA Journal, Vol. 32, pp. 1310-1315.

[79] Iyengar, M. and Bar-Coohen, A. (1998) "Least-Material Optimization of Vertical Pin-Fin, Plate-Fin, and Triangular-Fin Heat Sinks in Natural Convective heat Transfer", *Proceedings of 1998 InterSociety Conference on Thermal Phenomena*, pp. 295-302.

[80] Wu, X., Li, F., Tang, K., Yeh, C. and Wyatt, K. (1998) "Solder Joint Design Optimization for Fine Pitch Component Applications", *Proceedings of 1998 InterSociety Conference on Thermal Phenomena*, pp. 236-240.

[81] Subbarayan, G. (1996) "A Procedure for Automated Shape and Life Prediction in Flip-Chip and BGA Solder Joints", Journal of Electronic Packaging, Vol. 118, pp. 127-133.

[82] Corbin, J. S. (1993) "Finite Element Analysis for Solder Ball Connect (SBC) Structural Design Optimization", IBM J. Res. Develop., Vol. 37, No. 5, pp. 585-596.

[83] Zhang, G. Q., Bisschop, J. and Maessen, P. (2001) "Virtual Thermo-Mechanical Prototyping of Microelectronics Products – Towards Optimized Designing in Reliability", Journal of Advancing in Microelectronics, pp. 9-12.

[84] D. van Driel, W., Zhang, G. Q., Janssen, J. H. J. (2003) "Response Surface Modeling for Nonlinear Packaging Stresses", ASME Journal of Electronic Packaging, 125(4), 490-497.

[85] Zhang, G. Q., Maessen, P., Bisschop, J., Janssen, J., Kuper, F. and Ernst, L. (2001) "Virtual Thermo-Mechanical prototyping of Microelectronics – the Challenges for Mechanics Professionals", *Proceedings of EuroSIME* 2001, pp. 21-24.

[86] Fan, X. and Zhang, G. (2002) "Multi-Physics Modeling in Virtual Prototyping of Electronic Packages – Combined Thermal, Thermo-Mechanical and Vapor Modeling", *3rd Int. Conf. on Benefiting from Thermal and Mechanical Simulation in (Micro-) Electronics, EuroSIME* 2002, pp. 29-37.

[87] Vandevelde, B., Beyne, E., Zhang, G. and Caers, J. (2001) "Parameterised Modelling of Thermo-Mechanical Reliability for a 5x4 CSP", *Proceedings of EuroSIME* 2001, pp. 87-94.

[88] Chang, C. and Lin, W. (2001) "Robust Multiple Criteria Optimization of Thermally Enhanced PQFP", *Proceedings of EuroSIME 2001*, pp. 123-130.

[89] Fish, J. and Belsky, V. (1995) "Multi-grid Method for Periodic Heterogeneous Media, Part 2: Multiscale Modeling and Quality Control in Multidimensional Case", Comput. Meth. Appl. Mech. Eng. Vol. 126, pp. 17–38.

[90] Yu, Q. and Fish, J. (2002) "Multiscale Asymptotic Homogenization for Multiphysics Problems with Multiple Spatial and Temporal Scales: a Coupled Thermo-viscoelastic Example Problem", International Journal of Solids and Structures, Volume 39, Issue 26, December 2002, pp. 6429-6452.

[91] Cross, M. (1996) "Modelling of Industrial Multi-Physics Processes - A Key Role for Computational Mechanics", IMA Journal of Mathematics for Business and Industry, Vol. 7, pp. 3-21.

[92] Bailey, C. and Boettinger, W. J., (1999) "Modelling the Fillet Lifting Defect", EEP Advances in Electronic Packaging, Vol. 26, No.1 (ASME 1999), pp. 405-412.

[93] Wheeler, D. and Bailey, C. (1999) "Modelling the Melting and Solidification of Solder Material", EEP Advances in Electronic Packaging, Vol. 26, No.1 (ASME 1999), pp. 397-404.

[94] Bailey, C., Wheeler, D. and Cross, M. (1999) "An Integrated Modeling Approach to Solder Joint Formation", IEEE Transactions on Components and Packaging Technology, Vol. 22, No. 4, pp. 497-502.

[95] Wheeler, D., Bailey, C. and Cross, M. (1998) "A Computational Modelling Framework to Predict Macroscopic Phenomena in Solder Joint Formation", Journal Of Engineering Manufacture, Proceedings of the Institution of Mechanical Engineers, Vol. 212, Part B, pp. 141-157.

[96] Bailey, C., Lu, H. and Wheeler, D. (2002) "Computational Modeling Techniques for Reliability of Electronic Components on Printed Circuit Boards", Applied Numerical Mathematics, Volume 40, Issues 1-2, January 2002, pp. 101-117.

[97] Parry, J., Bailey, C. and Aldham, C. (2000) "Multiphysics Modelling for Electronics Design", *ITHERM 2000*, Las Vegas, pp. 86-93.

[98] Cross, M. *et al* (1998) "Multiphysics Modelling – A Vital Component of Virtual Prototyping", *AGARD SMP Meeting on Virtual Manufacturing*, AGARD-R-821.

[99] Cross, M., Chow, P., Bailey, C., Croft, N., Ewer, J., Leggett, P., McManus, K., Pericleous, K. and Patel, M. (1996) "PHYSICA – A Software Environment for the Modelling of Multi-Physics Phenomena", ZAMM - Journal of Applied Mathematics and Mechanics, Vol. 76, pp. 105-108.

[100] Bailey, C., Taylor, G. A., Bounds, S. M., Moran, G. and Cross, M. (1997) "PHYSICA: A Multi-physics Framework and its Application to Casting Simulation", Computational Fluid Dynamics in Mineral & Metal Processing and Power Generation, Eds. M.P. Schwarz, M.R. Davidson and A.K. Easton, pp. 419-425.

[101] Croft, N., Pericleous, K. and Cross, M. (1995) "PHYSICA: A multiphysics environment for complex flow processes", Numerical Methods in Laminar and Turbulent Flow, Vol. 9, Part 2, pp. 1269-1280.

[102] Bailey, C., Cross, M. and Pericleous, K. (1998) "A Finite-Volume Computational Mechanics Framework for Multi-Physics Coupled Fluid-Stress Problems", Proceedings of the 27th Israel Conference on Mechanical Engineering, May 19-20, 1998, Haifa, Israel, pp. 136-138.

[103] Croft, N. (1998) **Unstructured Mesh - Finite Volume Algorithms for Swirling, Turbulent, Reacting Flows**, Ph.D. Thesis, University of Greenwich, London, UK.

[104] Bailey, C., Taylor, G. A., Cross, M. and Chow, P. (1999) "Discretisation Procedures for Multi-physics Phenomena", Journal of Computational and Applied Mathematics, Vol. 103, Issue 1, pp. 3-17.

[105] Taylor, G. A. (1996) **A Vertex-based Discretisation Scheme Applied to Material Non-linearity within a Multi-physics Finite Volume Framework**, Ph.D. Thesis, University of Greenwich, London, UK.

[106] Taylor, G. A., Bailey, C. and Cross, M. (1995), "Solution of the Elastic/Visco-plastic Constitutive Equations: A Finite Volume Approach", Appl. Math. Modelling, Vol. 19, pp. 746-760.

[107] Fallah, N. A., Bailey, C., Cross M. and Taylor, G. A. (2000) "Comparison of Finite Element and Finite Volume Methods Application in Geometrically Nonlinear Stress Analysis", Applied Mathematical Modelling, Vol. 24, No. 7, pp. 439-455.

[108] Slone, A. K., Bailey, C. and Cross, M (2003) "Dynamic Solid Mechanics using Finite Volume Methods", Applied Mathematical Modelling, Vol. 27, Issue 2, pp. 69-87.

[109] Bailey, C. and Cross, M. (1995) "A Finite Element Procedure to Solve Elastic Solid Mechanics Problems in Three Dimensions on an Unstructured Mesh", International Journal for Numerical Methods in Engineering, Vol. 38, pp. 1757-1776.

[110] Bailey, C., Wheeler, D. and Cross, M. (1997) "Multiphysics Simulations for Solder Joint Formation", EEP-Vol. 19-2, Advances in Electronic Packaging – 1997, Vol. 2 (ASME 1997), pp. 1413-1420.

[111] Parry, J., Marooney, C., Warner, M., Bailey, C. and Pericleous, K. (2002) "An Integrated Approach to Flow, Thermal and Mechanical Modeling of Electronics Devices", *2002 Inter Society Conference on Thermal Phenomena*, pp. 862-868.

[112] Schäfer, M. and Teschauer, I. (2001) "Numerical Simulation of Coupled Fluid-Solid Problems", Computer Methods in Applied Mechanics and Engineering, Vol. 190, No. 28, pp. 3645-3667.

[113] Davey, K. and Rodriguez, N. J. (2002) "Solidification Modelling with a Control Volume Method on Domains Subjected to Viscoplastic Deformation", Applied Mathematical Modelling, Vol. 26, No. 3, pp. 421-447.

[114] Athavale, M., Yang, H. Q. and Przekwas, A. J. (1999) "Coupled Fluid-Thermal-Structural Simulations in Microvalves And Microchannels", *Second International Conference on Modeling and Simulations of Microsystems* (MSM-99), San Juan, Puerto Rico, USA, April, pp. 19-21.

[115] Schafer, M., Teschauer, I., Kadinski, L. and Selder, M. (2002) "Numerical Approach for the Solution of Coupled Fluid-Solid and Thermal Stress Problems in Crystal Growth Processes", Computational Materials Science, Vol. 24, pp. 409-419.

[116] Bailey, C. *at al* (1996) "Multiphysics Modelling of the Metals Casting Processes", Proc. R. Soc. Lond A, Vol. 452, pp. 459-486.

[117] Taylor, G. A., Hughes, M., Strusevich, N. and Pericleous, K. (2002) "Finite Volume Methods Applied to the Computational Modelling of Welding Phenomena", Applied Mathematical Modelling, Vol. 26, pp. 309-320.

[118] PHYSICA, User Guide and Documentation, Multi-Physics Software (MPS) Ltd., London, UK, http://www.multiphysics.com.

[119] FEMGV, Femsys Ltd., http://www.femsys.co.uk/

[120] Hrennikoff, A. (1941) "Solution of Problems in Elasticity by the Frame Work Method", Journal of Applied Mathematics, Vol. 8, No. 4, pp. 169-175.

[121] McHenry, D. (1943) "A Lattice Analogy for the Solution of Plane Stress Problems", Journal of Institution of Civil Engineers, Vol. 21, pp. 59-82.

[122] Turner, M., Clough, R., Martin, H. and Topp, L. (1956) "Stiffness and Deflection Analysis of Complex Structures", Journal of Aeronautical Sciences, Vol. 23, No. 9, pp. 805-824.

[123] Clough, R. (1960) "The Finite Element Method in Plane Stress Analysis", *Proceedings of American Society of Civil Engineers, 2-nd Conference on Electronic Computation*, Pittsburgh, pp. 345-378.

[124] Martin, H. (1961) "Plane Elasticity Problems and the Direct Stiffness Method", The Trend in Engineering, Vol. 13, pp. 5-19.

[125] Gallagher, R., Padlog, J. and Bijlaard, P. (1962) "Stress Analysis of Heated Complex Shapes", Journal of the American Rocket Society, Vol. 32, pp. 700-707.

[126] Melosh, R. (1963) "Structural Analysis of Solids", Journal of the Structural Division, Proceedings of the American Society of Civil Engineers, pp. 205-223.

[127] Zienkiewicz, O., Watson, M. and King, I. (1968) "A Numerical Method of Visco-Elastic Stress Analysis", International Journal of Mechanical Sciences, Vol. 10, pp. 807-827.

[128] Zienkiewicz, O. C. and Taylor, R. L. (1991) **The Finite Element Method**, Vols 1 and 2, 4-th edition, McGraw-Hill, New York.

[129] Zienkiewicz, O. C. (1992), "Computational mechanics today", Int. J of Num. Methods Eng., No. 34, pp. 9-33.

[130] Patankar, S. V. (1980) **Numerical Heat Transfer and Fluid Flow**, published Hemisphere, Washington DC.

[131] Versteeg, H. K. and Malalasekera, W. (1995) **An Introduction to Computational Fluid Dynamics: The Finite Volume Method**, Longman Group Ltd.

[132] Dow, J. O., Jones, M. S. and Harwood, S. A. (1990) "A New Approach to Boundary Modelling for Finite Difference Applications in Solid Mechanics", Int. J. numerical Methods in Engineering, Vol. 30, pp. 99-113.

[133] Rhie, C. M and Chow, W. L. "Numerical Study of the Turbulent Flow Past an Airfoil with Trailing Edge Separation", AIAA Journal, 21(11) (1983), pp. 1525-1532.

[134] Langstangen, H. P. (1999) **Computational Partial Differential Equations**, Springer-Verlag Berlin Heidelberg, pp. 387-401.

[135] Chow, P. M-Y. (1993) <u>**Control Volume Unstructured Mesh Procedure for Convection-Diffusion Solidification Processes**</u>, Ph.D. Thesis, University of Greenwich, London, UK.

[136] Lau, H. J. (1994) **Ball Grid Array Technology**, *McGraw-Hill Professional* (ISBN: 007036608X).

[137] Lau, H. J. (ed.) (1991) **Solder Joint Reliability: Theory and Applications**, published by Van Nostrand Reinold, New York.

[138] Lau, H. J. and Pao, Yi-Hsin (ed.) (1997) **Solder Joint Reliability of BGA, CSP, Flip Chip, and Fine Pitch SMT Assemblies**, *McGraw-Hill*.

[139] Surface Evolver, http://www.geom.unm.edu/software/evolver

[139] Lorenzen, T. J. and Anderson, V. L. (1991) **Design of Experiments**, Marcel Dekker.

[140] Plackett, R. L. and Burman, J. P. (1946) "The Design of Optimum Multifactorial Experiments", Biometrika 33, pp. 305-325.

[141] Taguchi, G. and Konishi, S. (1987) "Orthogonal Arrays and Linear Graphs", American Supplier Institute, Dearborn, MI.

[142] Box, G. P. and Draper, N. R. (1987) **Empirical Model-Building and Response Surfaces**, Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons Inc., New York.

[143] Vanderplaats, G. N. *et al.* (2002) VisualDOC 3.0 *Theoretical Manual* and *DOC User's Manual*, Vanderplaats Research & Development Inc., Colorado Springs, http://www.vrand.com

[144] LMS Optimus Rev.4.0 *Manual*, LMS International, http://www.lmsintl.com

[145] Cartuyvels, R. and Dupas, L., DEBORA *User's Manual*. IMEC, May 1993.

[146] *TMA WorkBench Version 2.0 User's Manual*, Technology Modeling Associates, Inc., Sunnyvale, California, Mar. 1996

[147] *VWF Interactive Tools User's Manual*, SILVACO International, first ed. Santa Clara, CA, Mar. 1994.

[148] COMPACT 3.0.0 *Manual*, Cosinus Computing BV , http://www.cqm.nl

[149] Box, G. P., Hunter, W. G. and Hunter, J. S. (1978) **Statistics for Experimenters** , Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons Inc., New York.

[150] Kodiyalam, S., Lin, J.S. and Wujek, B. A. (1998) "Design of Experiments Based Response Surface Models for Design Optimization", *Proceedings of the 39-th AIAA/NASA/USAF/ISSMO Symposium of Multidisciplinary Analysis and Optimization*, pp. 143-149. St. Louis, Sept. 1998.

[151] Kelley, J. E. (1960) "The Cutting Plane Method for Solving Convex Programs", J. SIAM, Vol. 8, pp. 702-712.

[152] Brekelmans, R., Driessen, L., Hamers, H. and den Hertog, D. (2001) "A New Sequential Optimization Approach to Product and Process Design Involving Expensive Simulations", *Proceedings of the First ASMO UK/ISSMO Conference on Engineering Design Optimization*, pp.56-63.

[153] Lauenberger, D. G. (1984) **Linear and Non-Linear Programming**, Addison Wesley, New York, pp. 167-292

[154] Corne, D., Dorigo, M. and Glover, F. (1999) **New Ideas in Optimization**, McGraw-Hill Int. (ISBN 007 7095065).

[155] Michalewicz, Z. and Dasgupta, D. (editors) (1997) **Evolutionary Algorithms in Engineering Applications**, Pub. Springer-Verlag.

[156] Venter, G. and Sobieski, J. (2002) "Particle Swarm Optimization", Published by AIAA 2002, paper No. 1235

[157] Dorigo, M., Di Caro, G. and Gambardella, L. (1998) "Ant Algorithms for Discrete Optimization", Technical Report 98-10, IRIDIA, University Libre de Bruxelles.

[158] Leon, X., Reinikainen, T., Ren, W., *at al.* (2003) "A Simulation-Based Multi-Objective Design Optimisation of Electronic Packaging under Thermal Cycling and bend Loading", *Proceedings of the 4-th International Conference on Thermal and Mechanical Simulation and Experiments in Micro-Electronics and Micro-Systems (EuroSIME* 2003), pp. 485-490.

[159] Nemhauser, G. L. and Wolsey, L. A. (1988) **Integer and Combinatorial Optimisation**, Chapter 3, John Wiley & Sons.

[160] Parry, J. *at al* (2001) "Simulation-Based Design Optimisation Methodologies Applied to CFD", available from http://www.flomerics.com

[161] Ghosh, D. and Vanderplaats, G. (1997) "Development of a Flexible Design Optimization Capability", *Conference Proceedings* "Optimization in Industry", Palm Coast, Florida, March 23-27.

[162] Ghosh, D., Garcelon, J. H., Balabanov, V., Venter, G. and Vanderplaats, G. (2000) "VisualDOC – A Flexible Design Optimization Software System", 8-th AIAA/USAF/ NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization, 6-8 Sept 2000, Long Beach, CA, USA, Paper No. AIAA-2000, Paper No. 4931.

[163] Venter, G. and Watson, B. (2000) "Efficient Optimization Algorithms for Parallel Applications", 8-th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization, 6-8 Sept 2000, Long Beach, CA, USA, Paper No. AIAA-2000, Paper No. 4819.

[164] Vanderplaats, G. (2000) "Very Large Scale Optimization", 8-th AIAA/USAF/ NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization, 6-8 Sept 2000, Long Beach, CA, USA, Paper No. AIAA-2000, Paper No. 4809.

[165] Ghosh, D., Garcelon, J. H., Balabanov, V. and Vanderplaats, G. (1998) "Development of a Flexible Design Optimization Study Tool", In AIAA-98, Paper No.4726.

[166] Lu, H., Stoyanov, S., Bailey, C., Hung K. C. and Chan, Y. C. (2001) "A Modelling and Experimental Analysis of the No-Flow Underfill Process for Flip-Chip Assembly", *Proceedings of the Fourth International Symposium on Electronic Packaging Technology* (ISEPT), August 8-11, 2001, Beijing, China, pp. 338-343.

[167] Lu, H., Hung, K. C., Stoyanov, S., Bailey, C. and Chan Y. C. (2002) "No-Flow Underfill Flip Chip Assembly – An Experimental and Modeling Analysis", Microelectronics Reliability, Vol. 42, pp. 1205-1212.

[168] Wong, C. P., Shi, S. and Jefferson, G. (1997) "High Performance Low Cost Underfills for Flip Chip Applications", *IEEE 47th ECTC Proceedings*, San Diego, Calif. 1997, pp. 850-858.

[169] Shi, S. H., Yao, Q., Qu, J. and Wong, C. P. (2000) "Study on the Correlation of Flip Chip Reliability with Mechanical Properties of No Flow Underfill Materials", International Symposium on Advanced Packaging Materials, pp. 271 – 277.

[170] Wong, C. P., Baldwin, D., Vincent, M. B., Fennell, B., Wang, L.J. and Shi, S. H. (1998) "Characterization of a No-Flow Underfill Encapsulant during the Solder reflow Process", *Proceedings 48-th Electronic Components and Technology Conference '98*, pp. 1253-1259.

[171] Shi, S. H. and Wong, C. P. (1999) "Recent Advances in the Development of No-Flow Underfill Encapsulants – a Practical Approach towards the Actual Manufacturing Application", IEEE Transactions on Electronics Packaging Manufacturing, Vol. 22, No.4, pp. 331-399.

[172] Fan, L. H., Shi, S. H. and Wong, C. P. (2000) "Incorporation of Inorganic Filler into the No-flow Underfill Material for Flip Chip Application", International Symposium on Advanced Packaging Materials, pp. 303-310.

[173] Tu, P. L., Chan, Y. C. and Hung, K. C. (2001) "Reliability of the micro-BGA Assembly using No flow Underfill", Microelectronics Reliability, Vol. 41, pp. 1993-2000.

[174] Dieter, G. E. (1998) **Mechanical Metallurgy**, McGraw-Hill Book Company, New York, pp. 241-246.

[175] Dudek, R. *et al.* (1997) "An Efficient Approach to Predict Solder Fatigue Life and its Application to SM and Area Array Components", *Proceedings of the 47th Electronic Components and Technology Conference*, pp. 462-471.

[176] Lu, H., Hung, K. C., Stoyanov, S., Bailey, C. and Chan Y.C. (2001) "Reliability Analysis of No-Flow Underfill Materials" *Proceedings of InterPACK* 2001(CD proceedings, paper No. IPACK2001 15814): The Pacific Rim/ASME International Electronic Packaging Technical Conference and Exhibition, July 8-13, 2001, Kauai, Hawaii, USA.

[177] Stoyanov, S., Bailey, C., Lu, H. and Cross, M. (2002) "Integrated Computational Mechanics and Optimization for Design of Electronic Components", Published in **Optimization in Industry**, Ian C. Parmee and Prabhat Hajela (Eds.) Publisher: Springer-Verlag London Ltd., pp. 57-70 (ISBN: 1-85233-534-3)

[178] Stoyanov, S., Bailey, C. and Lu, H. (2001) "Optimization Tools for Flip-Chip Design", *Proceedings of InterPACK* 2001 (CD proceedings, paper No. IPACK2001-15741): The Pacific Rim/ASME International Electronic Packaging Technical Conference and Exhibition, July 8-13, 2001, Kauai, Hawaii, USA.

[179] Stoyanov, S., Bailey, C. and Cross, M. (2001) "Integrating Computational Mechanics and Numerical Optimization for the Design of Material Properties in Electronic Packages" *TMS Proceedings of Computational Modelling of Materials, Minerals and Metals Processing*, September 23-26, 2001, San Diego, California, USA, pp. 551-561

[180] Stoyanov, S., Bailey, C., Lu, H. and Cross, M. (2002) "Solder Joint Reliability Optimization", *Proceedings of APACK 2001 International Conference on Advances in Packaging*, December 5-7, 2001, Singapore, pp. 178-186.

[181] Stoyanov, S., Bailey, C. and Cross, M. (2002) "Optimization Strategies in Reliability for Electronic Packages", Emerald Journals, Soldering & Surface Mount Technology, Vol. 14, No. 1, pp. 49-58.

[182] Stoyanov, S. and Bailey, C. (2003) "Optimisation Modelling for Electronic Product Design", *Proceedings of PREP2003: Postgraduate Research Conference in Electronics, Photonics, Communications and Software*, 14 April - 16 April at Exeter University, Exeter, UK, pp. 131-132.

[183] Hong, B. Z. (1998) "Thermal Fatigue Analysis of a CBGA Package with Lead-free Solder Fillets", *Proceedings of InterSociety Conference on Thermal Phenomena*, pp. 205-211.

[184] Wong, C. P., Shi, S. H. and Jefferson, G. (1998) "High Performance No-flow Underfills for Low-cost Flip-chip Applications: Material Characterization", IEEE Transactions on Components, Packaging and Manufacturing Technology, Part A, Vol 21 No 3, pp. 450-458.

[185] Stoyanov, S. and Bailey, C. (2002) "Response Surface Modelling and Optimization for Reliable Electronic Packaging", The Fifth International IEEE Symposium on High Density Packaging and Component Failure Analysis in Electronics Manufacturing, HDP 2002, June 30 July 3, 2002, Shanghai, China, pp. 49-57 (ISBN: 0-7803-9822-X ; ISBN: 0-7803-9823-8) * Keynote paper

[186] Strauss R. (1998) **SMT Soldering Handbook**, Published by *Newness* / Butterworth-Heinemann.

[187] Frear, D. R., Burchett S. N., Morgan H. S., and Lau J. H. (editors) (1994) **The Mechanics of Solder Alloy Interconnects**, Published by *Chapman & Hall*, New York.

[188] Stratton, P. and Adams, S. "Protective Atmospheres for Electronic Packaging and Circuit Assembly", BOC Gases Document.

[189] Wang, B. (2001) "Parameter Optimisation in Multiquadratic Response Surface Approximations", 6th U.S. National Congress on Computational Mechanics, Symposium on Design Optimisation Applications in Industry, Dearborn, MI.

[190] Fourie, P. C. and Groenwold, A. A. (2000) "Particle Swarms in Size and Shape Optimization", *Proceedings of the International Workshop on Multidisciplinary Optimization*, Pretoria, South Africa, Aug 7-10, pp. 97-106.

[191] Reid, R., Prausnitz, J. and Sherwood, T. (1977) **The Properties of Gases and Liquids**, McGraw-Hill Book Company.

[192] Punch, J. and Dalton, T. (1999) "Thermal Management of Electronic Systems", Woodlands Park Hotel Cobham, 11-12/03/1999, Surrey, UK, PEI Technologies Seminar Proceedings.

[193] Stoyanov, S. and Bailey, C. (2003) "Optimisation and Finite Element Analysis for Reliable Electronic Packaging", *The 4th International Conference on Thermal & Mechanical Simulation and Experiments in Micro-Electronics and Micro-Systems, EuroSIME* 2003, Aix-en-Provence, France, March 30 - April 2, 2003, pp.391-398 (ISBN (book): 0-7803-7054-6, ISBN (CD-ROM): 0-7803-7056-2) * Keynote paper.

[194] Warner, M. *et al* (2002) "FLO/STRESS: An Integrated Software Module to Predict Stress in Electronic Products", Computing & Control Engineering Journal, June 2002, pp. 143-148.

[195] Parry, J. et al. (1998) "The Development of Component-level Thermal Compact Models of a C4/CBGA Interconnect Technology: the Motorola PowerPC 603 and 604 RISC Processors", IEEE Trans. Comp. Pack. & Man. Tech., March 1998, Vo. 21, Part A, (1), pp. 104-112.

[196] Parry, J., Rantala, J. and Lasance, C. (2001) "Enhanced Electronic System Reliability - Challenges for Temperature Prediction", 7th THERMINIC Workshop, Paris France, Sep 24 – 27.

[197] Parmee, I. and Hajela, P (editors) (2002) **Optimisation in Industry**, Published by Springer-Verlag.

[198] Bendsoe, M.P. (1989) "Optimal Shape Design as a Material Distribution Problem", Structural Optimization, Vol. 1, pp.193-202.

[199] Swan, C.C. and Kosaka, I. (1997) "Reuss and Voigt Mixing Rules for Variable Topology Material Layout: Linear Elasticity", International Journal for Mum. Meth. in Eng., Vol. 40, pp. 3033-3057.

[200] Leiva, J.P., Watson, B.C. and Kosaka, I. (1999) "Modern Structural Optimization Concepts Applied to Topology Optimization", *Proceedings of the 40th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Material Conference*, April 12-15, 1999, St. Louis, MO, pp. 1589-1596.

[201] Flomerics Ltd, http://www.flomerics.com

[202] Foulds, L.R. (1981) **Optimization Techniques: An Introduction**, Published by Springer-Verlag New York Inc.