Power Cycling of Flip Chip Assemblies

By
Andrew Ochana

A Doctoral Thesis submitted in partial fulfilment of the requirements for the award of Doctor of Philosophy of Loughborough University

September 2004

Abstract

This thesis demonstrates the feasibility of power cycling a "Flip Chip" assembly for reliability assessment. The assemblies studied were Si on Si Multi-Chip Modules (MCMs) that were mounted on either an organic FR4 or a metallic (copper) substrate. The aim of the work was to investigate how anisothermal temperature distributions caused by local power inputs could influence the reliability of devices that would not be expected to be effected by thermal cycling. This work was performed using two complementary techniques: physically manufacturing assemblies in order to perform "real" power cycles, and utilising Finite Element Analysis (FEA) to perform "virtual" cycles.

The MCMs consisted of "heater chips" into which electrical power could be dissipated to heat the device locally. These heater chips were flip-chip bonded to Si carrier chips by solder interconnections and the entire assembly was then mounted onto a substrate. The thermal performance of the MCMs as a result of power input was characterised under steady state and cycling conditions using a number of techniques including thermal imaging. In addition, many devices were power cycled to evaluate their reliability.

In addition to the evaluation of real devices, a three dimensional finite element model was developed of the same structures. The model initially provided thermal data that was validated against that obtained from the real devices operating under the same environmental and power input conditions. In addition, it enabled the stress level within the solder joints to be evaluated so that insight to the long-term reliability of the assemblies could be gained.

The results of the experimental and modelling work have shown that the thermal performance and reliability of the devices depend strongly on the substrate onto which the MCMs are bonded. It was found that, using a copper substrate, the temperatures reached within the assemblies were greatly reduced and that the reliability during power cycling was enhanced.

Keywords: Flip Chip, Finite Element Analysis, Power Cycling, Reliability, Multi-Chip Modules, Solder.

Acknowledgements

First of all I would like to thank God for granting the opportunity for me to complete this, without whom this would not be possible.

I would particularly like to give special thanks to my supervisors, Dr David Hutt and Mr David Whalley for their continual guidance and support through my time completing this work. I would also like to express my gratitude to the technical staff, in particular Neville Carpenter and Andy Sandaver who have repeatedly gone beyond the call of duty to help me complete some of the practical work. I would like to thank Dr Amin Al-Habaibeh for the extended loan of some of the specialist thermography equipment, and also Furo Jumbo for offering me a valuable lifeline towards the end of my work.

I would also like to say thank you to the rest of my department, in particular my colleagues within my research group for continual support, the Students Union for keeping me occupied in my time away from my studies. Lastly I would like to thank all my friends and family for their relentless patience and understanding.

If there's anyone I have left out, it was not intentional so thank you

I would like to dedicate this thesis to my parents: - Richard and Mary Ochana, just to say thank you for continually believing in me when it seemed that everyone (including myself) had doubted my potential.

Happy Retirement

Contents

Glossary

Abreviations

BGA — *Ball Grid Array:* A packaging method where the intermediate body has external connections made by attaching solder balls to large diameter pads, and the final connection is made typically by reflowing the component.

CTE — *Coefficient of Thermal Expansion:* The rate at which a material expands when it is heated up.

DCA — *Direct Chip Attach:* (also known as flip chip on board (FCOB)) A flip chip packaging technology where the chip is placed directly on the circuit board. This method has the benefit of offering the lowest packaging profile.

DNP — *Distance from Neutral Point*: Commonly used when describing the location of a feature on a die (such as a connection joint) with respect to some location (normally central point).

FE — *Finite Elements:* Generic name for the science that incorporates FEM and FEA; this involves providing an numerical approximate solution to a problem that may otherwise be unsolvable. This method is applicable to a wide range of physical and engineering problems.

FEA — *Finite Element Analysis:* Practical implementation for the FEM. This may be divided in to the pre-processing, analysis and post-processing stages. Pre-processing involves first discretising a given domain (geometry) into smaller sub regions (elements), applying the conditions (boundary conditions, material properties), the analysis involves submitting the job to be solved and the post processing is extrapolating the useful information from it.

FEM — *Finite Element Method:* A method of solving partial differential equations, where the equations may involve a function u(x) for all x values defined in the domain with respect to a given boundary condition. The purpose od the method is to determine an approximation for the function u(x)

FR4 — *(FR= Flame Retardant)* Typical material used as an insulating base for circuit boards. It is manufactured from woven glass fibres that are bonded together with an epoxy resin. Subsequently the board is cured using temperature and pressure causing the resin to melt and bond such that the board is rigid.

IC — *Integrated Circuit*: A device on which a number of components, typically transistors are formed on the surface on a single piece of semiconductor

IPA — *Isopropyl alcohol:* (alternative name 2-propanol) an alcohol commonly used for cleaning applications

MCM — *Multi-Chip Module:* A packaging technology where one or more die is mounted on to an intermediate body, which is then mounted on to a parent substrate.

PCB — *Printed Circuit Board* (also known as a Printed Wiring Board in the US): A type of circuit board that has artwork superimposed or "printed" on one or both sides; it may also contain internal signal layers as well as power and ground planes.

SEM — *Scanning Electron Microscope:* A powerful microscope where images are created using electrons as opposed to light waves.

SMT *Surface Mount Technology:* A technique for populating hybrids, multi-chip modules and circuit boards where packaged components are mounted directly on to the surface of the substrate. A layer of solder paste is stencil printed onto the pads and the components are attached by placing them on the paste (the viscosity of the past is normally sufficient to temporarily hold the components. They are then soldered using a either a vapour phase or reflow soldering method.

THT *Through Hole Technology:* A technique for populating part or all of circuit board where components are inserted into plated through holes (vias). When all the components have been inserted they can be subsequently soldered to the board, typically either a wave or hand soldering technique is utilised

UBM *Under Bump Metallurgy:* A layer of material that is able to wet with solder (typically gold or nickel) that is superimposed on a material that the solder would otherwise not wet with. The material is superimposed on the designated area by a process such as electroplating.

Names for parts of an electronic assembly

(adapted from http://www.maxmon.com/glossary.htm)

Artwork Actual circuit layout (including pads and tracks) as used to manufacture a circuit board

Circuit Board Generic name for several interconnection methods; the board may be rigid flexible and may be either single sided, double sided, or multilayered.

Die Generic name for an unpackaged integrated circuit (IC)

Flux Chemical applied to the solderable surfaces (e.g. PCB pad or UBM) that removes the oxides that will inevitably impede the solder "wetting", therefore a good solder joint to form.

Lead frame A metallic frame containing leads and a base such that an unpackaged IC may be attached. Once correctly positioned the outer part of the frame can be cut away and the leads bent into the required shape.

Pad A conductive area on the substrate where one of the following may occur: a) the pad may be a designated area where a component is to be placed; b) the pad may connect to a via (e.g. in the case of double sided/ multi-layer PCBs) c) the pad may allow for external probing.

Schematic Name for a theoretical (drawn) circuit diagram

Solder An alloy consisting of tin and other metal(s) used to join less fusible materials together. The solder has a low melting temperature (compared with the materials it is joining) and this allows the tin in the solder to "wet" the other materials (typically either copper, gold or nickel) and form the intermetallic layer that is necessary for an adequate solder joint.

Solder paste microscopic solder balls combined with flux that have high viscosity and can be dispensed on to a designated area. Normally this area is a pad on the board where a component is intended to be placed later.

Solder bump Solder that has been placed on a pad and then reflowed (heated to a sufficient temperature that the solder melts) and the surface tension of the molten solder pulls the solder into a spherical shape. Upon cooling down, this spherical shape is retained.

Solder joint Solder that is connecting the pad on the substrate with the corresponding connection on the component

Solder Mask Layer applied to surface of the substrate to prevent solder wetting any other metallic surface; holes are patterned into the mask to correspond with the pads on the substrate.

Substrate The base layer for any integrated circuit, hybrid, multi- chip module or circuit board. A substrate may be manufactured from several materials depending on the product specification requirements. Typical materials include FR4, semiconductors and those that are ceramic based.

Track A narrow conductive connection between two different pads such that a component may emit or receive signals. The tracks are superimposed on the circuit board and may relay signals with other components or they may run to an external connection to a parent/ child board or component.

Via A hole drilled in a circuit board to link two or more conductive layers of a substrate.; the hole may be bare in the case when used in through hole technology or may be filled/lined with a conductive material when used for a double sided/ multi-layered PCB

Underfill An epoxy based material utilised in flip chip technology applied between the chip and the parent substrate. The purpose is to mechanically couple the chip and substrate such that stresses resulting from thermal expansion mismatches are distributed.

Wafer A thin slice cut from a pure semiconductor

IC package configurations

Wire Bonding The process of connecting the pads on an unpackaged IC to corresponding pads on a substrate using fine wires. Wire bonding may also be used to connect the pads on an unpackaged IC, hybrid or multi chip module to the leads of a component package.

Area Array A packaging technology where the area of a components base is utilised for interconnection purposes.

Grid array A packaging technology where components external connections are arranged as an array on the base of the package; there are several types

- Pin Grid Array The interconnection layer consists of conducting pins or leads
- Pad Grid Array The external connections are pads
- Ball Grid Array Similar to that of the Pad Grid Array however the pads have solder balls attached to them.

Peripheral array A packaging technology where only the area near the edges of the board are used for interconnection purposes.

Chip Assembly Procedures

Reflow A method of soldering surface mount components attached to a board with solder paste. The populated board is passed through an oven on a conveyor where it is progressively heated to a temperature sufficient to melt the solder paste such that a good connection is formed.

Solder Wetting The alloying procedure that occurs when the tin element of solder comes into contact with a compatible solid material. The tin atoms join with the pad/ UBM material and a resultant intermetallic compound is formed; the intermetallic is a necessary product to ensure that an adequate joint has been formed.

Stencil printing A technique for depositing solder paste on to either a substrate (surface mount) or a wafer (flip chip). The wafer/ substrate is placed on the surface and a stencil with holes (for the pads) is correctly aligned. Solder paste is then placed on top of the stencil and a squeegee is then passed across the pattern such that the solder paste is forced through the holes and onto the pads. Following this the wafer or substrate is inspected and either is populated with components or reflowed (depending on method used).

Vapour phase A method of soldering surface mount components attached to a board with solder paste. The populated board is lowered into a tank containing boiling hydro carbons that consequently form a vapour cloud. The temperature of the vapour cloud is sufficient to melt the solder. However this method is now becoming less popular due to environmental legislations.

Wafer bumping A flip chip technique where spheres are formed on the dies pad. Solder paste is deposited on the pads using a stencil printing technique. The wafer is then passed through a reflow oven such that the solder melts and solidifies resulting in spherical solder "bumps". Once complete, the wafer can be diced and individual die can be subsequently placed on the appropriate pad.

Wave soldering A method of soldering through hole components to a circuit board. A wave of molten solder is generated in a tank and the board is passed over the solder wave by a conveyor, the wave then catches the bottom and the components are soldered.

Miscellaneous

Solder Wetting The alloying procedure that occurs when the tin element of solder comes into contact with a compatible solid material. The tin atoms join with the pad/ UBM material and a resultant intermetallic compound is formed; the intermetallic is a necessary product to ensure that an adequate joint has been formed.

Thermal Cycle A product verification method whereby a component is physically heated and cooled according to a predefined specification. Thermal cycling is normally carried out in an oven such that the temperature of the component is dictated by the temperature of the surrounding environment.

Power cycle A product verification method whereby a component generates heat internally as a result of a power load applied to it.

Symbols used

R	Resistant measurement	(Ω)
t	Time measurement	(s)
T	Temperature measurement	(°C OR °K)
U	Fluid velocity measurement	(m/s)
k	Thermal Conductivity	(w/M^2xK)
h	Convection heat transfer coefficient	(w/M^2xK)

R_{obs}	Measured resistance of assembly with power applied
R_{RT}	Measured resistance of assembly at room temperature
T_s	Temperature of a body surface
T_∞	Ambient temperature (temperature of surroundings)
U_∞	Free flow fluid velocity
δ	Velocity boundary layer
θ	Thermal boundary layer
v	Kinnematic viscosity
h_x	Specific convection heat transfer coefficient
x	length

Re_x	Reynolds number for given conditions
Nu_x	Nusselt number for given conditions

Chapter 1: Introduction

1.1: Trends in the Manufacture of Electronic Products

Every 18 months the number of transistors found on a typical silicon chip doubles
(Moore's law 1965)

In the present consumer electronics climate, there are persistent demands for a product to be of superior quality and performance, yet smaller size than its predecessor. To illustrate the effects of just 6 years of product improvement, table 1.1 shows the volume reduction of a typical domestic "camcorder" manufactured at two different periods in the 1990's (1,2). In addition, the seemingly continuous increase in computer processing speeds and memory capacity are well known.

Feature	1992	1998
No. of components	1800	848
Density	12 components/cm^2	21 components/cm^2
Volume	2380cm^3	1598cm^3
Viewing unit	B/W charge coupled device	Colour LCD

Tables 1.1a: *Illustration of electronic trends over 6 years (1)*

Many of these reductions in size and increases in performance were due to persistent changes in manufacturing methods and endeavour to reduce package profile. Possibly the most significant initiative in package reduction was the introduction of the **Printed Circuit Board** (PCB). PCBs are typically manufactured from FR4, which is a glass fibre epoxy laminate though they may also be ceramic based. The PCB initially had its electronic components secured to it by passing the connecting leads through the board in specified positions and then soldering them to tracks that were superimposed on the PCB (Through Hole Technology fig 1.1a). The components were originally attached to the board by hand soldering where an operator soldered each connection with a soldering iron; this method was not only slow but also potentially introduced human error. Therefore a more efficient manufacturing method was sought, which greatly reduced the risk of human error and allowed several components on a PCB to be soldered simultaneously. This initially lead to the introduction of *wave soldering* where this allowed a board populated with through-hole components to be mass-soldered by a molten solder wave passed underneath the board; the subsequent drive

for lower profile led to the development of Surface Mount Technology (SMT). In contrast to the through hole technology, SMT PCBs were manufactured with their tracks on the same side of the board as the soldering was carried out using either a *vapour phase* or *solder paste "reflow" oven* soldering where all the mounted components could be soldered simultaneously. These simultaneous soldering methods increased the production rate as well as offering greater assurance of product quality.

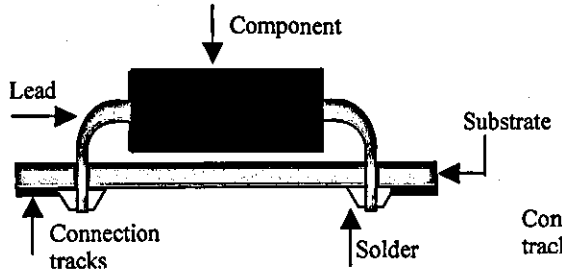


Figure 1.1a:
Illustration of through–hole assembly. Note how the leads from the component pass through the substrate through holes drilled for this purpose (vias). The lead ends are hand-soldered or wave soldered to connect them to the relevant tracks.

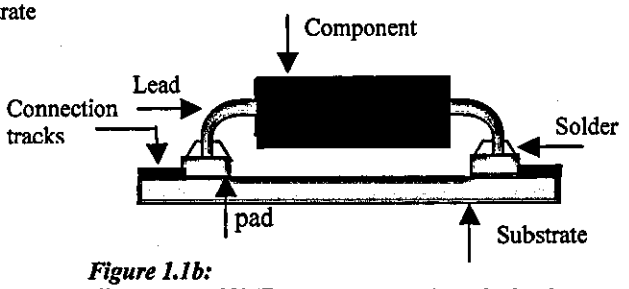


Figure 1.1b:
Illustration of SMT assembly. Note how the leads from the component are connected to pads on the same side it is mounted on. The connections are soldered simultaneously, typically by passing the board through a reflow oven.

Initially the components connected to the board were single passive devices such as resistors, capacitors and transistors. In subsequent years, there were developments of integrated circuits based on silicon that required several input and output connections. As the capabilities of a single packaged electronic component increased, the number of necessary Input/Output (IO) connections also inevitably increased. These were accommodated for by utilising a lead–frame where several leads would pass from the component (typically a silicon die) and connect to the board as shown in figure 1.1c. To connect the pads of the silicon die to the lead frame, wire bonding was used and until the present day has been used to successfully secure dies with many IO connections. This method was available for through hole assembly, however, its application to SMT allowed for smaller lead profiles and smaller pad dimensions to be implemented in the chip/board design, therefore providing a greater number of potential connections for a given area. In order to avoid the additional packaging associated with the lead frame, the die was directly wire bonded to the PCB in some devices.

In recent years the drive towards smaller products with more functionality has led to the introduction of other packaging methods that enable the maximum number of IO

connections in the minimum area. Typically this has required the introduction of area array devices where an entire face of a chip can be utilised for connections to and from the substrate. However this is a limitation of the conventional wire bonding interconnection where only area close to the face edges can be used for connection pads, therefore new technologies have been developed to accommodate for this connection method. For SMT the area array connections are made using solder balls in place where the connections are to be made; this connection methodology is known as Ball Grid Array (BGA). Variations of this methodology exist, a common example is when the solder balls are substituted for pins resulting in a Pin Grid Array (PGA). The ultimate extension of the reduction in package size is the removal of all packaging and direct connection of the Si chip to the PCB surface. The technique of flip chip assembly enables bare die to be connected to a PCB using solder or a conducting adhesive as shown in figure 1.1d. At present, there is a considerable interest in flip chip assembly; in particular, the mobile telecommunications industry has benefited due to increased research in flip chip resulting in economic mass production of smaller devices (1). Flip chip was first researched in the 1950's by Lucent Technologies (formerly AT&T), and later enhanced when IBM Yasu developed the Controlled Collapse Chip Connection (C4) process in the 1960's. In contrast to the mature wire bond manufacturing method, flip chip involves mounting a chip with its topside down and interconnections made by solder bumps to a matching pad array on the substrate (typically a PCB) for the case of Direct Chip Attach (DCA) or it may be attached to an intermediate body in the case of a Multi-Chip Module (MCM) where connections to the substrate are typically made utilising BGA technology.

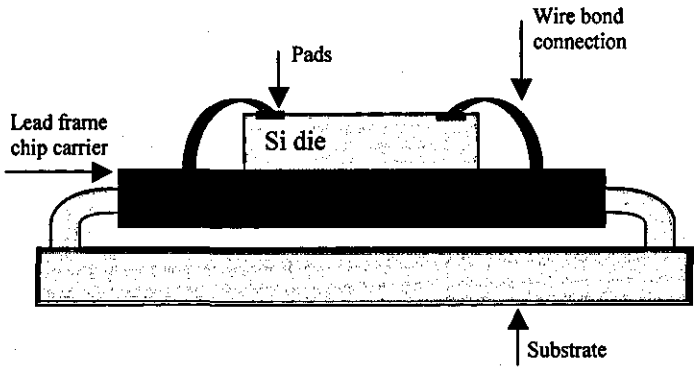


Figure 1.1 c:
Illustration of wire bond chip assembly. The example shows a die attached to the substrate via a chip carrier. The connection pads face upwards on the die where they are connected to the chip carrier utilising wire bonding. Wire bonding directly to the substrate is less common.

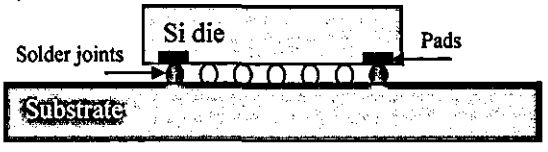


Figure 1.1d:
Illustration of flip chip on board assembly method or DCA. In this instance, the chip has been directly attached to the board, though can be attached to a chip carrier first. Note the much shorter distance an electronic signal would have to travel through compared with the wire bond example as the pads are themselves closer to the substrate. The grey solder balls show the more common peripheral array solder joints while the clear solder balls represent the extra potential connections if an area array grid pattern is used.

1.2: Advantages of Flip Chip Technology

With increasing performance and consumer demands, flip chip is likely to become ubiquitous in electronics packaging and assembly, especially as the limits are being reached for wire bonding technology in terms of IO connections per available area. This enables the continuation of the packaging trends witnessed in recent years. As previously stated, the manufacturing of mobile phones greatly utilises flip chip technology and it is anticipated that other technologies will soon adopt this trend. There are several reasons why flip chip is an attractive manufacturing method compared with SMT or wire bonding (3,4), among the obvious are:

Lower package profile: As stated in section 1.1, the consumer demands smaller packages with increasing functionality. Flip Chip On Board (FCOB) or DCA is the smallest packaging technology as it eliminates the peripheral wires or leads that are present in wire bond and SMT manufacturing methods respectively. In addition as the

connections are underneath the chip (in contrast to around the perimeter) they not only allow denser connections, but also free up extra space on the substrate.

Area Array packages: The ever-increasing number of IO connections required from a die are a significant factor when considering manufacturing processes; flip chip enables full area array interconnect as opposed to only the peripheral areas of a die as is the case with both wire bonding or conventional surface mount devices. Therefore extra area of the board can be utilised for IO connections without having to reduce the pad size on the peripheral array chips. Figure 1.2 shows the available pad area for peripheral array and the potential connection space when area array is implemented, it can be seen that a much larger area on the die can be utilised for interconnections.

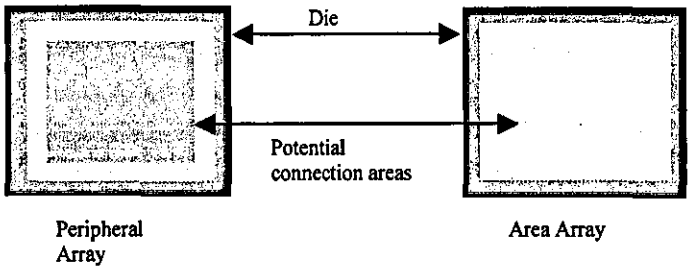


Figure 1.2:
Illustration of the potential connection space available if an area array pattern is used, compared with the more conventional peripheral array available for wire bond. The grey area represents the die while the red area shows the potential connection area for the respective methods

Shorter IO connections: The reduced distance travelled by the electrical signal (a flip chip solder joint height is typically less than 100μm tall, compared to the wire bond or lead frame which may be several mm long) enables signals to be relayed faster; this feature is becoming more critical as the speeds of microprocessors are constantly increasing. In addition to the speed increase, the shorter distance implies less impedance and noise.

Manufacture economy: According to Lau (4), the cost of attaching a die to a PCB using wire bond technology is largely dependent on the number of connection pads the chip contains; for example, assuming all other aspects of a die are constant, it may be the case that only 4 wire bonders are necessary for a die with 30 IO pads, whereas 70 wire bonders are required for a die with 600 pads (4). Clearly as the amount of connection points per chip increases an unacceptably large number of wire bonders

will be needed. Conversely, for flip chip assembly, only one reflow oven is required regardless of the amount of pads on chips, it can therefore clearly be seen that as the amount of pads increases (as it will inevitably, as consumer demands continue to grow) the wire bond method will become very impractical and uneconomical if not totally prohibitive. Furthermore, as a reflow soldering process is used to secure the flip chip to the board it is directly compatible with conventional SMT methods.

1.3: Disadvantages of Flip Chip Technology

Though there are several disadvantages of flip chip (4), the fundamental drawback of flip chip is the potential impact of thermal effects on product reliability; when an electronic die is in service (or powered on) it is well known that the temperature will increase in proportion with the power level supplied. Consequently the die will expand according to its Coefficient of Thermal Expansion (CTE); in addition, the substrate may also heat up and expand with respect to its own CTE. The CTE values may be substantially different between the die and the substrate and this will result in a shear stress in the small solder joints as shown in figure 1.3. While materials thermally expanding at different rates is not a new concept in electronics reliability assessment, it is especially emphasised from the flip chip perspective due to the small geometries of the solder joints (height is typically <100μm); resulting in little room for flexibility. The wires and lead frames used in wire bonding and surface mount technology respectively, are long and flexible enough to absorb the CTE mismatches between the die and substrate.

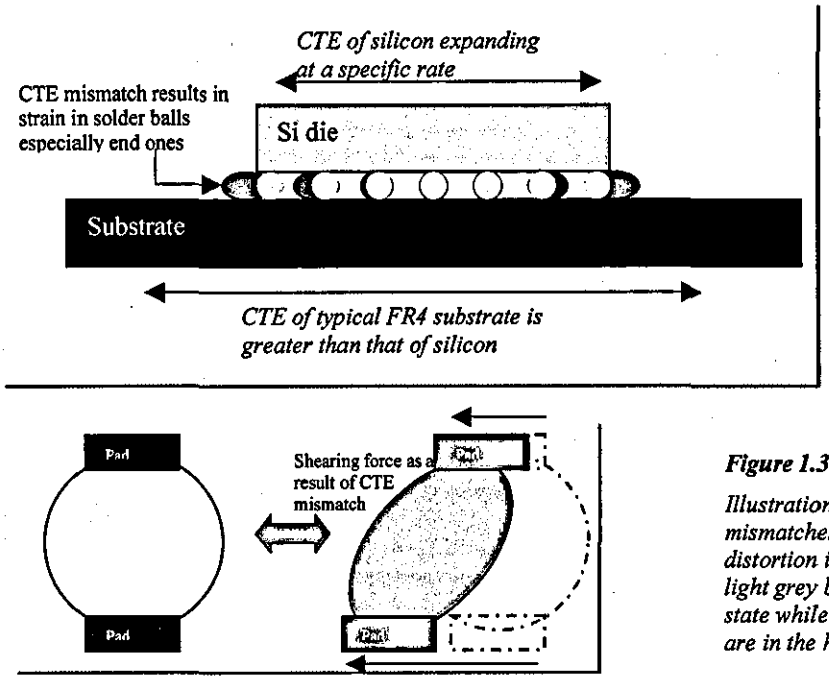


Figure 1.3:

Illustration of effects of CTE mismatches on solder joints. The distortion is exaggerated. The light grey bumps are in the cool state while the dark grey bumps are in the hot state

As the product experiences temperature changes either due to testing or use in service, the die and substrate will be continuously expanding and contracting, therefore increasing and decreasing the shear stress. There are several failure mechanisms that may be invoked by this shearing action; they may be instantaneous such as plastic deformation, or they may be time dependent as is the case with creep and fatigue (5,6). These effects shall be looked at in more detail in subsequent chapters.

Product Reliability Evaluation methods

Obviously products require some form of testing to investigate the extent these mechanisms mentioned contribute to the demise of a product. A common method used for assessing damage mechanisms is thermal cycling. This involves heating and cooling a component isothermally and holding it at the high/low temperatures. The isothermal temperature of the assembly has been considered adequate to investigate the CTE mismatches of the die and the substrate and to invoke the time dependent mechanisms described. This method has been used for many years to demonstrate the reliability of electronic devices including flip chip. However this is a somewhat artificial test condition and may be a misrepresentation of the stresses an assembly would endure in a real environment. The main issue regarding thermal cycling is that the heat source is supplied from an external source, hence is independent of the test-vehicle (they are normally placed in an oven and heated this way) resulting in an isothermal temperature distribution. In reality the heat from a component will be generated internally as a result of the power supplied to it. As a consequence, the temperature of the component may be significantly greater than the substrate; hence an anisothermal temperature distribution exists. Such an anisothermal temperature distribution is more likely to be captured using an alternative product testing method, namely **power cycling**. Power cycling involves supplying a specified power level to an assembly for a given time period and then removing the power for an equivalent "off" period. The heat involved in a power cycle is generated internally (i.e. from within the device) and the temperature change endured by the component when the power is supplied/removed forms a complete ON/OFF power cycle which is equivalent to the conventional high/low thermal cycle; the fundamental drawback of power cycling is that in most cases the ambient temperature cannot be regulated, nor can a strict maximum/minimum temperature be enforced. Compared with

conventional thermal cycling, the temperature of the assembly is "an-isothermal" where the substrate is typically cooler than the component. This leads to a range of stress distributions within the device that are difficult to predict and can be unexpected.

For devices where there is no CTE match between the component and the substrate, power cycling is likely to show similar results to thermal cycling. However in situations where components and substrates are of the same material, which is particularly important for Multi-Chip Modules (MCMs) composed of Si chips flip chip attached to a Si substrate, thermal cycling is unlikely to stress the device and therefore power cycling will be more representative, as has been demonstrated by Trigg & Corless (7). They thermally cycled silicon on silicon assemblies (hypothetically a good CTE tailored scenario) and found that after 500 cycles that there was only one failure, that was attributed to earlier rework of the chip. However, when the components were power cycled, a greater number of assemblies were observed to fail, implying a different stress distribution from the case when they were thermal cycled. It was anticipated that there was an anisothermal temperature distribution present as the chip may both heat up faster than the substrate chip (7) and reach a higher temperature; this effect may be enhanced depending on the type of substrate the assembly is mounted on and the external cooling conditions. Trigg and Corless have highlighted the potential hazards of being totally dependent on thermal cycling and highlights why power cycling may be a more reliable testing method.

1.4: Thesis Aims

The aim of this thesis is to investigate the effect of power cycling on the temperature, stress and reliability of a flip chip assembly. The work involved the manufacture of silicon on silicon multi-chip module assemblies that were power cycled to investigate their reliability and such that the temperatures of the different "parts" of the assembly could be seen/measured. The assemblies were attached to either an FR4 or a copper substrate such that the different thermal properties of the substrates were also investigated. FR4 is a conventional electronics substrate while the copper substrate represents a thermally enhanced assembly such as when a heat sink is fixed to a component. In order to control the external variables, the experiments were conducted in a wind tunnel such that the airflow could be regulated. Finite element modelling of the assembly was also carried out to understand the thermal effects and consequent stress distributions.

References

1) **J.H. Lau** "Low Cost Flip Chip Technologies" *McGraw-Hill, 1999, Chapter 1*
2) **J.H. Lau & S.W. Ricky Lee** "Chip Scale Packaging, Design, Materials, Process, Reliability and Applications" *McGraw-Hill, 1999, Chapter 1*
3) **S.G. Konosowski, & A.R. Helland** "Electronic Packaging of High Speed Circuitry" *Electronic Packaging and Interconnection Series McGraw-Hill, 1997*
4) **J.H. Lau** "Low Cost Flip Chip Technologies" *McGraw-Hill, 1999, Chapter 4*
5) **Peter M. Stipan, B. C. Beihoff, M.C. Shaw** "Electronics Package Reliability and Failure Analysis: A Micro Mechanics Approach" *The Electronic Packaging Handbook IEEE press, 2000, Chapter 15*
6) **Robert D Malucci** "Accelerated Testing of Tin Plated Copper Alloy Contacts" *IEEE Transactions on Components and Packaging Technology Volume 22, March 1999, pages 53-60*
7) **A.D. Trigg & A.R. Corless** "Thermal Performance and Reliability Aspects of Silicon Hybrid Multi-Chip Modules" 40th *Electronic Components and Technology Conference, 1990, pages 592-9*

Chapter 2: Literature Review

As outlined in the introduction, the area of electronics reliability is strongly affected by the thermal issues regarding an assembly. This is especially apparent in the area of flip chip technology where the CTE mismatch may cause excessive shear strain in the joints. These effects however, can be minimised by applying a cooling mechanism such as a fan or heat sink. This review starts off by analysing the more novel approaches to thermal management of high power devices, before looking at reliability testing (thermal/power cycling methods) in more detail.

2.1: Higher Power Flip Chip Assemblies

Prior to a discussion of issues regarding high power flip chip assemblies it is first necessary to define what is meant by "high power." Kranz (1) defines a high power device as one with a power density greater than 6.2 Watts/cm^2. Though clearly this is an arbitrary figure, it can serve as an approximate guide to differentiate between high and moderate power levels (e.g. much greater than 6.2 Watts/cm^2 may be confidently considered high and lesser values moderate). An obvious application area of high-powered flip chip is an alternative packaging method for computer microprocessors. Computer technology (especially microprocessors) are approaching the limits that conventional wire bonding may offer with respect to the increasing number of IO connections required from a given density. This increasing number corresponds with a smaller connection pitch required resulting in microprocessor manufacturers considering flip chip as viable alternative to packaged components (2, 3, 5).

General Application Areas

There have been some studies considering the possibility of flip chip packaging for microprocessors. For example Darnauer et al (2), considered the feasibility of flip chip packaging for the next generation of microprocessors by comparing low cost organic ball grid array against thin film ceramic land grid array. The feasibility study was conducted with respect to electrical noise control and specifically the study focused on two types of noise emissions, namely a) core noise, and b) simultaneous

switching output (SSO) noise. The noise measurements were recorded and compared against noise recorded from the conventional wire bonded packages and the results were analysed for each parameter separately. The results showed no significant reduction in core noise over conventional wire bonding technology, nonetheless, positioning a chip capacitor near the noise source may compensate for this. SSO emissions, on the other hand, had significantly improved over the previous generation microprocessor, with the resulting noise emission reduced by a factor of 2. In addition, the ceramic land grid array demonstrated superior performance to plastic ball grid array due to the lower impedance of its power plane.

This investigation is useful as it certainly gives credibility to the next generation microprocessor being flip chip based, though there are several crucial issues that have not been considered. Furthermore, it should be noted that Darnauer only considers the feasibility from an electrical perspective (that is to say "*is* the product capable of functioning adequately using flip chip technology?") and does not address the mechanical and thermal performance of such packages.

Thermal Management

There is limited published data on thermal management for high power flip chip assembly, however most methods involve thermally enhancing the package, and typically enhancing the substrate material. For most electronic applications, the utilised substrate material is FR4, which has a thermal expansion rate approximately 4 times that of silicon. In order to address the CTE mismatch issues highlighted, a potential substrate material should have a thermal expansion rate close to that of silicon, and good thermal conductivity.

Gilbert et al (3) illustrated the temperature reached by high power modules when they performed a Finite Element (FE) simulation of a signal processor for military applications. The module consisted of 16 GaAs chips 650μm thick mounted on a copper and polyimide flex 100μm thick mounted on a 100μm thick alumina substrate. There were two 6W power chips in the central region of the board and it was calculated that left untreated the chips would exhibit a thermal resistance of 7°C/W, resulting in a potential temperature rise of 42°C per chip. An array of metal vias was

inserted into the substrate such that each chip had a via running from the base of the chip vertically through the polyimide and into the substrate. The via was in contact with approximately 7% of the chip area and subsequent simulations indicated that the chips would have a thermal resistance of only 1°C/W. The thermal performance was greatly improved by the vias providing a highly conductive path between the chips and the substrate.

Traditional substrate materials for high power MCMs include approximately 1mm thick alumina or 600-650μm silicon wafers. However in more recent times, alternative materials including sapphire and metal matrix composites containing silicon carbide are becoming popular. According to Wu et al (4) sapphire is popular as it is available at reasonable costs and in large wafer sizes, whereas silicon carbide is desirable as it provides excellent thermal conductivity. In addition to this, the overall performance of the assembly can be improved by altering the die package.

Ulrich and Rajan (5) foresaw a novel approach to thermal management of high power flip chip assembly. Rather than provide a path for heat flow through the joints and the substrate and finally to a heat sink, their theory was to conduct as much heat from the die as possible therefore bypassing the interconnection layer completely. The only way this could be done was to immerse the chip in liquid nitrogen. Operating a component at these cryogenic temperatures offers many additional benefits, for instance the thermal conductivity of the silicon and copper are greatly increased and furthermore, the rate of many temperature dependent failure mechanisms, such as corrosion and electro-migration are greatly reduced therefore resulting in good package improvements. Initial tests (5) involved bare chips directly exposed to liquid nitrogen and the silicon yielded excellent thermal resistance (0.15°K/W), however it is undesirable to have bare chips in a microelectronic assembly from a reliability perspective. After a metal lid was placed on the test vehicle the thermal resistance rose to 4°K/W due to a gas gap between the lid and the chip; this gap increased the thermal resistance to 18°K/W resistance and forced 80% of the heat generated to flow through the solder bumps and the substrate.

2.2: Thermal Cycling and associated Phenomena

Thermal Cycling Description

The reliability of electronic components and assemblies is often verified by thermal cycling test vehicles. Traditionally a batch of specimens or test vehicles is manufactured, placed in a cycling chamber and physically heated and cooled between a designated maximum and minimum temperature, however it is becoming more common to perform virtual thermal cycling using numerical computer methods such as FE modelling. When an assembly is cycled there are two types of time periods: a ramp time and a *hold* or *dwell* time. The ramp time involves the controlled isothermal temperature increase to the maximum or decrease to the minimum temperature. The ramp rate is typically constant, e.g. 11°C/minute as shown in figure 2.2a, though the ramp rate is occasionally varied to provide specific conditions (6). The dwell time is the length of time the component is held at the designated maximum/minimum (e.g. a ramp and dwell time of 15 minutes may be implemented as shown in figure 2.2a). The temperature limits used vary depending on the solder used, the test vehicle requirements and its intended operating environment. Typical examples for 63/37SnPb solders are 0-100°C for commercial equipment operating in benign environments, and –40 to 125°C for those operating in hostile environments (7); it should be noted that lead free solders may have a significantly higher maximum temperature due to higher melting points. The trials normally run continuously for the required cycle period (20 000 cycles or 2½ years in 2.2a), though after a designated number of cycles (e.g. 500, 1000, 2000) some test vehicles may be removed from the trial and relevant physical analysis is performed on these test vehicles.

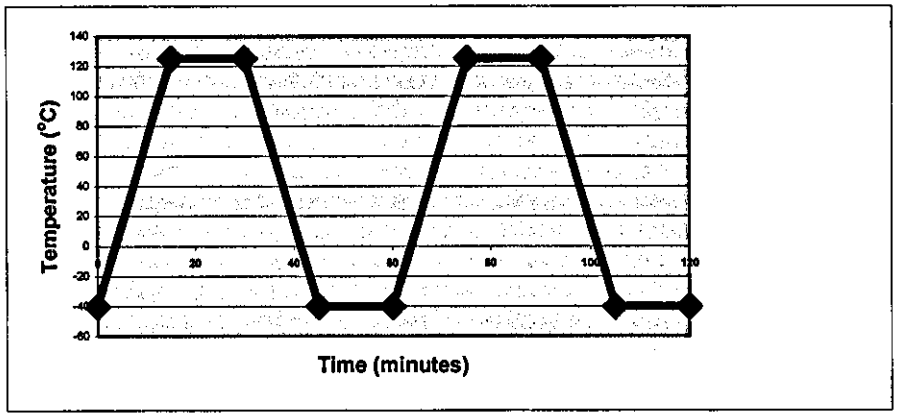


Figure 2.2a: *Example of a typical time temperature profile of two complete thermal cycles*

Stresses and Strains for Solder Joints

When a flip chip is thermal cycled, there is likely to be a shear stress produced in the solder joint resulting from the different CTE values of the die and substrate. Principally, mechanical materials subject to a tensile or shear force will exhibit some form of extension, the amount of which is governed by the material properties and may either be instantaneous as in the case of elastic or plastic deformation or it may be time dependent such as those as a consequence of creep influences (8,9).

Most material deformation data is obtained from manufacturing and deforming a test piece. Typically these are tensile test pieces that are manufactured in a "*dog-bone*" shape where the central section is of a fixed diameter and the larger ends are configured in a way to allow the test piece to be secured in a tensometer as shown in figure 2.2b_i. However, the typical dimensions and the nature of the stresses of the dog-bone test piece fail to accurately represent the shear stresses endured by a typical solder joint, so test pieces can also be manufactured from two metallic pieces positioned accordingly and joined by solder to form a lap joint as shown in figure 2.2b_{ii}. Once a large force has been applied to the test piece, it may then be examined for the relevant elastic and plastic strain data.

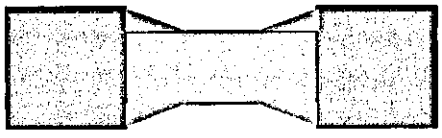


Figure 2.2.b_i:
A typical dog-bone specimen sample. Central region is of a fixed length and diameter. Criteria for ends depends on the nature of the "stress tensor"

Figure 2.2b_{ii}:
Example of the amended specimen test piece used to capture small amounts of solder and to simulate shear stress

The plastic deformation demonstrates the immediate effects large stresses have on a material, however materials may also fail if they are subject to stress levels below the yield stress if the applied stress is held for a significant time period; this phenomenon is known as creep. Unlike plastic deformation, which is fundamentally dependent on the magnitude of the applied stress load, accumulated **creep strains** show a strong dependence on the applied stress, the temperature of the material and the length of time the stress is applied for.

The significance of creep strain to a material is dependent on the materials homologous temperature; the homologous temperature refers to the absolute material temperature, T (°K) with respect to the melting point of the material or T_m (°K) and therefore has values between 0 and 1. At homologous temperatures less than $0.4T_m$, the creep strains for most metallic materials including tin based solders is known to increase logarithmically with time, hence total creep strains are typically very small and seldom result in failure. On the other hand, at temperatures above $0.4T_m$, creep strains are known to be much more prevalent and may well result in fracture if the stress is applied for a sufficient time period. The phenomenon of creep is particularly important for tin based solders as the Sn63-Pb37 eutectic solder has a homologous temperature of 0.65 at room temperature due to its low melting point; lead free solders despite the higher melting temperatures, may be $0.5T_m$ at room temperature. Figure 2.2c shows characteristic time/creep strain curves for a material subjected to stress at constant temperature above and below $0.4T_m$. It can be seen that for temperature ($T>0.4T_m$) creep strains there are three distinct regions of the curve, namely the primary, secondary and tertiary regions. The strain rate can be seen to reduce during the primary stage until it is constant throughout the secondary stage or steady state. Within the tertiary stage creep strain rapidly accumulates again and the rate increases until eventual failure. Most constitutive relationships such as the sinh law (commonly used) assume the secondary stage to be the most significant.

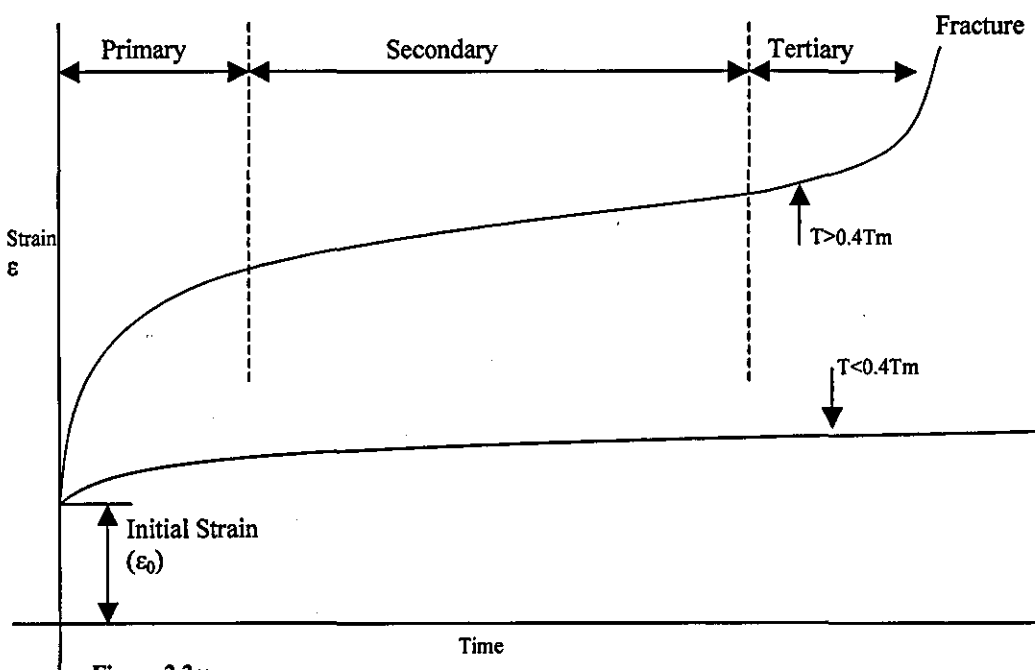


Figure 2.2c:
Schematic of how creep strain accumulates with time for a material at a low homologous temperature and a material at high homologous temperature.

Creep mechanisms in solder are known to occur in either one of two ways: **dislocation climb** refers to diffusion within the grain or crystal while **grain boundary sliding** refers to diffusion at the surface, either mechanism is dependent on the temperature and the nature of the applied stress. Dislocation climb has been identified when a large temperature change has occurred suddenly, or at low temperatures, while grain boundary sliding is associated with small temperature changes and warmer temperatures (9,10).

From observing a typical time/creep strain curve, there can be some ambiguity in determining the exact point when primary creep ends and secondary creep begins; the same argument may be made for the secondary and tertiary regions. Due to difficulty in characterising the primary creep (i.e. the rates and exact time length), it is often neglected from solder analysis. However referring to figure 2.2c, it can be seen that within the secondary region, the creep strain increases linearly with time. Within this region, the accumulated creep strain can be predicted for a given time period. Therefore it is the secondary creep region that is commonly used to model or characterise the creep in solder when testing electronic assemblies. Most relationships used are based on the power law, hence are of the form: -

$$\gamma' = A\tau^n \exp\left(\frac{-Q}{RT}\right)$$

Key
γ'= shear strain rate
A= constant
τ=shear stress
n= stress exponent
Q=Activation energy
R=Universal gas constant
T= absolute temperature

Ubiquitous solder testing methods such as thermal cycling have the relevant parameters modified such that the necessary conditions to assume steady state creep are implemented. Hence most thermal cycling regimes have long hold times and rapid temperature change is eliminated.

Grain boundary sliding is the prevalent creep mechanism for solder within the steady state creep stage, and the rapid increase in strain rates in the tertiary region are thought to be aggravated by the formation of cavities. Such cavities nucleate at points where typically three or more grains meet. Large and irregular grains may also increase the stress concentration at these junctions. Plastic strain may also occur within the tertiary region, however for the soft-low melting solder alloys this does not

result in failure as they can accommodate very large amounts of plastic strain within this region. The reason for this is that they are known to deform readily at a given stress level compared to their higher melting counterparts. Therefore these low melting alloys should be used when potentially large amounts of strain need to be accommodated as the higher melting alloys may stress the joints by not relaxing sufficiently (11-13).

Grossmann and Weber (10) investigated the effects of altering the ramp rate and the dwell time in thermal cycling on resultant crack growth on various SMT devices joined to a substrate using Sn62-Pb36-Ag2 solder. The control test conditions were a ramp rate of 80°C/min, with lower and upper limits of –20 and 110°C respectively. The standard dwell time was 30 minutes and the thermal cycle trials lasted for 4000 cycles for all cases. Following the control tests, the crack lengths were found to vary from 10% to 65% of the entire connection. The experiment was repeated with two different scenarios: one where the dwell time was reduced to 0 minutes and in the other, the ramp rate was reduced to 2°C/min. The dwell time was found to have the greatest reduction in crack growth, where all examined specimens had a crack growth of less than 5% of the entire joint. This effect was anticipated as most of the shear strains imposed on the solder joint were stored elastically and there was not sufficient time for the stress relaxation to occur. The slower ramp rate was found to marginally reduce the crack growth rate as the failure mechanism was thought change from dislocation climb (associated with fast ramp times and the more damaging of the two) to grain boundary sliding due to the slow ramp rate. Whilst altering the temperature would have been useful, it was accepted that reducing the minimum temperature would have offered little benefit due to the very low creep rates associated with solder at low homologous temperatures.

Solder Fatigue: Description and Relationships

The repeated application, removal and reversal of the stress on an object can lead to an eventual failure mechanism, namely fatigue which involves crack initiation and propagation resulting in material fracture after a number of stress cycles (8,14). It has always been desirable to have data detailing the number of stress reversals an assembly will last in service and therefore many relationships have been derived to

predict the number of cycles until failure. To begin with, solder strains were obtained experimentally utilising strain gauges attached to their devices, however with the ever decreasing profile of the packages combined with their increasing complexity it became necessary for computer simulations such as FE modelling to be used. The successful life prediction of an interconnection normally consists of choosing an appropriate constitutive relationship (i.e. one that describes how strain relates to stress, temperature and time) and implementing it in the computer simulation. The resultant strains can then be obtained and an appropriate model can be used to predict the number of cycles to failure. Once a value is obtained, the strain rates and reliability predictions can then be verified by thermal cycling manufactured devices to failure.

Failure Prediction Relationships

Lee et al (14) conducted a comprehensive review of the 14 major fatigue models used in solder joint reliability and separated them in to 5 different categories based on the fundamental failure mechanism used. They were a) stress based, b) plastic strain based, c) creep strain based, d) energy based and e) damage accumulation based. The prevalent methods used tend to be those that are strain based so the main focus shall be on plastic and creep strain based models.

Plastic Strain

Coffin Manson: This is possibly the most widely used fatigue prediction model. It assumes that the cycles to failure is fundamentally due to the plastic strain observed within a stress cycle, and is shown below.

$$\frac{\Delta\varepsilon_p}{2} = \varepsilon_f (2N_f)^c$$

Key
$\Delta\varepsilon_p$= Plastic strain amplitude
ε_f= Fatigue ductility coefficient (Typically Fracture coefficient)
c= Fatigue ductility exponent (typically 0.5<c<0.7) (12)
N_f= Cycles to failure

The constants are determined experimentally by obtaining the observed plastic deformation and correlating this with the number of cycles untill failure, however such tests are inevitably time consuming, and results are typically only applicable to a specific solder joint geometry. The largest criticism of the original Coffin Manson relationship is that it only assumes plastic deformation is responsible for joint failure. This led to the equation being modified by combining it with **Basquin's** stress dependent relationship such that the elastic strains are considered, therefore the elastic contributions are also accounted for and is shown below.

$$\frac{\Delta\varepsilon_{tot}}{2} = \frac{\sigma_f}{E}(2N_f)^b + \varepsilon_f(2N_f)^c$$

Key
$\Delta\varepsilon_{tot}$= Total strain amplitude
ε_{ff}= Fatigue ductility coefficient (typically true strain at fracture)
c= Fatigue ductility exponent (typically -0.7<c<-0.5)
σ_f= Fatigue strength coefficient (typically true stress at fracture)
b =Basquin's exponent (typically –0.12<b<-0.05) (12)
E= Young's modulus (Pa)
N_f= Cycles to failure

The combined characteristics are shown in figure 2.2d (12,14), where it can be seen that the during the low cycle range, the plastic strain is the dominant failure mechanism while for high cycles, the elastic strain is the main contributor.

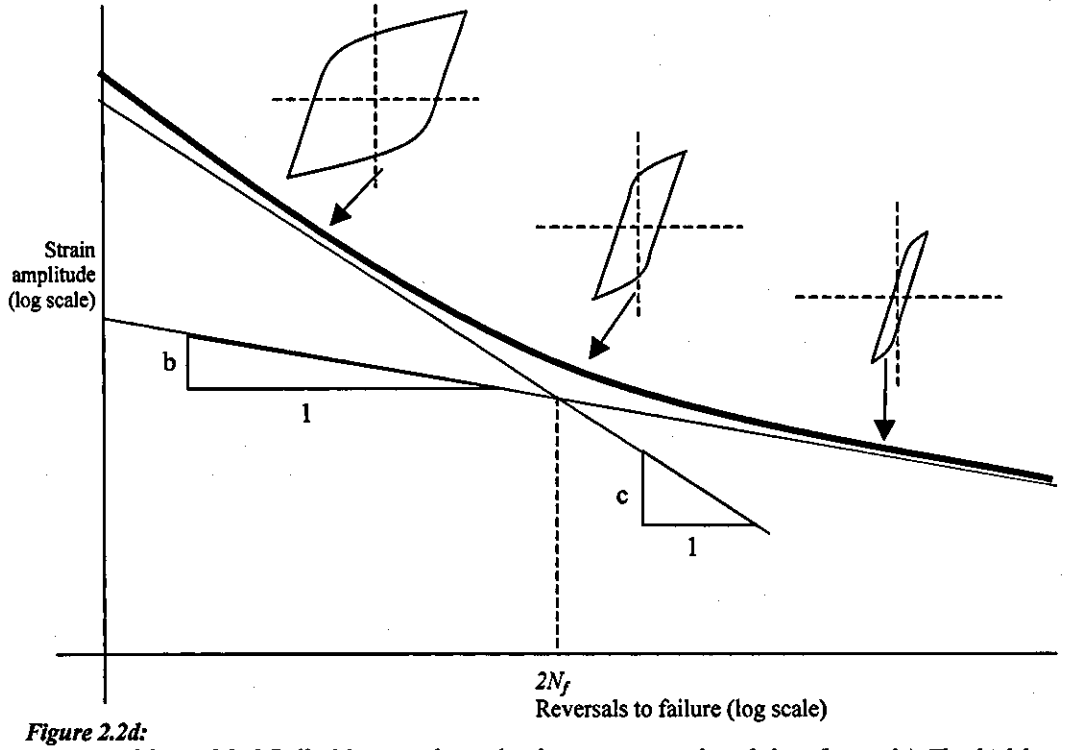


Figure 2.2d:
Diagram of the modified Coffin Manson relationship for stress reversals to failure (log scale). The thick line shows the total strains to failure, the red line shows the failure rates from the original Coffin Manson law (plastic deformation) while the green line shows those anticipated from Basquin's law. Also shown are the corresponding changes in the hysterresis loop and size (12,14). It can be seen that plastic deformation dominates early cycle failures while higher cycles tend to be influenced by elastic properties

Solomon's model relates plastic shear strain imposed on the specimen to low cycle fatigue prediction and the plastic strain range must be determined. However creep strain is not considered in the model so its use is limited to very short cycle times.

$$\Delta\gamma_p N_p^{\alpha} = \theta$$

Key
$\Delta\gamma_p$= Plastic shear strain range
N_p= Number of cycles to failure
α=materials constant
θ=inverse of the fatigue ductility coefficient

The **Engelmaier** fatigue model relates the total number of cycles to failure by relating the total shear strain and improves on Coffin Manson and Solomons relationships as it accounts for the cyclic frequency effects, the temperature effects and the elastic-plastic strain rates. However, it is based on isothermal fatigue data, and is only valid for specific criteria such as eutectic solders, hence cannot be used for lead free.

$$N_f = \frac{1}{2}\left[\frac{\Delta\gamma_t}{2\varepsilon_f}\right]^{\frac{1}{c}}$$

Key
$\Delta\gamma_t$= Total shear strain range
ε_f= Fatigue ductility coefficient $2\varepsilon_f \sim 0.65$ for underfilled flip chip (12)
N_f= Cycles to failure
c= $-0.442-(6\times10^{-4}\,T_s)+(1.74\times10^{-2})\ln(1+f)$
T_s= Mean cyclic solder temperature (°C) (in c constant)
f= cyclic frequency (cycles per day) (in c constant)

Creep Strains

As previously stated, creep may be separated into two different types: dislocation climb and grain boundary sliding. **Knecht** and **Fox** proposed a basic creep model based on the microstructure of the solder, and the dislocation climb shear strain range.

$$N_f = \frac{C}{\Delta\gamma_{DC}}$$

Key
C= a constant based on solder microstructure & failure criteria
$\Delta\gamma_{DC}$= Strain range due to dislocation climb creep
N_f= Cycles to failure

Syed created a model that accounted for both dislocation climb and grain boundary sliding where the equation was partitioned into two parts. The results published showed that the dominant failure mechanism changes from grain boundary sliding to dislocation climb when fast ramp times are used. One limitation of this model is the absence of the plastic deformation, of which Syed claims plastic deformation is not applicable to thermal cycling due to the high homologous temperatures and the low temperature gradients (i.e. the temperature change rates are very small in thermal

cycling such that there is only elastic strain developed and the subsequent creep strains).

$$N_f = \left([0.022 D_{GBS}] + [0.063 D_{DC}]\right)^{-1}$$

Key
D_{GBS}= Equivalent strain from grain boundary sliding
D_{dC}= Equivalent creep strain from dislocation climb
N_f= Cycles to failure

Discussion:

All the models have their relative merits and flaws, namely the models based on plastic strain tend to neglect the creep properties and vice versa. However, there have been attempts to merge two of the models that have neglected the two parts. For example **Miner** has managed to merge Solomons equation with that of Knecht and Fox whereby: -

$$\frac{1}{N_f} = \frac{1}{N_p} + \frac{1}{N_c}$$

where N_p and N_c refer to the lifetime predictions based on the plastic and creep components respectively. All the models considered require some form of previous experimental work and models such as the Coffin Manson equation require knowledge of the strain range, which is dependent on the solder geometry. The creep-based models were derived from the microstructure of solder; many of these were based on traditional specimen shaped samples, such as the "dog-bone" and may be several millimetres long. Therefore tensile stresses may not be a true representative of mechanisms present in flip chip assemblies due to the potential grain size to overall dimensions ratio. They do not account for impurities that may be present within the solder joint. Furthermore, given the small amount of solder present in a single flip chip joint, there may be very few actual grains present therefore influencing the grain boundary sliding models (12,14).

Other models for life prediction are based on energy levels and damage fatigue. The energy level models are based on the obtained data from the stress-strain hysteresis loops and offer the benefit of capturing the test conditions with greater accuracy.

However the models considered are only capable of predicting the number of cycles to crack initiation and could not capture crack propagation or failure. Damage models are based on either fracture or creep mechanisms and the models are based on the theoretical constitutive relationships, and are therefore strongly dependent on the created computer model.

Also there is the ambiguity in the definition of failure; failure may simply be cyclic strains that cause some permanent damage in the solder, it may alternatively be defined when a change in resistance is recorded, or the first sign of crack initiation or even total joint failure. Some of these mechanisms are extremely difficult to identify and track especially given the dimensions of flip chip solder joints. For example cracks are known to arise from microscopic inconsistencies, therefore pinpointing the exact time a crack initiates may be very difficult. This may be compensated by altering the failure criteria, for example a designated change in resistance measurements or total failure (open circuit).

2.3: Investigated Phenomena using Thermal Cycling

Thermal cycling of flip chip assemblies has been used to investigate the impact of intermetallic formations, CTE mismatches, die cracking, size limits of packages, the effects of underfill, and, especially following recent legislation, lead free solder properties. The following section is a summary of some of the typical studies that have focused on these topics through the use of thermal cycling.

Intermetallic Formation

When a tin based solder wets a solderable material, there is the inevitable formation of intermetallic compounds, which are necessary to form a substantial connection between the two materials. For example when 63/37 SnPb eutectic solder wets a copper surface, two intermetallics are formed: Cu_6Sn_5 (η phase) on the solder side and the Cu_3Sn (ε phase) on the copper side (15-17). When they are initially formed during the soldering process, the layers are very thin, however they are known to grow over time. The elevated temperatures of an electronic product in use are known to increase this growth rate, and if a product is *"active"* for a long enough period, then the intermetallics may eventually consume an entire joint. Repeated thermal cycling is an adequate method of accelerating the growth of these intermetallics (15-17). Pang et al (15) thermally cycled test vehicles consisting of a single SnPb eutectic joint connecting two copper pads each on a separate FR4 substrate as shown in figure 2.3b. The assemblies were thermal cycled as opposed to simple high temperature aging as these would simulate the thermal fluctuations of a "real" assembly. The test vehicle was designed in a way such that the shear effects resulting from CTE mismatches were negligible. The batch was thermal cycled and test vehicles were taken out after 500, 1000 and 2000 cycles. Along with control samples (i.e. those that had not been cycled) all of the test vehicles were shear tested and examined for microstructure coarsening and intermetallic growth. As anticipated, the examinations showed a growth in intermetallic thickness corresponding with an increasing coarse microstructure. It was found that the joints shear strength after 1000 cycles were reduced by 8 times that of the control specimens, and those subjected to 2000 cycles were reduced by a further factor of 6. These findings were in good agreement with

other intermetallic growth studies such as those of high temperature aging of a flip chip joint (16), the cycling of tin-plated copper contacts (17) and the generic studies of intermetallics (18,19).

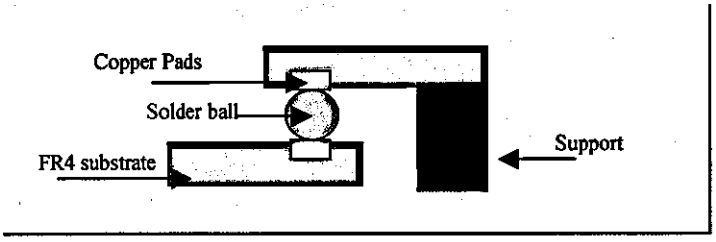


Figure 2.3b: *Illustration of specimen assembly as used*

The previously described experiment offers a useful insight as to how mechanisms influence the strength and reliability of a joint. However, it is noted that the geometry of a real assembly is neglected. It would be useful to investigate how the stress distribution varies with respect to the location of a joint and also how other factors such as die size influence joint fatigue.

Die Size

There are many failure mechanisms that contribute to product failure in addition to intermetallic growth. For example voids and die cracking are prominent failure modes throughout SMT electronics (20), however flip chip must now address new phenomena that may compromise the reliability of a product. For example with the fine pitch and high IO count required in many applications, there may be larger die sizes required. It is anticipated that by 2012 computer processing applications may require die sizes to be 40mm square (21).

This raises the question: is there a maximum die size for DCA assemblies? The issue must be addressed as the corner joints are under the greatest strain and there may be some fundamental limit of stress/strains a joint may endure. Schubert et al (21) examined the strains in 20x20, 30x30 and 40x40mm dies and compared these with those of the currently more common 10x10mm. This was performed using FEA, with all dies modelled as being on FR4 substrates, with solder ball heights of 90μm at a pitch of 500μm. The underfills, solder mask, copper pads, UBM and silicon were all modelled as linear elastic materials. The eutectic solder however, had to be modelled using the non-linear constitutive sinh law such that creep mechanics could be considered. Two types of simulations were performed, namely a clamped and non-

clamped board. In the clamped simulation, the board was not permitted to deform in the z axis, however, in the non-clamped board deformation was permitted. The simulations were thermally cycled a number of times (cycle number unspecified) and the results implied the following: though creep strains were largest in the bumps nearest to the die edges, they appeared to be independent of the actual Distance from Neutral Point (DNP), that is to say as the chip gets larger, the strain in the corner joints remains the same. It was concluded that larger dies will be more prone to failure, however, this was due to high IO count, resulting in a higher probability of joint failure. Applying the underfill is likely to be more complex, therefore defects such as voids are likely to require great attention.

Underfill and Voids

The concept of mechanically coupling the die to the substrate by means of an underfill was initially proposed by Nakano et al 1987 and there have been many subsequent investigations regarding its influence on solder joint fatigue. As previously stated, the role of underfill is to enhance the reliability of a flip chip package and this improvement was found to be optimum when the CTE was approximately that of the solder balls. Though this effect may reduce the thermal stresses there may be additional impact on the chip especially if voids are present.

Madenci et al (22) investigated the effects underfill had on the stress concentration therefore identifying possible failure sites using FEA. The experiment involved creating two models of DCA flip chip assemblies, with one model without underfill, and the other with underfill. Both models employed their purpose built Global-Local (GL) elements that combined the properties of the solder, underfill and pads. GL elements are elements that contained a combination of material properties at the relevant interfaces (i.e. chip and pad) such that the necessary model characteristics could be accommodated for. These elements were used for the case with underfill and versions with properties of only the solder and the substrate were used for the case without underfill. The GL elements were positioned at the corners of the outermost solder bump to capture the material properties of both the solder and pad, and they were used in the same manner for the underfilled model. As expected, the chip without underfill showed the potential cracking sites at the corner where the ball was

subjected to a tensile shear stress. Though also identified as potential cracking points on the chip with underfill, the critical region was noted as the point where underfill meets the die. The underfill was also noted to reduce the maximum stress concentration factor by half in any of the global elements within the model.

Doi et al (23) used FE thermal cycling to investigate the effects underfill had on chips with heat spreaders attached. The model was created with an alumina substrate as opposed to the traditional FR4. There were several parameters that were to be considered. The study focused primarily on two parameters: a) whether there was underfill present or not and b) whether the chip had a heat spreader. There were initial tests to determine which underfill would be appropriate. The underfill with the lowest CTE and the highest Young's modulus was chosen in accordance with a previous study by a different author (23). There were 4 models created with the control condition being without a heat spreader or underfill. Table 2.3 demonstrates the varying conditions tested.

Case	Underfill present	Heat spreader attached
1	no	no
2	yes	no
3	no	yes
4	yes	yes

***Table 2.3:** Summary of conditions used in the experiment by Doi (23)*

The outermost solder ball was then analysed for areas with the highest stress, as these are most vulnerable to crack propagation. The conditions without underfill (i.e. case 1 and 3) both showed the highest stresses, both at least a factor of four greater than those with underfill present. Both case 1 and 3 showed the areas under the most stress to be the outermost top corner. The cases where underfill was used, had the stresses moved to the side of the solder joint and the stresses were much lower and closer together. In addition, it was found that when a heat spreader was attached, delamination between the chip and the underfill was the more critical failure mode. The size and shape of the fillet influenced the small fillets, with a curved fillet having the optimum shape.

These results compare well with those obtained by Madenci (22) as they highlight identical stress areas for all chips without underfill, though Doi predicts that the presence of underfill will displace the stress distribution in the solder bump. Madenci does not stress such a prediction.

Niu & Sammakia (24) investigated types of voids that may exist once the underfill has been applied and the impact they have on product reliability. FE models of thermal cycling were used to investigate the effect of voids. The voids were grouped into four different categories: -

- **Case 1** Small voids (covering no more than 2 bumps)
- **Case 2** Medium voids (occupying maximum of 6 bumps)
- **Case 3** Large central voids (occupying all but the outermost (peripheral) bumps)
- **Case 4** Edge/peripheral voids (where the outermost edges of the voids are not covered by underfill)

The tests were compared against an ideal scenario (no voids present in the underfill), and the results from a previous test with no underfill (worst case). The results were then analysed for strain depending on whether they were small or large DNP. For small DNP (<6mm) with the exception of case 4, the measured strains were close to those of the void free module. However, for large DNP (>10mm), the strains in case 3 become larger than the edge void, which may compromise the reliability of the solder joints. The results therefore implied that a product should be acceptable even if voids of case 1 and 2 were detected on chips as large as 20mm.

These results give confidence in the role of underfill for a peripheral array chip, provided the underfill is adequate around the edges of the chip, thereby reducing the interconnect fatigue. However, underfill is known to also improve the temperature distribution of an assembly by increasing the volume that may be used for the thermal path, thus reducing the temperature difference between the die and the substrate. Consequently if voids are present in an assembly then they may impede the heat transfer from the die to the substrate and a greater temperature difference may exist therefore resulting in a greater CTE mismatch between the die and the substrate.

Furthermore, if case 3 (a large central void) was present on a sufficiently large die, it may cause the board to bow and therefore may be subjected to additional tensile stresses. It is therefore appropriate to examine the influence of stresses, where they originate from, and their impact on the assembly integrity.

Die Stresses

It was previously assumed that a flip chip is in its stress free state when the component is cool, however little consideration was given to the fact that the solder solidifies at a temperature significantly higher than the operating temperature. Hanna & Sitaraman (25) used FEA to investigate the role underfill contributes to chip assembly stresses. This study uses complex geometry parameters whereby the stress free temperature of the solder, substrate board and silicon die are the temperature the eutectic solder solidifies (183°C) therefore the assembly is considered to be in a stressed state at room temperature. The underfill is in its stress free state at its cure temperature (150°C). Two separate set-ups were considered: -

- **Chip 1**:was 6.35mm had 35 solder bumps at 381μm diameter at 762μm pitch.
- **Chip 2**:was 5.8mm had 441 solder bumps at 145μm diameter at 254μm pitch

All other variables were kept constant, i.e. stress measurement positions were identical for both chips.

Chip 1 was modelled with the viscoplastic models and creep behaviour of the solder neglected and this model had the stress values zeroed after reflow and the assembly was cured. The product was allowed to cool to room temperature and the stresses and warpage of the die were measured. Four different underfills were used, and the results showed that the lower CTE of the underfill participated as a factor in the reduction in compressive stresses at the die.

The model of Chip 2 included time dependent models for stress relaxation to allow underfill and solder to creep while at room temperature. The stress measurements for the die were zeroed at the reflow temperature. It was seen through simulations that it

only took 2 cycles for the characteristics to stabilise and once these had stabilised they showed a monotonic decrease in terms of warpage as the cycling progressed.

In contrast to traditional thermal stress studies, where the primary interest is the stresses in the solder joint, Michaelides & Sitaraman (26) investigated the effects of materials and geometries to the stresses and subsequent cracks in the die. Though the maximum stress a flawless die can sustain is large, it should be noted that silicon exhibits brittle failure mechanisms, therefore cracks and voids resulting from the manufacturing processes may compromise the die reliability. As the focus of this study was the die, the solder bumps were modelled as time independent to reduce the computational time and error margin. The parameters that were manipulated to see the impact on the die stress were: die thickness, substrate thickness, standoff height, and bump interconnect pitch.

The most significant parameter influencing the die stress was found to be the ratio of die and substrate thickness. The die size/thickness ratio was kept constant while the substrate thickness was varied resulting in ratios between 7:2 and 17:100 (die first) Figure 2.3b shows the schematics when the ratio is 3:1 and 1:3 and 1:1. The largest stress was found to be when the ratio was approximately 1, i.e. when the die thickness was approximately that of the substrate. As the thickness of the die was increased (i.e. the ratio reduced) the stresses were reduced. According to their study it is even possible to induce compressive stresses on the backside of the die therefore reducing the risk of cracking. In addition, as the substrate thickness is reduced, there was a notable decrease in stress though this is a more gradual change than if the substrate is increased. The results were verified by keeping the substrate constant and varying the die thickness. The effects of the stand-off height and the solder bump pitch were seen to have minimal effect on the die stress.

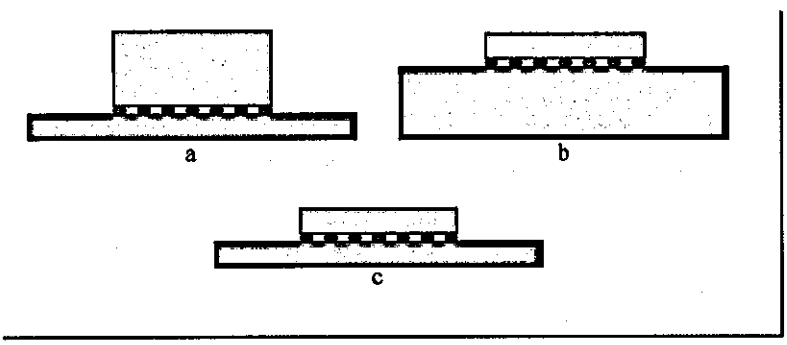


Figure 2.3b:
Examples of different die: substrate ratios used.
a) Ratio is 3:1
b) Ratio is 1:3(optimum)
c) Ratio is 1:1 (worst case)

NOT drawn do scale

The different die and substrate combinations were physically thermally cycled to obtain the first failure of a specific combination. Results varied from less than 100 cycles for a 5:8 ratio to over 500 cycles for a ratio of 1:3. It is therefore recommended that the die be made as low profile as possible.

Design Optimisation

As well as identifying and evaluating specific failure mechanisms, computer simulations of a thermal cycle can be used to optimise the design of an existing assembly by amending specific features and observing how fatigue life can be improved.

For example Tee et al (27) simulated the thermal cycle of an area array wafer level chip scale package of an assembly and subsequent lifetime predictions were made from the obtained non-elastic strains. Following this, specific features of the assembly were altered to investigate the subsequent effect on lifetime prediction. The parameters altered were: - a) the bump height, b) the bump diameter, c) the diameter of the UBM pad, d) the PCB pad diameter, e) the CTE of the PCB and f) the die thickness. The changes were first applied to each parameter individually such that the effects of changing a particular variable may be evaluated. The two most significant changes observed were increasing the bump standoff height and reducing the PCB CTE rate, improving the cycle lifetime by 17.1% and 22.7% respectively. The greater standoff height helps to reduce the shear strain in the joints, while lowering the CTE of the PCB towards that of silicon will inevitably reduce the shear strain rate endured by the joints. When the UBM diameter was increased and PCB pad sizes reduced individually they were found to improve the cycle time by 13.5% and 12.3% respectively, however when they were both implemented in the model, the predicted cycles to failure had improved by 44.9%. As the area prevalent for crack propagation is near the solder –UBM interface, it is logical to increase the diameter such that the crack takes longer to "grow" throughout the joint, while the smaller PCB pad helps to redistribute the higher stresses from the critical bump. After the individual assessments, all the optimised parameters were implemented in one model and the fatigue life improved from an initial 1014 cycles to 2167 cycles.

Stoyanov et al (28-31) coupled a computer simulation package (PHYSICA) with design optimisation software (VisualDOC) in order to optimise a flip chip design. The initial design was first entered and the **design variables** were modified using iterative techniques. Design variables are certain criteria that can be altered by the optimisation software; these may be material properties such as Young's Modulus/ CTE or they may be geometric data such as the standoff height. The variables were altered in accordance with the **Design Constraints**, which are typically the upper/lower limits a particular value may be, based on the available/possible materials. The aim was to optimise the **objective function** or the final outcome. For the case of flip chip assemblies the objective was to minimise the creep strain, hence maximising the cycles to failure.

Two types of iteration processes were used for the design optimisation, namely the results from a *control* Direct Design Optimisation approach were compared against those from an *experimental* Design of Experiments technique (28-31). **Direct Design Optimisation** (DDO) is a numerical technique that is reliant on improving existing designs by iterative means; the process starts with VisualDOC submitting the original parameters (design variables) to PHYSICA, then obtaining the solution and checking the results. Following this updated values are subsequently submitted to PHYSICA for analysis. As DDO is a first order (gradient descent) technique, VisualDOC uses gradients obtained from the solver to specify in which direction the relevant design variables should be changed (i.e. should the updated variable be increased or decreased). It is limited as VisualDOC typically has to submit numerous sets of design variables to PHYSICA, where analyses must be run; hence it is impractical for use with large models with several design variables. Whereas DDO first starts with a single set of values (design variables) the **Design of Experiment** (DOE) method involves VisualDOC submitting several different sets of values to PHYSICA where FEA is performed independently for each set. The number of different sets and their values are determined by implementing a relevant design algorithm. The response data is then returned to VisualDOC where the data can be used to generate approximations or Response Surfaces (RS); these are mathematical formulae for each response considered. VisualDOC solves the given response surfaces by using a simple numerical technique thereby saving several different submissions to PHYSICA

having to be made. The final set of design variables can then be used to evaluate the optimum objective function.

The material properties altered in this study were the underfills Young's modulus and the CTE, while the geometric parameters were the standoff height, the die height and the substrate thickness. In (29), both the iteration techniques were used to optimise the fatigue life of the solder joints. The cycle life of the initial specification and those obtained from the final optimised design are shown in table 2.3b, also shown are the number of cycles needed to achieve the optimised values. It can be seen that the number of analyses using both techniques increased the life time predictions to almost identical values, however the DOE method optimised the design in less than half the number of analyses needed for the original DDO method, therefore saving the resources needed to perform several analyses.

Variables	***Initial Design***	***Direct***	***Design of Experiments***
Number of Cycles (life time estimate)	4540	19744	19742
Number of analyses needed to optimise the design	N/A	65	27

Table 2.3b: *Original and optimised lifetime predictions (cycles to failure) and the number of times the analysis was run*

The studies of Stoyanov et al are useful as the parameters that produced the optimum results can then be passed on to PCB manufacturing/assembly plants as design recommendations. In addition to the reduced number of calls to the FE program, the DOE approach also offers the benefit of allowing sensitivity analysis to be performed. This allows the user to see the impact that changing one design variable has on the objective function. While the study of Tee et al may not thoroughly optimise a potential design in the way of Stoyanov et al (28-31), Tees study does offer some basic sensitivity information similar to that of the DOE approach. Furthermore Tee et al, highlights the combined effect of optimising two or more features, such as the pad sizes where the two optimised pads were found to have a much greater impact than they had individually. This sensitivity data may be useful for electronic companies as the efforts to change the specification of all the manufacturing equipment may be prohibitive (in terms of costs and loss in production time) but to implement some of the changes may be feasible.

2.4: Power cycling

Physically heating up and cooling down an assembly as described in the thermal cycling section may seem like an adequate method of obtaining the stress data from a flip chip assembly but what exactly does it achieve? Surely such stresses would be a mis-representation of the actual thermo-mechanical properties of an in service package where both the heat source and path(s) are likely to produce a substantially different thermal (and subsequent stress) distribution from those of conventional thermal cycling. In many cases, results obtained from thermal cycling may be misleading with inaccurate stress predictions (32-4).

An excellent example of such a misrepresentation can be found in the parametric tests of Trigg & Corless (32). They compared the thermal and power cycling trials of a silicon on silicon MCM. Theoretically this was the best CTE scenario and indeed, there were no genuine failures obtained during the thermal cycling trials. However, when the assemblies were power cycled, there were found to be several failures observed within the same time period. During the power cycling the die was likely to have endured a higher and more rapid temperature change than the substrate. Furthermore, these effects may have been further enhanced if the substrate was attached to a heat sink or if some other cooling mechanism was implemented, also, considering the probable size/mass difference between the die and substrate there is always likely to be a significant temperature difference between the two and therefore, a shear stress in the joints.

Another parametric study was conducted by Towashiraporn et al (33), where area array dies were mounted on FR4 substrates, and the finished assemblies were both thermal and power cycled. Both trials were configured such that the die was subjected to a 100°C temperature change regardless of whether it was thermal or power cycled. The power cycled assemblies endured an ON transient temperature increase 15 times faster than that of the equivalent ON ramp rate during the thermal cycle trial as well as exhibiting the anisothermal temperature distribution associated with power cycle trials. However, in contrast to the results of Trigg & Corless, the results from the power cycle trials implied better reliability than for the thermal cycling trials. The

contradiction in this case may be due to the isothermal conditions of the thermal cycle and the comparatively large CTE of FR4 compared to silicon; during the power cycling trials the substrate was significantly cooler than the die, so it expanded less and therefore a smaller shear stress was generated in the joints. They also conducted 3D FE models of these studies and were found to be in good agreement with the experimental results.

In addition to providing a more accurate stress distribution, power cycling may offer additional information as to the manner in which an assembly is likely to heat up/ cool down in service. For example Sur and Turlik (34) performed power cycling trials on a silicon die mounted on an aluminium nitride substrate, where the temperature of each body was monitored. The results revealed the behaviour of the components during the power ON stage, where it was observed that both the die and the heat sink would heat up very quickly, while the substrate and interconnection layer would heat up more slowly. Upon cooling down the heat sink cooled the fastest, while the substrate took the longest to cool. It was found that on power-up the assembly reached steady state after only 20 seconds but the cooling stage took much longer. Subsequent FE models characterised the transient behaviour and it was found that for the first three seconds the solder joints endured a shear stress in the opposite direction to that anticipated from conventional thermal cycling studies, before the substrate heats up and the larger CTE of the substrate sheared the joint in the *"correct"* direction. This phenomenon was attributed to the time required for the heat to pass through the interconnection layer and reach the substrate due to its low conductivity. The simulation provided a better representation of the transient behaviour a flip chip is likely to be subjected to in real life and many of these observations are undetectable in traditional thermal cycling.

A further advantage of power cycling is that it offers potentially large savings on the long test times needed to perform an adequate number of thermal cycles. These long test durations are costly to manufacturers who are required to perform such cycles to verify their product before it is commercially released. Lenkkeri and Jaakola (35) investigated whether power cycling is a viable alternative to thermal cycling. The cycle time may be reduced to as low as one minute, therefore the number of cycles per day could potentially be as large as 1440; the equivalent number of cycles may take

60 days when using traditional thermal cycling (based on the thermal profile as shown in figure 2.2a). They also found that by altering the power level that the temperature amplitude can be varied to an extent.

In addition, this allows the stages or the nature of a failure to be analysed. Zhiang and Baldwin (36) power cycled area array chips that were mounted onto an organic substrate where both individual bumps and complete daisy chains were monitored by measuring the resistance changes. They identified 4 stages of joint failure: a) gradual resistance growth- the resistance of a joint increases from its initial measurement; b) stabilisation, where the bump resistance stops increasing and stabilises for a period; c) intermittent failure, characterised by when a joint fails at a high temperature, though continuity is detected at the low temperature (characterised by spikes or sudden jumps in resistance readings) and; d) permanent failure when an open circuit is detected.

As power cycling is known to stress joints in a different manner from thermal cycling, there may be different failure mechanisms apparent. Roesch and Jittinorasett (37) power cycled gallium arsenide chips on a laminate substrate flip chip bonded with copper plated bumps on both the chip and substrate UBMs adhered with solder and they noted two types of failure in the joints. The first was cracking along the interface between the UBM and copper bumps typically associated with their bump shear tests. The second failure was solder fatigue associated with traditional thermal cycling results, however in contrast to crack propagation in the middle of the joint, cracks were observed in the intermetallic compound.

In addition to the stresses induced by thermal characteristics, power cycling can also be used to investigate high power levels where void nucleation may be instigated by electro-migration. Liu and Irving (38) investigated this effect by power cycling two FE models of the same device. The control model considered only the thermal influences while the second model had the electro migration criteria implemented. They were left to complete 4 cycles and the results were analysed. It was found that the electro migration model had stress levels double those of the thermal model, and the total fatigue life was therefore quartered.

Though power cycling is a perfectly viable reliability testing method, conventional thermal cycling is still very much a preferred means of proving a product, as power cycling has its associated limitations; although there has been an increase in interest in the past 5 years (33,35-39). For example, power cycling is commonly conducted in a natural environment where the ambient temperature may vary significantly on a daily basis. Thermal cycling has the temperature boundaries specified for a time period, allowing for repeatability and hence greater compatibility with alternative verification methods such as FE simulations. Furthermore, many of the lifetime prediction methods used are based on an isothermal profile (39) as is the case when a component is thermal cycled. While the number of studies primarily focused on power cycling is somewhat limited, it is still often used as a means of verifying other thermal cycling results. For example Barton et al (40) used power cycling and thermal cycling along with several other reliability tests to prove their novel MCM design.

While the number of power cycling trials reported are small compared to the number of those utilising thermal cycling, there is even less work reported that uses FE modelling as a means to simulate power cycling. Sur and Turlik (34) simulated a single power cycle of their assembly using FE models. However, they obtained the device temperatures from measurements of their "real" devices and these temperatures were then entered into the model as boundary conditions such that the thermo-mechanical data could be obtained. They did not however model the initial power generation in the chip with the convection modes used, so their model was adequate for obtaining resultant stresses of a thermal model, but could not be used to obtain a thermal profile of an assembly from first principles. For example, using an FE model to indicate how a given set of power and cooling parameters can result in a specific temperature/stress distribution. Towashiraporn et al (33) successfully modelled the area array examples they power cycled, where the heat was generated inside the die and a convection heat transfer mechanism was implemented with several assumptions. Their model accurately predicted the maximum temperature of the die, but failed to correctly characterise the transient profile. Likewise Ham et al (41) performed FE simulations of an assembly that had been tested by experimental power cycling. The FE simulation was performed first as a steady state thermal analysis and then a second stress analysis was run. The steady state analysis featured the heat generated internally and a temperature dependent convection coefficient was

used. After the thermal profile was obtained the thermal profile was entered into the stress model in the same way as Sur and Turlik. However, while this model may characterise the stress profiles once the assembly reaches steady state, it completely omits the transient behaviour of an assembly, including the transient stress profile.

There has been little work published regarding the influences of material properties on the thermo-mechanical profiles of a particular assembly during power cycling. Lenkkeri and Jaakola (35) used different materials to investigate the resultant temperature differences, though it appears their objective was to investigate how a given substrate may create a different cycling condition, and were less concerned with how it may influence the properties of an assembly subjected to a specific power level.

The ethos of FE is that it allows potential amendments to an existing design to be analysed without the task of actually building the model. Finite element analysis has been used to some degree in power cycling studies though its use has been restricted to evaluating a model that has been built, i.e. all of the examples from the literature required the assembly to be physically built such that necessary thermal details could be obtained. There has not been an FE model capable of obtaining the temperature profile of a flip chip assembly for a given power level from first principles (they all required an assembly be manufactured and thermally profiled first) and the additional effects of various material properties have not been fully investigated.

2.5: Conclusion

The work presented so far has evaluated the scope of research into flip chip reliability. There is much published work investigating the effects of phenomena such as intermetallic growth, voids and the role of underfill. The recent practices of assessing the nature of these failure mechanisms appears to be favouring virtual "Finite Element" simulations as opposed to practical trials saving on the long times needed to cycle a batch of assemblies.

Thermal cycling has been the favoured method of evaluating component reliability though there has been a recent burst of interest in power cycling that is believed to offer more realistic operating conditions and the additional benefit of shorter trial periods. The published work of accurate computer modelling of power cycling work is somewhat limited, and the work published so far, in one way or another has a strong dependence on "real" power cycling trials.

References

1) **Robert Kranz** "Thermal Interface Materials for High Power BGA" *Advanced Packaging Penwell, April 2000, pages 29-32*

2) **Joel Darnauer, Dave Chegson, Bill Schmit, Ed Priest, David Hanson and William Petefish** "Electrical Evaluation of Flip Chip Packages for Next Generation Microprocessors" *IEEE Tranactions on Advanced Packaging Volume 22 No 3, August 1999, pages 407-15*

3) Barry K Gilbert and George W Pan " Packaging of GaAs Signal Processors on Multi-Chip Modules" *IEEE Transactions on Components Hybrids and Manufacturing Technology Volume 15 no 1, February 1992, pages 15-28*

4) Yi Feng Wu, David Kapolnek, James Ibertson, Primit, Karikh, Bernd P. Keller, and Umesh K Mishra "Very High Power Density AlGaN/GaN HEMTs" IEEE *Transactions on Electronic Devices Volume 48 no 2, March 2001, pages 586-90*

5) **Richard K Ulrich and Sanjay Rajan** "Thermal Performance of a MCM Flip Chip Assembly in Liquid Nitrogen" *IEEE Transactions on Components, Packaging and Manufacturing Technology Part A Volume 19 No 4, December 1996, pages 451-7*

6) **Seong-ming Lee and Donald S Stone** "Grain Boundary Sliding in Surface Mount Solders during Thermal Cycling" *IEEE Transactions on Components, Hybrids and Manufacturing Technology CHMP-14, September1991, pages 628-32*

7) **R. David Greke, Gary B Kromann,** member IEEE "Solder Joint Reliability of High IO Ceramic Ball Grid Arrays and Ceramic Quad Flat Packs in Computer Environments: The PowerPC 603tm and PowerPC 604tm microprocessors" *IEEE Transaction on Components and Packaging Technology Volume 22 No 4, December 1999, pages 488-96*

8) **Peter M Stipan, Bruce C. Beihoff and Michael C. Shaw** "Electronics Package Reliability and Failure Analysis: A Micro Mechanics Approach" *The Electronic Packaging Handbook, IEEE press, 2000, Chapter 15*

9) **R.W. Evans and B Wiltshire** "Introduction to Creep" *The Institute of Materials, 1993*

10) **Gunther Grossmann and Ludger Weber** "Metallurgical Considerations for Accelerated Testing of Soft Solder Joints" *IEEE Transactions on Components Packaging and Manufacturing Technology Part C Volume 20 No 3, July 1997, pages 213-8*

11) **Darrel Frear, Harold Morgan, Steven Burchett, John Lau** "The Mechanics of Solder Alloy Interconnects" *Van Nostrand Reinhold New York (1994)*

12) **Milton Ohring** "Reliability and Failure of Electronic Materials and Devices" *Academic Press, 1998, Chapter 9*

13) **John Lau (editor)** "Solder Joint Reliability, Theory and Applications" *Van Nostrand Reinhold New York, 1991*

14) **W. W. Lee, L. T Nguyen, G.S. Selvaduray** "Solder Joint Fatigue Models: - Review and Applicability to Chip Scale Packages" *Microelectronics Reliability 40, 2000, pages 231-44*

15) **John H. Pang, H.T. Tan, X. Shi, Z.P. Wang** "Thermal Cycling Aging Effects on Microstructual and Mechanical Properties of a Single PBGA Solder Joint Specimen" *IEEE Transactions on Components and Packaging Technologies Volume 24 No 1, March 2001, Pages 10-15.*

16) **J. H. Okura, A Dasgupta** "Effect of High Temperature Aging on Joint Degradation in FCOB Assemblies" *EEP Volume 26-2, Advances in Electronic Packaging Volume 2 ASME, 1999, pages 150 –57*

17) **Robert D. Mallucci** "Accelerated Testing of Tin-plated Copper Alloy Contacts" *IEEE Transactions on Components and Packaging Technology Volume 22, March 1999, pages 53-60*

18) **Ralph W Woodgate** "The Handbook of Machine Soldering" *John Wiley and sons, 1996*

19) **Anthony J Bilotta** "Connections in Electronic Assemblies" *Marcel Dekker inc, 1985*

20) **Peter M Stipan, P.C. Beilhoff, M.C. Shaw** "Electronics Package Reliability and Failure Analysis A Micro-Mechanics Approach" *The Electronic Packaging Handbook IEEE press, 2000, Chapter 15*

21) **Schubert, R Dudek, R Leutenbauer, R Doring, H Opperman, B Michel, H Reichi, D Baldwin, J Qu, S Sitaraman, M Swaminathan, C.P. Wong, R Tummala** "Do Chip Size Limits Exist for DCA?" *1999 International Symposium on Advanced Package Materials, 1999, pages 150-57*

22) **E. Madenci, S Shkaeayev and R Mahajan,** "Potential Failure Sites in a Flip Chip Package With and Without Underfill" *Journal of Electronic Packaging Volume 120, December 1998, pages 336-341*

23) **H. Doi, K Kawano, A Yasukawa, T Sato** "Reliability of Underfill Encapsulated Flip Chips with Heat Spreaders" *Journal of Electronic Packaging Volume 120, December 1998, pages 322-8*

24) **Tyan-Min Niu, Bahgat G Sammakia (fellow IEEE) Sanjeev Sathe** "Void Effect Modelling of Flip Chip Encapsulation on Ceramic Substrate" *IEEE Transactions on Components and Packaging Technology Volume 22 No 4, December 1999, pages 484-7*

25) **Carlton E Hanna, Suresh K Sitaraman** "Role of Underfill Materials and Thermal Cycling on Die Stresses" *EEP-Vol 26-1 Advances in Electronic Packaging Volume 1 ASME, 1999, pages 795-801*

26) **Stylianos Micaelides, Suresh Sitaraman** "Die Cracking and Reliable Die Design for Flip Chip Assemblies" *IEEE Transactions on Advanced Packaging Volume 22 No 4, November 1999, pages 602-13*

27) **Tong Yang Tee, Hun Shen Ng, and Zhaowei Zhong** "Design Optimisation of Wafer Level CSP Solder Joint Reliability" *EMAP2003 Proceedings 5th International Conference on Electronic Materials and Packaging Singapore, November 17-19 2003, pages 184-9*

28) **S, Stoyanov, C. Bailey, H Lu and M Cross** "Solder Joint Reliability Optimisation" *APACK 2001 Conference on Advances in Packaging, Singapore, 2001*

29) **S. Stoyanov, C. Bailey and H Lu** "Optimisation Tools for Flip Chip Design" *Proceedings of IPACK2001, The pacific RIM/ASME International Electronic Packaging Technical Conference and Exhibition Hawaii USA, Jul 8-13 2001, pages 1-10*

30) **S. Stoyanov, C. Bailey and M Cross** "Optimisation Modelling for Flip Chip Solder Joint Reliablilty" *Soldering & Surface Mount Technology (Emerald), 14th January 2002, pages 49-58*

31) **S Stoyanov and C Bailey** "Optimisation and Finite Element Analysis for Reliable Electronic Packaging" *4th International Conference on Thermal and Mechanical Simulation and Experiments in Micro-Electronics and Micro-systems (EuroSIME 2003), April 2003, pages 391-98*

32) **A.D. Trigg & A.R. Corless** "Thermal Performance and Reliability Aspects of Silicon Hybrid Multi-Chip Modules" *40th Electronic Components and Technology Conference, 1990, pages 592-9*

33) **P. Towashiraporn, G. Subbarayan, B. McIlvanie, B.C. Hunter, D. Love, B Sullivan** "Predictive Reliability Models through Validated Correlation between Power Cycling and Thermal Cycling Accelerated Life Tests" *Soldering and Surface Mount Technology, 14 March 2002, pages 51-60*

34) **Biswajit Sur and Iwona Turlik** "Power Cycling and Stress Variation in a Multi-Chip Module" *IEEE Transactions on Components Packaging and Manufacturing Technology part B Volume 10 no 2, May 1995, pages 388-395*

35) **Jaakko Lenkkeri and Tuomo Jaakola,** "Rapid Power Cycling of Flip Chip and CSP Components on Ceramic Substrate" *Microelectronics Reliability 41, May 2001, pages 661-8*

36) **Jian Zhang and Daniel F Baldwin** "Reliability Assessment of Flip Chip on Organic Board using Power Cycling Techniques" *52nd IEEE Electronic Components and Technology Conference, 28-31 May 2002, pages 1402-10*

37) **William J. Roesch and Suwanna Jittinorasett** "Cycling Copper Flip Chip Interconnects" *Microelectronics Reliability 44, July 2004, pages 1047-1054*

38) **Youn Liu and Scott Irving** "Power Cycling of an IC Package Considering Electro-Migration and Thermal Mechanical Failure" *53rd Electronic Components and Technology Conference, 2003, pages 415-21*

39) **Quan Qi** "Reliability Studies of Two Flip Chip BGA Packages using Power Cycling Test" *Microelectronics Reliability 41, April 2001, pages 553-62*

40) **J Barton, G. McCarthy, R. Doyle, K Delaney, E. Cabruja, M Lozano, A Collado, and J Santander** "Reliability Evaluation of a Silicon on Silicon MCM- D Package" *Microelectronics Reliability 41, June 2001, pages 887-99*

41) **S.J. Ham, M.S. Cho and S.B. Lee.** "Thermal Deformations of CSP Assembly during Temperature and Power Cycling" *International Symposium on Electronic Materials and Packaging, 2000, pages 350-57*

Chapter 3: Manufacture of Si on Si MCM assemblies

3.1: Introduction

3.1.1: Background

Typical failure mechanisms for a flip chip module have previously been identified as a result of fatigue in the solder joints. This fatigue is a direct consequence of the Coefficient of Thermal Expansion (CTE) mismatch between the chip and the substrate; they may heat up and cool down at different rates and this causes a shear stress in the small interconnecting solder joints eventually causing the joint fracture. The classic ways in which these mechanisms are identified and verified is by thermally cycling components as described in the literature review.

Theoretically, this phenomenon can be minimised, if not completely eliminated by matching the relevant thermal properties of the chip and the substrate. One such way to do this is to have identical materials for both the chip and substrate. Trigg & Corless (1) thermally cycled silicon on silicon assemblies (hypothetically a good CTE tailored scenario) and found that after many cycles that there was only one failure, which was attributed to earlier rework of the chip.

While matching the CTE of the chip and substrate may prove satisfactory for the case with thermal cycling this is not necessarily the case with power cycling, as an anisothermal temperature profile may exist. This anisothermal temperature distribution may be present as the chip may both heat up faster than the substrate chip (1) and reach a higher temperature This effect may be enhanced depending on the type of substrate the assembly is mounted on and the cooling conditions.

The aim of this part of the work was to investigate to what extent the fatigue failure mechanism would be present in an MCM assembly subject to power cycling conditions and to investigate what parameters may influence this mechanism. (e.g. power input and substrate material.) This section details the chips and substrates used in the MCM, how the chips were prepared for the final assembly and how the manufacturing method was refined.

3.2: Component Description

3.2.1: Multi-Chip Module Description (a general overview):-

The MCMs used in this study consisted of ASH0034 "heater" chips and ASH0045 "carrier" chips originally manufactured by **GEC Plessey** semiconductors. These devices were assembled and subsequently tested using conditions equivalent to those modeled in the FE simulations. Both the heater and carrier chips had their pads capped with solder by being dipped in a solder bath, as opposed to the conventional wafer bumping methods, resulting in a small volume of solder deposited. The MCM was assembled by placing the heater chip on top of the carrier chip which was then passed through a reflow oven with a suitable profile. Schematics of the carrier and heater chips used for the power cycling are shown in figure 3.2.1. In addition, a brief description of the various paths and connections within the MCM is presented in table 3.2.1, while the overall route for assembly of the heater onto the carrier chips is summarised in figure 3.2.2a.

Path name/Number.	*Typical Resistance*	*Purpose*
Thick	27Ω	Provides power to the heater, hence allowing the assembly to increase in temperature
1	4.3Ω	One daisy chain to be continuously monitored
2	4.1Ω	Second daisy chain to be continuously monitored
3-4	N/A	Aluminium tracks connected for 4-point resistance measurement. This provides an objective measurement to the temperature of the component independent of thermocouples and/or thermal imaging.
5-6	N/A	Second aluminium track for alternative 4-point resistance measurement.

***Table 3.2.1:** Description of the MCM paths, resistance values and purpose*

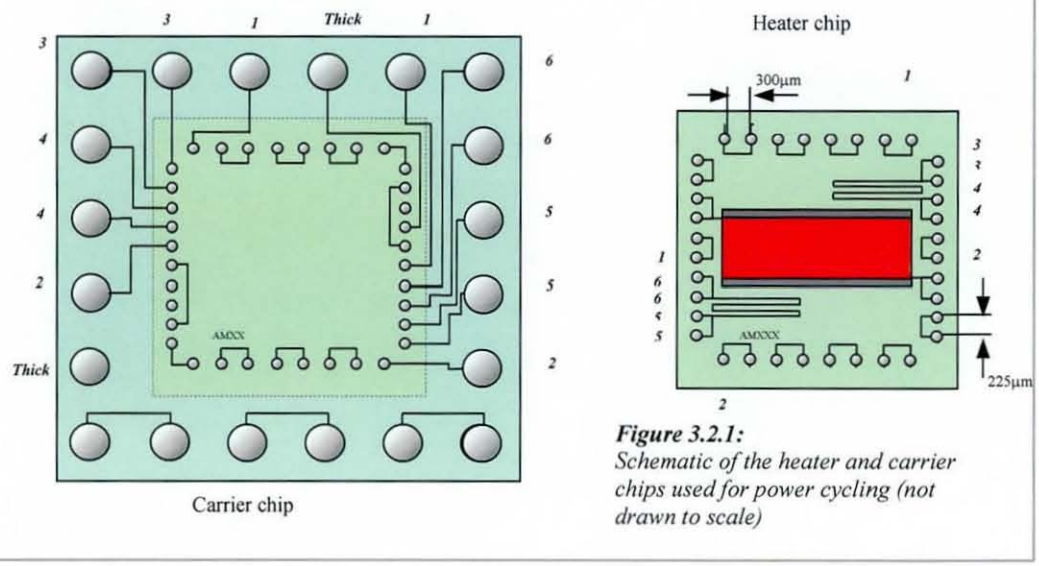


Figure 3.2.1:
Schematic of the heater and carrier chips used for power cycling (not drawn to scale)

3.2.2: Interconnection Pads

The interconnection pattern on the heater chip consisted of 36 aluminium bond pads in a peripheral array and the carrier chips had a matching pattern. The aluminium bond-pads were 1µm thick and were octagonal in shape with a circular opening in the silicon nitride passivation area that was 75µm in diameter as depicted in figure 3.2.2b. The pads on the heater chip were electroless nickel bumped to 15µm while those on the carrier chip were bumped to 5µm high in accordance with (2). They were then dipped in molten solder such that the pads were spherically capped with solder as demonstrated in figure 3.2.2.c.

The pitch of the bond-pads was 225µm pitch on sides of the pattern that included the connections to 4-point resistance measurements and the heater, whereas a 300µm pitch was used on the edges providing the majority of the daisy chains. Figures 3.2.2.d and e demonstrate the different pitches of the pads of the interconnection layer.

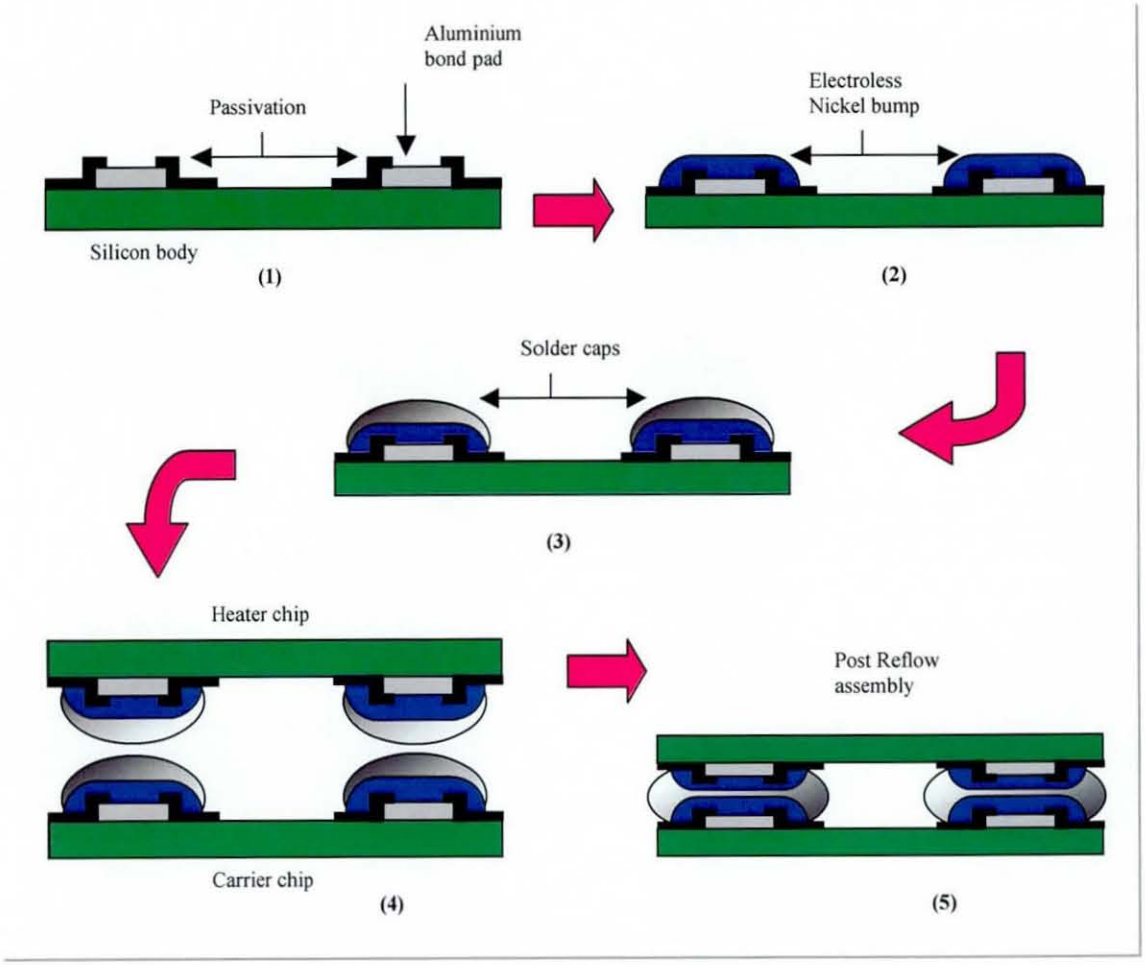
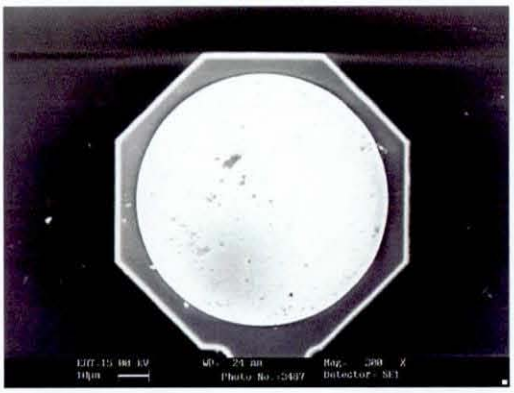
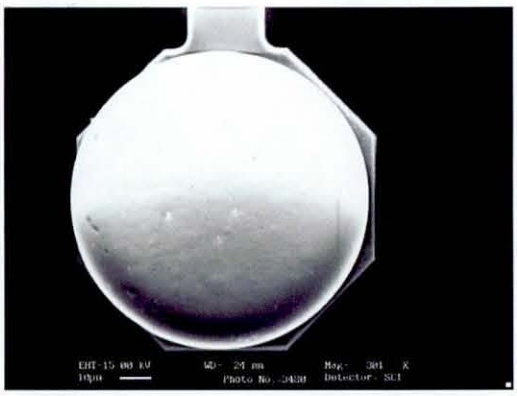


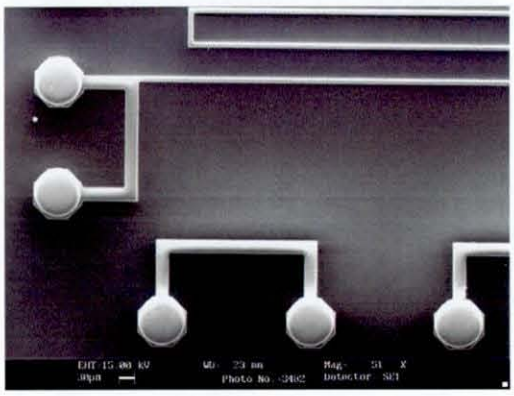
Figure 3.2.2a:
Overall schematic of MCM preparation; shown is bare die (1) the electroless nickel bump (2), the solder caps (3) and finally the placement and reflow (4&5)



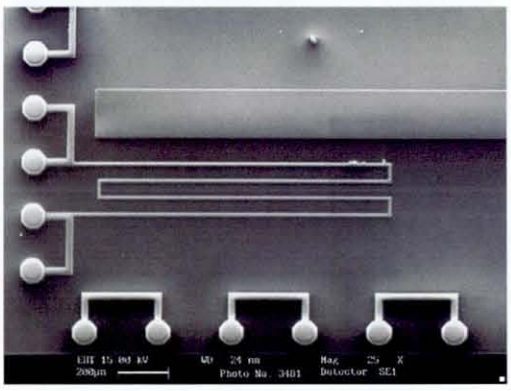
(b)



(c)



(d)



(e)

Figure 3.2.2 b, c, d & e:
SEM micrographs of the pads used in the assemblies. Aluminium bondpads before(b) and after(c) electoless bumping to 5μm and micrographs demonstrating the 300μm and 225μm pitches of the interconnection pads (d) and (e)

3.2.3: Chip and Substrate Descriptions

Heater Chip

The overall heater chip dimensions were 3x3x0.5mm. The heater chip contained the heater, two aluminium tracks that allowed for resistance measurements to be taken, and a pattern of bond pads and connecting tracks that formed half of the daisy chains. The heater chips had previously been diced and solder dipped in tin-lead eutectic solder.

Carrier chip

The carrier chips were 6x6x0.5mm. They had a peripheral array of bond-pads to match those of the heater chip such that they completed the daisy chain connections. In addition to these, the pads were also connected to larger BGA size pads that allowed for external connections to be made.

MCM assembly

To assemble the specimens, the heater chip was placed on top of the carrier chip using a fine placement machine. The assemblies were subsequently reflowed such that the interconnection layer would be complete and any initial misalignment would be corrected due to the surface tension of the molten solder. Figure 3.2.2a, demonstrates the overall manufacturing route.

Substrates: -

The MCMs were mounted on to substrates that were to serve two purposes: -

- To provide a base with suitable thermal characteristics and
- To provide a path for the electrical connections

The substrates were manufactured from either FR4 (poor thermal conductivity) or copper (good thermal conductivity). The dimensions were 40x40x0.8mm for the FR4 boards or 40x40x1mm for the copper boards. The MCM was positioned centrally on the substrate (see figures 3.2.3.c, d, and e.)

Once reflowed, the assemblies had to be glued to the relevant substrate and electrically connected. Finally external connections were made and the assembly was complete. Figures 3.2.3.c, d, and e show examples of finished assemblies. Table 3.2.1 displays the purposes and resistance values of each of the connections labeled in figure 3.2.1.a

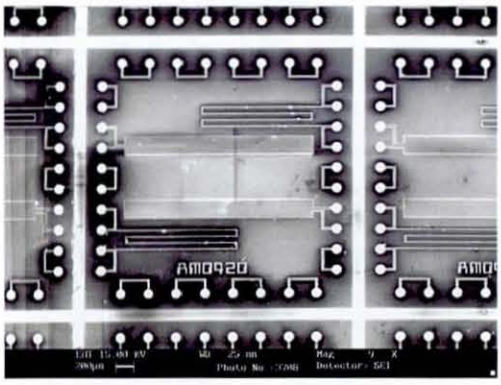


Figure 3.2.3a:
SEM micrograph of the heater chip used in the assembly. The connections for the heaters, the 4 point resistance measurements and half of the daisy chains can be seen. The image was taken from a sample still attached to a wafer

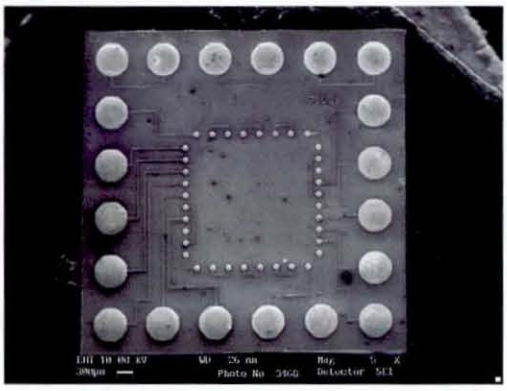
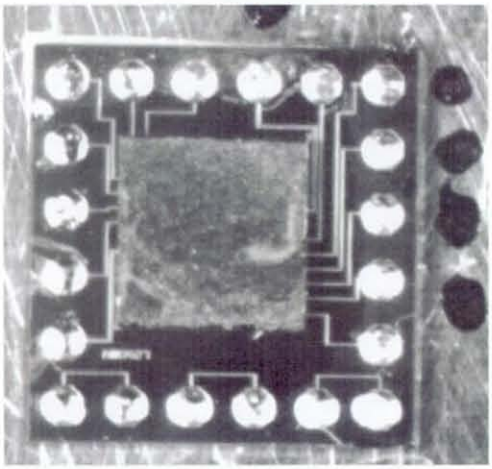
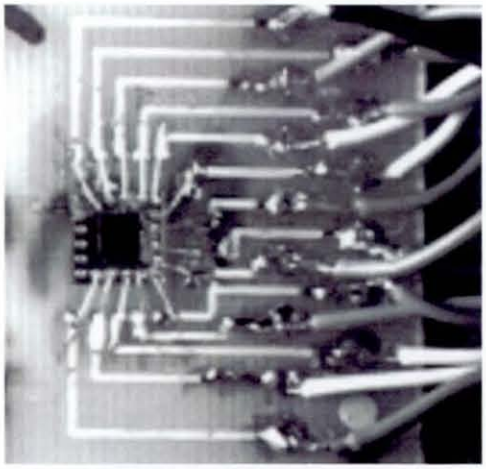


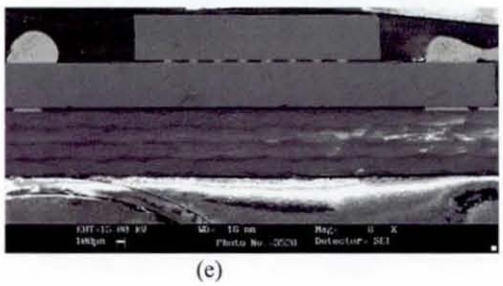
Figure 3.2.3b:
SEM micrograph of the carrier chip used in the assembly. It can be seen that the pads match those in figure 3.2.3.a In addition the carrier chip has connection wires between the small flip chip pads and the larger external BGA pads



(c)



(d)



(e)

Figure 3.2.3.c, d & e:
Examples of finished assemblies.(c) A complete assembly mounted on bare copper but not electronically connected (d) a close up of an assembly mounted on FR4 with wires attached (e), and the SEM micrograph of a sectioned assembly.

3.3: Chip Preparation

The heater and carrier chips were received in 100mm diameter wafers and the pads had been previously electroless nickel bumped. Prior to final MCM assembly, the chips had to be adequately prepared by dicing and/or solder bumping. This section describes these preparation stages.

The carrier chips were received electroless nickel bumped (but not solder dipped) and were in wafer form whereas the heater chips were received already diced and they had been previously solder dipped and stored for approximately 3 years. However, to ensure the consistency of the solder quality it was decided to dip all chips in a solder bath, regardless of whether they had or had not been solder dipped previously. Therefore before they could be assembled, the chips required wafer dicing, solder dipping, cleaning and inspection. The following section describes the necessary preparation process.

3.3.1: Chip Dicing

Due to limitations of the subsequent solder dipping procedure (the opening of the crucible for solder dipping was 40mm in diameter) it was a prerequisite to dice the wafer into individual chips prior to solder dipping. Furthermore, this ensured that each chip was submerged to the same depth for the same duration (chips at the bottom of the wafer would be submerged at a greater depth than those at the top). This gave greater confidence that factors that may influence the quantity of the solder on the pads (e.g. wetting angle) were made as constant as possible.

Dicing method

The wafer saw used was a **Microace 3 Series 2** Wafer Blade. The procedure was as follows:

- Wafer was placed on film
- Film was mounted in machine and blade was aligned
- Wafer was cut in one direction and the film was rotated 90° and blade was aligned and the wafer dicing was completed.

Using the wafer saw was much more desirable than hand scribing, with respect to both chip economy and size consistency (fewer chips were wasted and all were close to the required dimensions).

3.3.2: Solder Dipping

Solder dipping was used for the deposition of solder on to the electroless nickel bumps. This method of bumping the chips was chosen instead of conventional wafer bumping methods as the test chips were designed for RF application, requiring a low standoff height of approximately 30μm (from heater chip face to carrier chip face). This technique had been used before to successfully assemble these devices (2). Before the chips were solder dipped, they first had to be coated with flux such that any oxides present were removed and good wetting could occur. Actiec 2 (0.2% activator) was used as a flux for the chips and this was applied by brushing the surface of the chip with the flux just before dipping.

Solder dipping was carried out using an **RPS 6-sigma** wetting balance tester supplied by **Robotics Process Systems** (RPS). This particular wetting balance machine contained lead free solder from previous experiments, so the 60/40 SnPb near eutectic solder pellets were placed in a separate crucible, that was subsequently mounted in the lead free solder bath to heat the 60/40 solder to the required temperature of 250 C.

Initially a manual method of dipping the chips was performed. Once the chip had been coated with flux, it was gripped with tweezers and held above the molten solder for pre heat (a time of 30 seconds was used) before the chip was dipped in the solder for a further 30 seconds. This method produced good wetting results, however as this was a manual process, consistency could not be guaranteed (e.g. hang time and dipping time, angle of the chip on entry etc.) It was therefore considered desirable to have the procedure automated and the wetting balance itself was used for the solder dipping process, which enabled consistency as a fixed hang and dwell time could be ensured. Figure 3.3.2. demonstrates the dipping procedure, which was as follows. 1) chips were brushed with flux and held by the edges using a spring clip, the clip was mounted on the wetting balance tester, 2) machine moved the clip over the molten solder and positioned it 1mm above the solder bath for 30 seconds (pre heat stage) 3) the chip was then submerged into the solder for the desired time period (30 seconds) before it was withdrawn and allowed to cool.

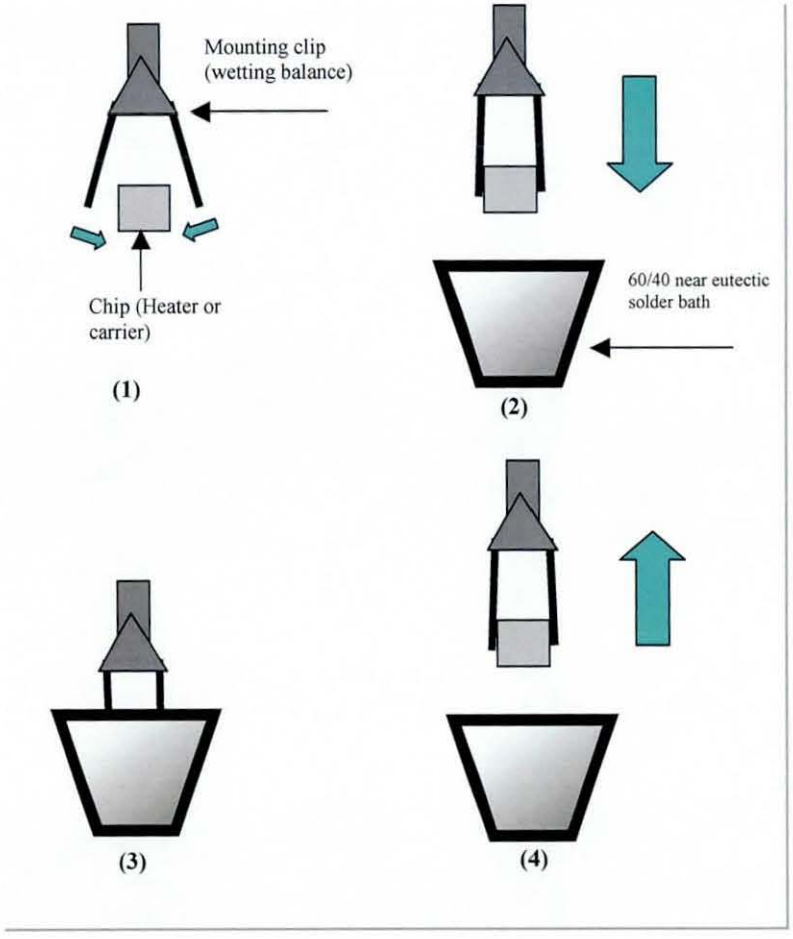


Figure 3.3.2:

Schematic of chip mounting and solder dipping procedure. The chip is put in the clip (1), then preheated(2) then submerged (3) and withdrawn (4)

3.3.3: Inspection

Once the chips had been solder dipped, they then had to be inspected to check that the solder pads had been correctly capped with solder. Before this could take place, the chips first required cleaning to remove remaining flux residues. This was done by placing a chip in isoproanol (IPA) solution and then applying ultrasonic energy for three minutes. This time proved long enough to remove most flux residue though if any flux residue remained then the process was repeated until all the chips were free of flux.

Following the cleaning, the chips were then inspected to verify that all the bond pads were properly capped with solder using an optical microscope. Satisfactory chips would have all the bond-pads spherically capped with solder. Any unsatisfactory chips (i.e. a chip with one or more pads that were only partially capped or didn't wet at all) were then solder dipped a second time while satisfactory chips were then referenced (i.e. it was assigned a number/letter) and considered for MCM assembly.

3.4: MCM Assembly

This section describes the experiments conducted to develop a suitable manufacturing process for the assembly of the MCMs. The assembly procedure involved placing the heater chip onto the carrier chip using a fine placement machine and then reflowing the device which was supported on a 90x90mm FR4 board.

The MCM assembly procedure was initially carried out in the simplest manner available on a single test device that was subsequently cross-sectioned and examined. The joints obtained are displayed in figure 3.4. It was evident that the manufacturing process was unsatisfactory as both misalignment and poor chip wetting were apparent. It was then decided to examine each stage of the manufacturing process by choosing to alter certain variables/parameters and then preparing a test device to evaluate the effect on the assembly. Once the most suitable process for each stage was found this was subsequently used in the manufacturing process.

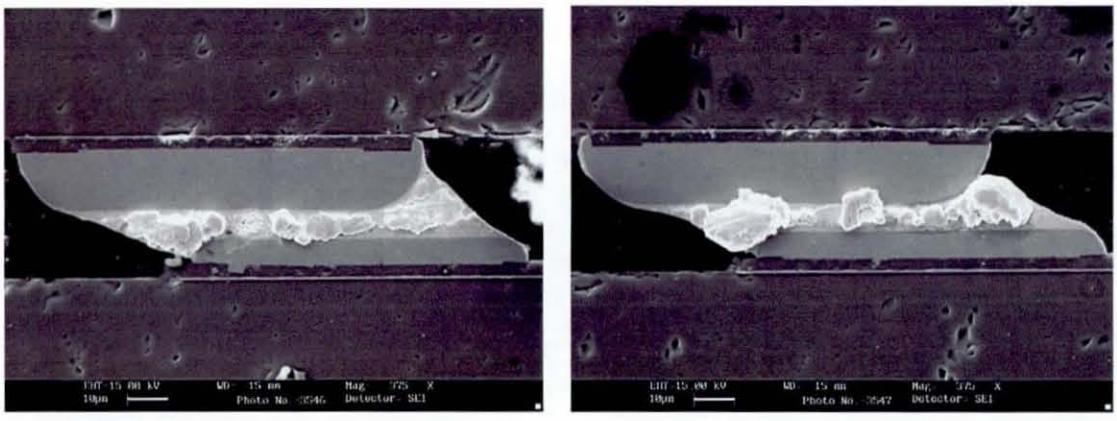


Figure 3.4:
SEM micrographs of the interconnection layer before the manufacturing process was refined

3.4.1: Fluxing

Before the chips could be placed and reflowed, it was first necessary to apply flux to ensure that the solder caps were free of oxides that would inevitably impede the reflow process. Ensuring precisely the right amount of flux was applied onto the chip was crucial to permit an accurate reflow procedure; if insufficient flux was applied to the chip then the flux was ineffective, conversely if there was too much flux then the heater chip failed to make contact with the carrier chip due to the very low standoff heights. For these reasons, it was decided that the flux should be applied manually using a hand brush such that the amount of flux going on to the chip could be controlled.

There were two fluxes available: Actiec 2 (0.2% activator) and Actiec 5 (0.5% activator). Three fluxing techniques were considered: -

- Actiec 2 applied to carrier chip only
- Actiec 5 applied to carrier chip only
- Actiec 2 applied to heater and carrier chips

Each technique was applied to three different test assemblies and they were subsequently reflowed and sectioned.

Actiec 2 applied to carrier chip only.

Actiec 2 was used initially, as it was the weaker strength flux and due to the relevant hazard and environmental concerns, it was favoured to use the weaker flux if possible. Furthermore it was desirable to avoid placing flux on the heater chip if possible as this could obstruct the subsequent placement stage (the flux may obscure the image of the solder bumps in the fine placer therefore making it difficult to align). Therefore Actiec 2 flux was placed on the carrier chip only and the test device was then reflowed and sectioned. The cross section images of the assembly produced in this way are shown in figure 3.4.1a. From the cross-section, it could clearly be seen that the solder had failed to adequately wet the carrier UBM and there was very poor wetting on the heater UBM. In figure 3.4.1a (ii), an adverse bulge of solder can be seen between the solder and the pad on the heater chip, this was possibly due to the flux failing to remove oxides on the pad of the heater chip. Such bulges may act as stress raisers that could potentially contribute to premature failure.

Actiec 5 applied to carrier chip only

Actiec 5 was used to investigate the influence of increasing the strength of flux used in the assembly. The application procedure was identical to that of the Actiec 2 flux described previously. The results of the cross-section through a sample are shown in figure 3.4.1b. It could be seen that increasing the strength of the flux from that of Actiec 2 to Actiec 5 had improved the overall solder wetting of the carrier chip, however in 3.4.1b (ii) the wetting of the heater chip was still unsatisfactory, and it could be seen that there was some misalignment of the bond-pads.

Actiec 2 applied to both chips

It was originally intended to avoid placing flux on the heater chips if possible as this would lead to visual problems when trying to align the heater and carrier chips during the chip placement stage. However it became apparent that the fundamental issue with the previous examples was that the flux was not cleaning the pads adequately on the heater chip. Therefore it was necessary to apply a very controlled amount of flux to the heater chip as well. Figure 3.4.1c shows images of cross-sections through samples prepared in this way. The application of flux to both the heater and carrier chips had

the most significant effect on the quality of the joints; in particular the improved wetting on the UBM of the heater chip was noted. It was anticipated that the wetting could be further improved by applying Actiec 5 on both the heater and carrier chips so it was decided to prepare subsequent assemblies by applying Actiec 5 to both the heater and carrier chips.

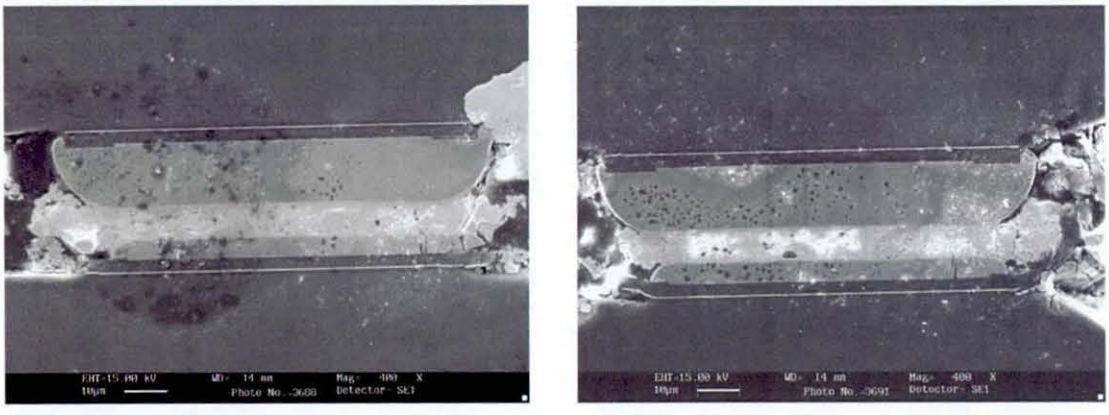


Figure 3.4.1a:
SEM micrographs of interconnection layers of MCM assembled with Actiec 2 applied to the carrier chip only

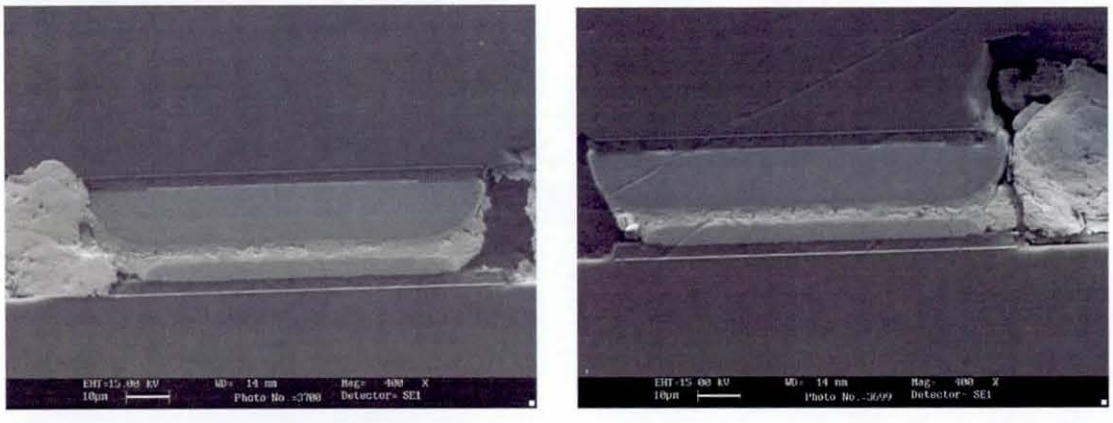


Figure 3.4.1b:
SEM micrographs of interconnection layers of MCM assembled with Actiec 5 applied to the carrier chip only. The clumps at the side of the bumps are excess gold due to a faulty gold coating process

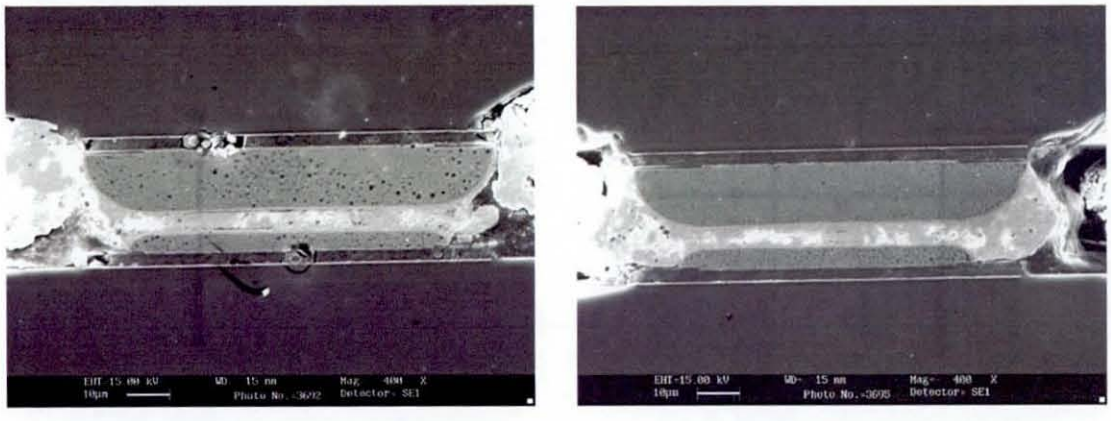


Figure 3.4.1c:
SEM micrographs of interconnection layers of MCM assembled with Actiec 2 applied to both the heater and carrier chip

3.4.2: Reflow Profile: -

Reflow was conducted in a conveyor reflow oven. There was a choice between two types of oven that could be used: -

- 3-zone **Surf Systems** infrared reflow oven
- 7-zone **Quad QRS VII** air convection reflow oven

Each oven was profiled by securing an MCM to a 90x90mm board using **loctite chip bonder** adhesive, and attaching two thermocouples: one to the board and the other to the MCM which were then connected to a DATAPAQ data logger. The assembly and the data logger were passed through the oven and once complete the recorded temperatures were downloaded to a PC and analysed with the DATAPAQ software.

The three-zone oven was the initial choice for assembly due to the simplicity of the set up, i.e. only 3 zones and short time required to pass through the oven. Once thermally profiled the initial results implied that this would be a suitable reflow profile. However upon inspection it was found that the joints seldom reflowed and on occasions when reflow did occur, subsequent sectioning revealed that the quality of the reflowed solder joint was poor. Several attempts were made to rectify this issue including increasing temperatures as well as slowing the conveyor down but these methods did not improve the quality of the joints. It was noted that silicon may be infra red transparent, therefore the only heat transfer was from the FR4 board through the carrier chip and consequently the assembly failed to reach the correct reflow temperatures.

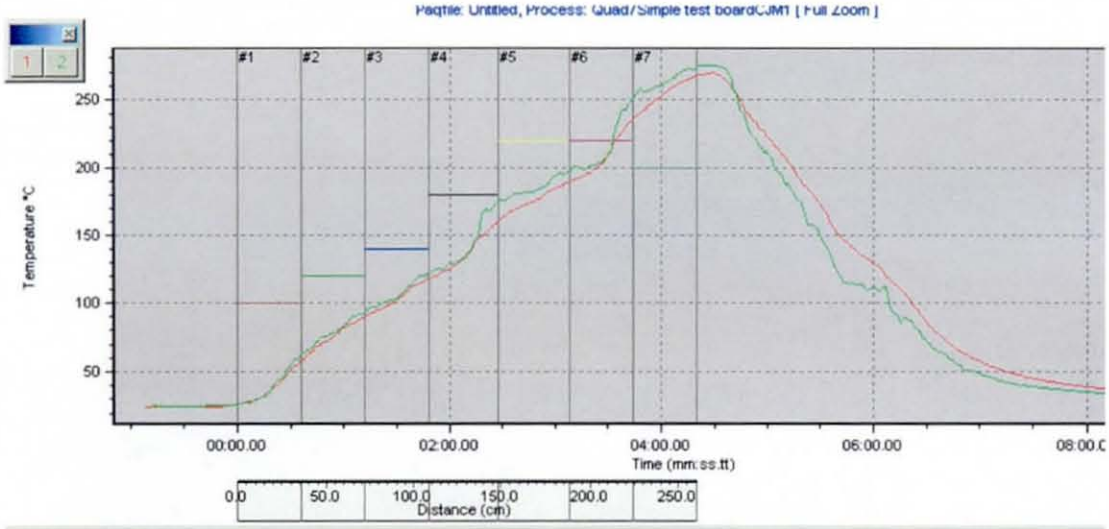
A 7-zone convection oven was subsequently analysed for suitability as this was seen to potentially eliminate the problem of infrared transparency. Furthermore, the 7-zone oven would allow for a finer control of the obtained temperature profile. The oven zone temperatures were used are shown in figure 3.4.2a.

Zone	1	2	3	4	5	6	7
Temp.(°C)	100	120	140	215	215	300	300

Figure 3.4.2a: *Profile of the final temperatures of the zones used.*

The prototype board was passed through the oven with a conveyor speed of 600mm/minute. Figure 3.4.2b displays the temperature plot obtained.

The red line represents the temperature of the board while the green line represents the temperature of the silicon assembly. Due to the small thermal mass of the MCM, the temperature was assumed to be uniform throughout the MCM. It could be seen that the temperatures of the assembly were actually greater than the general criteria for reflowing 63/37 solder. (i.e. for the assembly to be above 183°C for a designated period and to reach a peak temperature of approximately 220°C). However due to the previous difficulty in getting the joints to reflow, it was decided that as the assembly contained no live components the maximum temperature reached during the reflow was less critical and it was considered of greater importance to ensure that the solder melted correctly to ensure the alignment issues were corrected.



Reflow Results

Probe	Positive Slope (°C/sec)	Positive Slope Time (mm:ss.tt)	Rise Time (120.0 - 160.0°C) (mm:ss.tt)	Time Above Liquidus (183.0°C) (mm:ss.tt)	Peak Temperature (°C)	Delta T (°C)	Negative Slope (°C/sec)
#1 (°C)	2.35	08:18.00	00:37.00	02:28.00	269.5	5.5	-2.30
#2 (°C)	4.18	03:33.00	00:35.00	02:25.00	275.0		-3.64

Figure 3.4.2b:
Reflow profile of the test assembly used in preparation. The general profile shows a temperature regime greater than generally required for the eutectic solder used.

3.4.3: MCM Inspection

Once the components had been successfully reflowed, they then required inspection to verify that the solder had reflowed correctly. Probing the connections labeled in 3.2.1 with a multi-meter to measure the resistance was used as a method of testing the chips. Successful assemblies would have all connections working with resistance values close to those shown in table 3.2.1. One such assembly was tested and found to be satisfactory, this was subsequently cross-sectioned and the obtained images are shown in figure 3.4.3

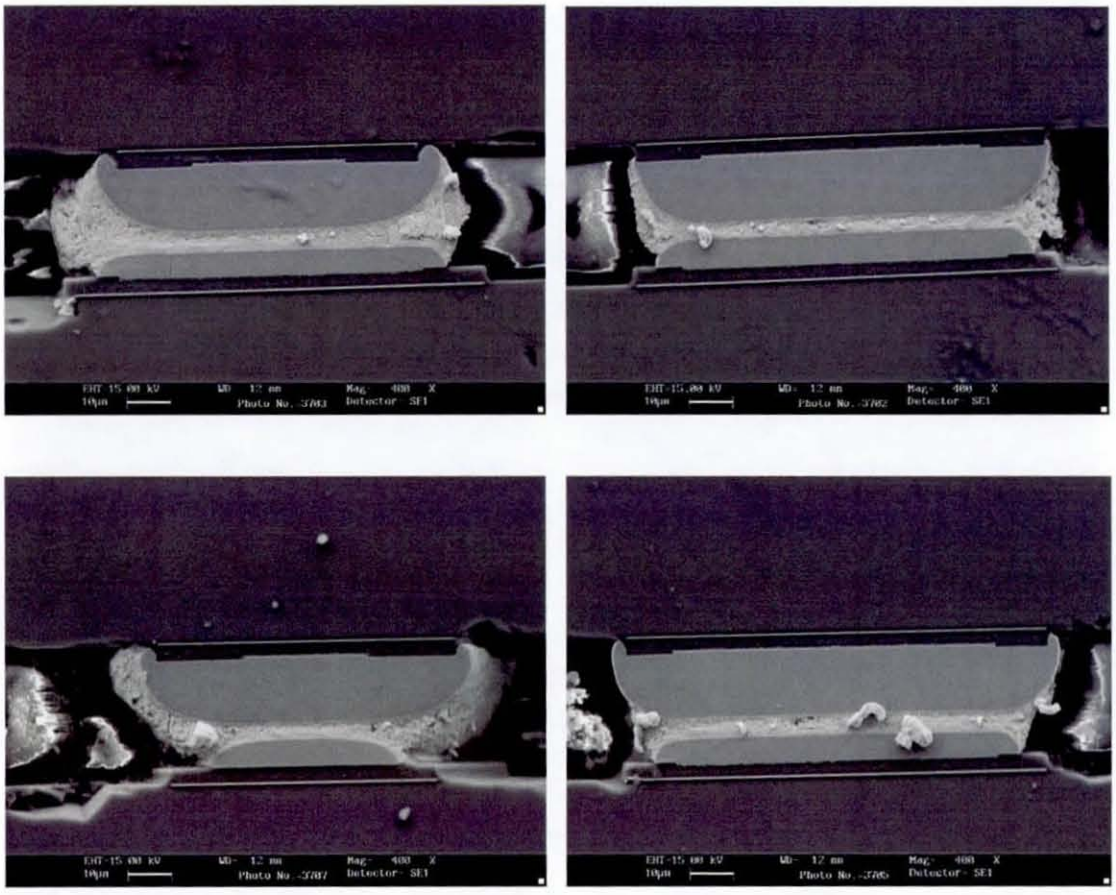


Figure 3.4.3:
SEM micrographs of a satisfactory test piece. The particles on the sample are dust that was contracted between sectioning and gold coating.

3.5: MCM to Board Connection

With the successful assembly of the MCM devices completed, they were then attached to a substrate, which had to support the electrical connections and to provide the desired thermal characteristics.

3.5.1: Chip Attachment to Substrate

In total, three different methods of securing the MCM to the substrate were considered: -

- Loctite® 3608 Chip bonder
- Araldite® Rapid epoxy
- Chomerics ® THERMATTACH T404 Thermal adhesive pad

Chip bonder: This was the initial choice for attachment as the adhesive was designed for similar purposes and also some of the necessary thermal data was available. The procedure was as follows: 1) clean surfaces on both MCM and board using IPA and lint free cloth; 2) apply chip bonder to the specified area on the board and finally 3) position MCM on board and apply the necessary heat (150°C for 30 seconds). This method was not suitable as it was difficult to control the amount of chip bonder applied to the board. It was considered important to have a controlled amount of chip bonder from both a manufacturing consistency and thermal modelling perspective. In addition, the chip bonder had to be cured at a high temperature, which was difficult to control effectively.

Araldite Rapid epoxy: This was considered a viable method, as the epoxy did not require high curing temperatures (room temperature cure). Araldite rapid was used because of the fast curing time. The procedure was as follows: - 1) surfaces were cleaned using IPA and lint free cloth 2) Araldite mixture was correctly prepared 3) Araldite was applied to the MCM (the back of the carrier chip) 4) MCM was placed on the board and left to cure for 15 minutes. This method provided adequate support for

the chips however, as with the chip bonder, the high viscosity of the glue resulted in difficulty in regulating the amount applied to the assembly and in addition, the thermo-mechanical properties of Araldite were unknown such that inclusion in the FE models was problematic.

Chomerics ® THERMATTACH T404 pad: This adhesive was purposely designed to attach plastic encapsulated or metallic components onto heat sinks, and was provided as a sheet of material such that it allowed for a controlled amount of the substance to be put in place. Furthermore, the relevant thermal properties were available on the accompanying datasheet. The procedure for use was 1) clean the surfaces of both the MCM and the board using IPA and a lint free cloth; 2) cut adhesive pad to size using a scalpel or knife; 3) remove the protective film from the correct side and place pad onto substrate surface; 4) remove the other side protective film and correctly position chip on top and 5) apply pressure for 30 seconds.

The procedure used is shown in figure 3.5.1 This method also had its associated problems primarily due to the final part of the assembly stage (the uniform pressure application). The manufacturer specified a uniform pressure of 0.069 MPa for 30 seconds. Several different methods of applying the uniform pressure (e.g. putting weights on the assembly) were tried but many of these resulted in destroying the test pieces. Tweezers were used by turning them to the side, placing them horizontally over the edges of the carrier chip and then pressing down on the tweezers. As the pressure was applied by hand, there could be no guarantee that the applied pressure was in fact 0.069MPa. In addition the pressure was supposed to be applied as a uniformly distributed load and in this case the pressure was applied in 4 individual loads (2+2). Figure 3.5.1 demonstrates how the loads were applied to the chip. As the silicon was rigid, it was assumed that the load was distributed uniformly across the carrier chip base, however it is possible that the pad may not have adhered to the chip correctly around the central region of the carrier chip.

Despite this flaw, the Chomerics adhesive pad was considered the best available method of securing the chip to the substrate as it allowed for a consistent and controlled amount to be placed down.

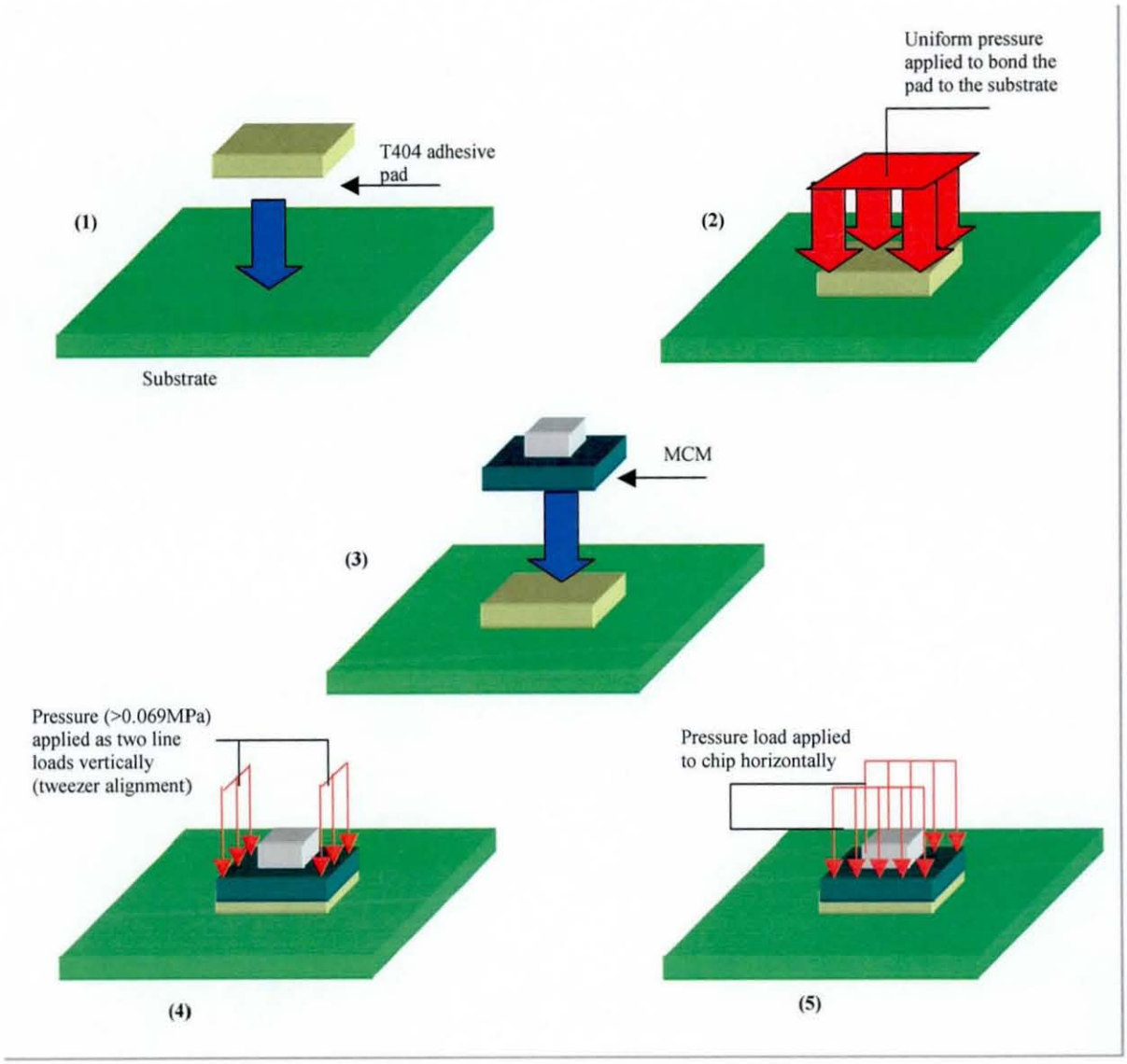


Figure 3.5.1:
Process route showing the securing of the MCM to the substrate. It can be seen that the T404 adhesive pad is first placed onto the substrate and then a uniform pressure is applied (1&2), then the chip is positioned correctly (3) and finally the required pressure to secure the MCM to the substrate is applied in two separate operations (4&5)

3.5.2: Attaching Wires from MCM to Board

It was initially intended to use a wire bonder to connect the larger BGA size pads from the carrier chip to the pads on the board, as this would be a controlled and automated procedure. However, as the carrier's external pads had already been solder capped from the dipping procedure this was unsuitable. It was therefore necessary to complete the connections from the MCM to the board by manually soldering each connection. To form the connections of the BGA pads on the MCM to the pads on the substrate a thin wire was used: An insulated wire was taken, the coating was stripped off and the individual wire strands were separated. Once separated a wire was tinned with solder using a conventional solder iron.

The route for attaching the wires to the necessary pads is shown in figure 3.5.2 and the procedure was as follows 1) solder paste was dispensed on top of the pads on the MCM, 2) the tinned wire was held close and the solder iron was placed on the joint. Once the solder had reflowed, the surface tension of the molten solder pulled the wire close to the pad and once the solder solidified, a connection was made to the pad as shown in figure 3.5.2. Once completed, the solder paste was then applied to the pads on the board and the solder iron melted the solder and the surface tension drew the wire to the pad. When a connection was finished, the remaining wire was clipped from the assembly and the process continued until all pads were complete (see figure 3.5.2).

Afterwards, the external connections to the board had to be completed. This was performed by simply hand soldering wires to the array of pads at the edge of the surface. Figure 3.2.3c, d & e show completed MCMs mounted on copper, FR4 and a section respectively.

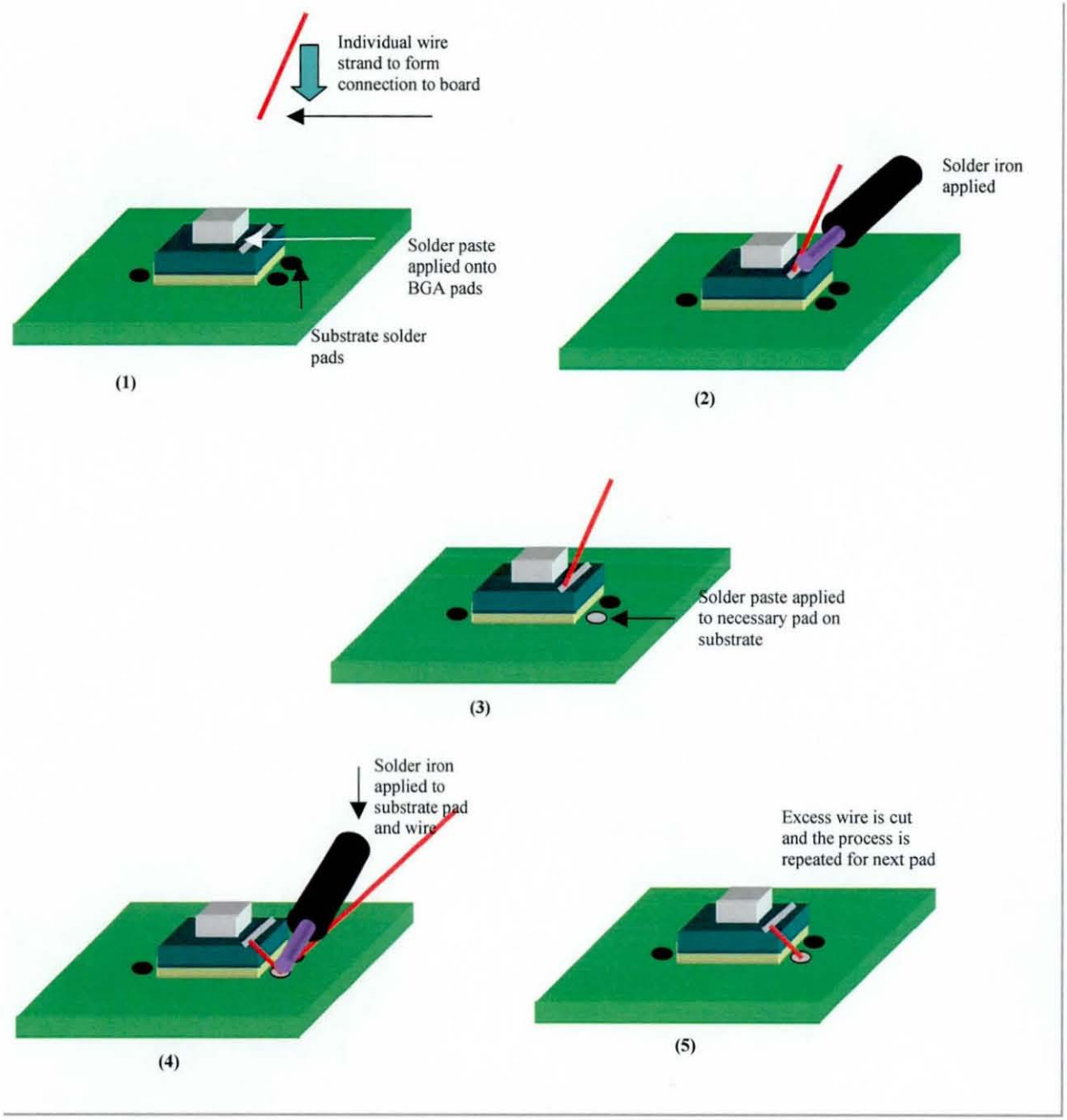


Figure 3.5.2:
Process route showing the electrical connections between the MCM and the substrate. Shown is the securing of the wire to the MCM (1&2) and the subsequent wire attachment to the board (3, 4 & 5)

3.6: Final Assembly Procedure

The analysis and refinement of each stage in the manufacturing procedure resulted in the following manufacturing process that was used to manufacture all subsequent test vehicles

1) Heater and carrier wafers were diced using a Microace 3 Series 2 Wafer blade.
2) Individual chips and substrates were coated with Actiec 2 flux by brushing the flux on top of the chip. The chips were then dipped in a 60/40 eutectic solder bath at 250°C for 30 seconds
3) Chips and substrates were cleaned: they were placed in a beaker of IPA and the beaker was put in an ultrasonic bath, for three minutes. They were subsequently examined for defective solder caps.
4) Satisfactory chips (i.e. chips that had all pads capped with solder) were then hand fluxed using Actiec 5 and aligned using a fine placement machine
5) The assemblies were reflowed using a Quad QRS VII 7 zone oven. All connections on the chips were checked and the resistance of each of the chips was recorded.
6) Assemblies were placed on the appropriate substrates. They were bonded using a Chomerics T401 adhesive pad as specified in section 3.5.1
7) Connection wires were then hand soldered from the connection pads to the appropriate pads on the board.
8) External wires were then connected to the pads on the board to allow for the chip to be connected to the relevant power sources and monitoring points.

References

1) **A.D. Trigg & A.R. Corless** "Thermal Performance and Reliability Aspects of Silicon Hybrid Multi- Chip Modules" *40th Electronic Components and Technology Conference, 1990, pages 592-9*

2) **David A. Hutt, Samjid H Mannan, David C Whalley and Paul Conway** "A Maskless Low-Cost Multi-Chip Module Assembly Process" *Advances in Electronic Packaging (InterPACK99) ASME EEP Volume 26-2, 1999, pages 1705-11*

Chapter 4: Thermal Data Collection from MCM Assemblies

As previously stated, the primary failure mechanism for flip chip assemblies has been identified as a consequence of the chip and substrate thermally expanding and contracting at different rates. This may happen as a consequence of:

a) The chip and substrate being manufactured from different materials and consequently having different CTEs.
b) The chip and substrate may heat up and cool down at different rates due to a potential non-isothermal temperature distribution within the assembly.

Either scenario, or both together are known to instigate the failure mechanisms within a solder joint depending on the type of device. With reference to this experiment, it is the latter scenario that best describes the applicable failure mechanism and it was therefore necessary to capture the temperature distribution to assess its impact and to provide data to compare with the FE models.

The aim of this chapter is to describe the hardware used to capture the thermal data and to present the results obtained from the MCM test vehicles.

4.1: Power Cycling Chamber

A controlled environment for the assemblies to be tested in was necessary in order to ensure that the tests were carried out under known and repeatable conditions. Therefore a wind tunnel was used to provide such a controlled environment within which the cooling mode of the chips could be regulated. The wind tunnel is shown in figures 4.1a to c and it had a rectangular cross-section with a fixed width of 250mm. The depth of the tunnel could be varied from 20 to 200 mm using a moveable platform, which had a leading edge of 1500mm. The chips were placed in the central "observation" region of the wind tunnel, where there was a removable panel that allowed for the area to be sealed or open. Figure 4.1a displays a simple schematic of the wind tunnel, which was situated in ambient conditions; i.e. air at room temperature and pressure (the cooling fluid).

The wind tunnel was designed and hand built by the student to a specification requirement for an earlier set of experiments (1). The experiments in question required that a constant measurable airflow rate should be obtainable within the middle section of the tunnel such that novel heat sink designs could be assessed. Upon evaluation, the wind tunnel was found to provide a good control of airflow for the region where the assemblies were placed. The air was drawn into the tunnel by a fan that had a speed control to allow different airflow rates. The air velocity was measured by an **Airflow TA35 anemometer** through several probing holes situated at the specimen area as shown in figure 4.1b. Unfortunately it was not possible to obtain a constant wind velocity for the whole specimen area (across the area the airflow velocity was observed to vary by 10%).

The wind tunnel was situated in a multi-purpose laboratory and the temperature within the laboratory varied depending on the time of day/year. Consequently while the airflow within the tunnel could be controlled, there was very little control over the ambient air temperature; the temperature of the laboratory was known to be between 17 and 25°C during the day while the tunnel was used. However, the temperature in the evening was not known.

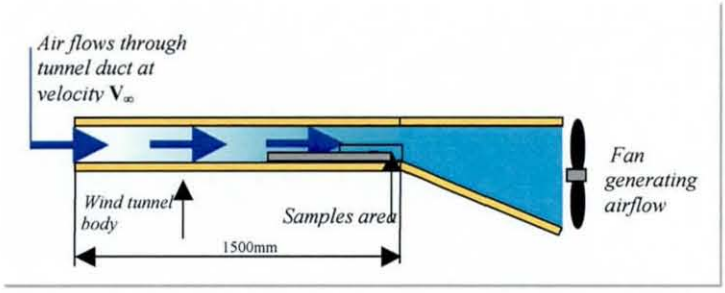


Figure 4.1a: *Schematic of wind tunnel used for power cycling experiments.*

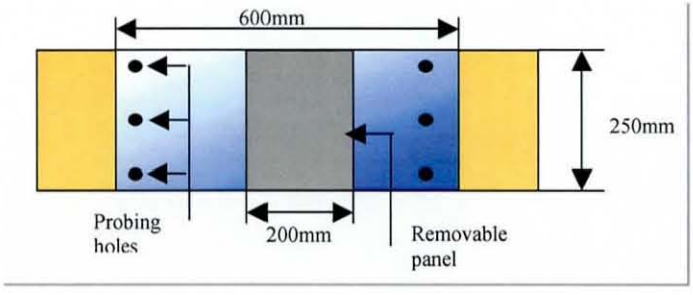


Figure 4.1b: *Schematic of sample observation/probing area on the wind tunnel body (top view)*

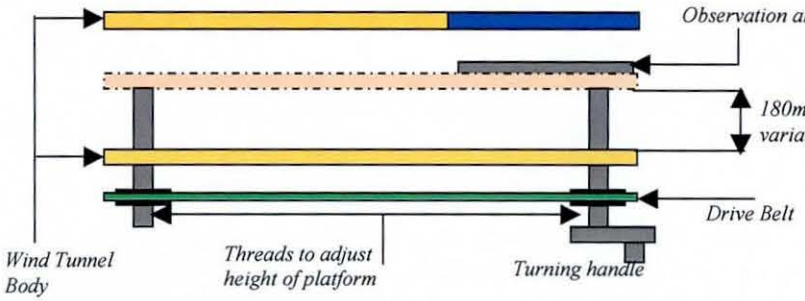


Figure 4.1c: *Schematic of the variable platform mechanism*

Time Controller

The MCMs could be heated simply by connecting a power supply across the resistance heater. However, in order to carry out the transient power cycling trials a control timer was manufactured such that the power could be turned on and off at regular intervals. The timer was set for a 128s cycle time (64s on / 64s off).

4.2: Assembly Temperature Measurement

In order to validate the thermal models it was necessary to obtain accurate measurements of the temperatures reached by the assemblies under different power/airflow conditions. Several techniques were evaluated as described in the following sections.

4.2.1: Thermocouples

Three K-type thermocouples were used and they were attached using **Loctite® 3608 chip bonder** adhesive to the heater chip, a corner of the carrier chip and on the substrate secured in place. Once a thermocouple was attached to its respective body, temperatures were then recorded when the test vehicle had been powered up, however while this was a convenient method, there were a number of factors that may have compromised the accuracy of the results.

Thermocouple Attachment: The chip bonder adhesive required curing at 120°C for 30 seconds and this was difficult to achieve in a reproducible manner. The adhesive was cured by using a hot air blower and this led to similar problems encountered when originally attaching the assembly to the substrate; there was no way of ensuring that this was the maximum temperature reached in the assembly. Moreover, this curing

procedure could compromise the joint integrity as well as adding further intermetallic growth to the solder joints.

Thermal Contact: It was assumed that all contact between the assembly body and the thermocouples was good such that an accurate measure of temperature was obtained, however there was no way of verifying that this was the case. If the contact was poor, due to a layer of the relatively low conductive adhesive intervening, then the recorded temperature reading may have significantly underestimated the true temperature.

Volume: It was noted that the thermocouples used were comparatively large compared with the dimensions of the assembly. Therefore the possibility of the thermocouple conducting heat away from the assembly could not be overlooked. Furthermore, the volume of the adhesive used was significant compared to that of the assembly, which may also have obstructed the convection cooling, reducing its efficiency (if not increasing the volume of the whole assembly to be heated.)

It was clear that readings obtained by thermocouples alone could not be relied upon to be totally accurate, especially for rapidly changing temperatures, therefore a more robust method of obtaining the temperatures was investigated.

4.2.2: Four Point Resistance Measurements

On each heater chip there were two aluminium tracks that were each connected to four bond pads. This enabled 4-point resistance measurements to be made using a **Keithley® 580 micro-ohmmeter**. It is well known that the resistance of a metal increases when the temperature is raised, therefore by monitoring the resistance of the track it was possible to determine the temperature of the heater chip.

The tracks were located very close to the heater, therefore they were expected to provide an accurate value of the maximum temperature reached. Unfortunately, there were no such tracks on the carrier chip so an equivalent reading could not be obtained.

In order to use this technique, it was first necessary to calibrate the resistance of the track as a function of the temperature. This was done by placing the test pieces in a **Siemens** fan oven; the oven used was a domestic cooking oven that had the controls converted such that the user could accurately control the temperature to within one degree. A fan forced the air to circulate around the oven that ensured a uniform temperature distribution.

The resistance readings were recorded for 5°C temperature increments. Once a temperature was reached, the oven was left to stand for a minimum period of 15 minutes to allow the samples to reach steady state. A thermocouple was attached to the assembly and also, a thermocouple was placed loosely in the area of the test vehicles to verify the air temperature of the oven. In addition to this, the resistance reading of the track was monitored to observe any changes over time. Once the change was seen to be minimal (i.e. less than 0.001Ω in one minute), it was assumed that the sample had reached steady state and the value was recorded. The calibration procedure was performed on four test pieces i.e. two samples for each type of substrate. Table 4.2.2a describes the resistance measurements for the test vehicles at room temperature, while figure 4.2.2 shows variation in resistance with temperature. Only one aluminium track was used for calibration though readings on the alternative track were checked periodically to verify that the resistance change was consistent with the track used.

Sample	***Substrate Material***	***Room Temperature Resistance***
001	FR4	5.717Ω (room temp=26°C)
003	FR4	6.194Ω (room temp=26°C)
A0	Copper	5.825Ω (room temp=23°C)
A1	Copper	6.335Ω (room temp=23°C)

Table 4.2.2a:
Table showing the resistance of the specimens at room temperature prior to being profiled in the oven. The difference in the start temperatures was due to theFR4/ copper profiles performed on different days.

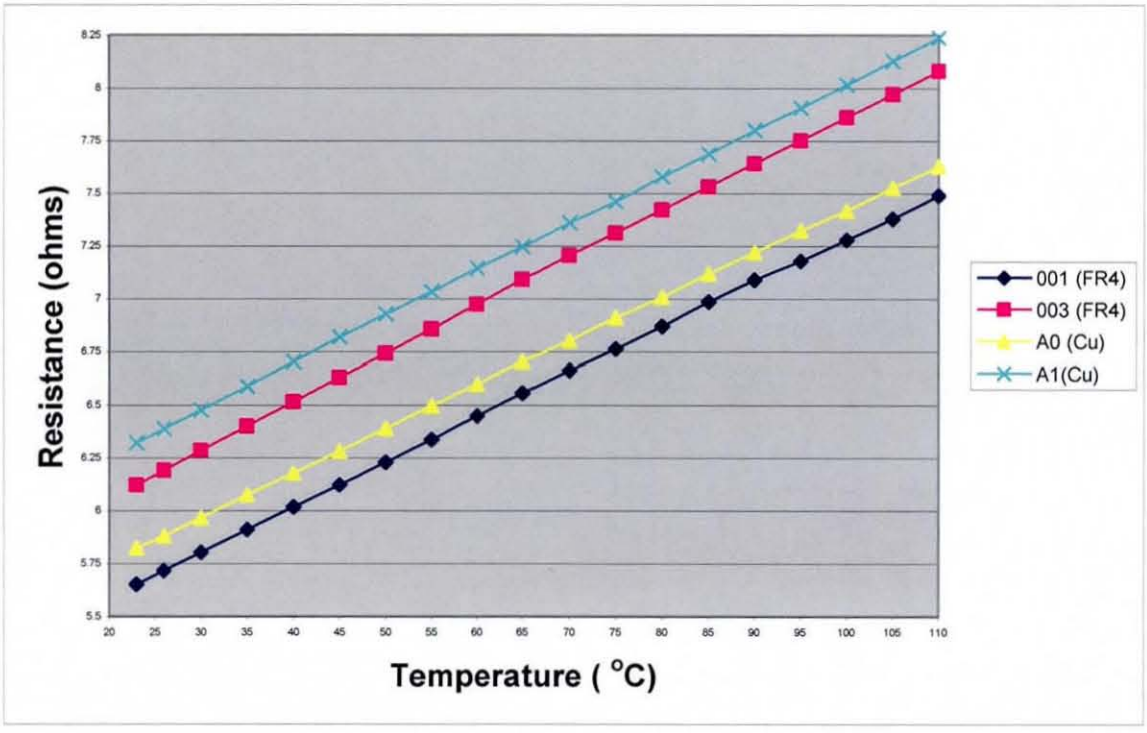


Figure 4.2.2:
Graph of 4-point resistance of aluminium tracks against the oven temperature

It can be seen that the resistance change with temperature was effectively constant with all the profiled assemblies. The gradients of the lines in figure 4.2.2 were compared to find the observed temperature change matching a 1Ω resistance increase for each test piece and the results for each assembly are shown in table 4.2.2b. It was assumed that the resistance change of aluminum tracks on subsequent assemblies would follow the same behaviour and therefore were used to provide an approximation for the maximum temperatures reached by assemblies in the subsequent power cycling tests.

Assembly	Observed change (°C/Ω)
001	47.88
003	44.91
A0	48.25
A1	45.33
Average	46.59

Table 4.2.2b:
Table showing the temperature rise for each assembly corresponding to 1 Ω observed rise in resistance.

Based on the above readings, an estimate for the mean temperature rise for 1Ω increase was 46.59°C/Ω with a standard deviation of 1.715°C/Ω. Therefore for the subsequent power cycling tests, each assembly could have a 4-point resistance measurement taken at the very end of each ON or OFF cycle to allow an estimate for the temperature change to be made.

4.2.3: Thermal Imaging Cameras

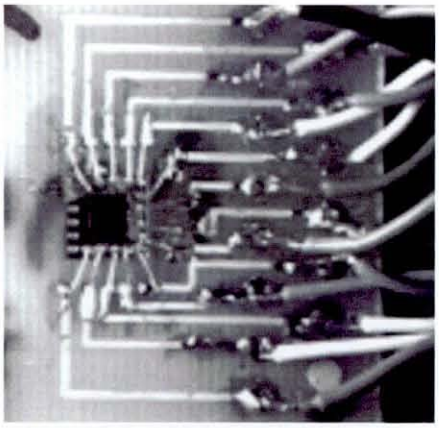
Though the 4-point resistance measurement provided an estimate of the temperature of the heater chip, it failed to provide any estimate for the temperature of either the carrier chip or the substrate or the overall temperature distribution. It was necessary to obtain a temperature plot of the whole assembly such that the heater chip temperature could be displayed relative to that of the carrier chip and the substrate. Thermal imaging cameras coupled to two-dimensional thermal imaging software were therefore used to present the thermal distribution of a given area. Two cameras were available: -

- High resolution thermal image camera
- Low resolution thermal image camera

Figure 4.2.3 shows examples of the data that could be captured from each camera. The specimen captured is the MCM mounted on FR4 board.

High Resolution: The **ThermaCAM®** P40 Infrared camera was used to capture the image of the specimen. The resolution of the obtained images was 320x240 pixels. The camera was held and focused by the user and a still image was taken. The camera used could only capture *still* images and was incapable of capturing transient thermal *movies*.

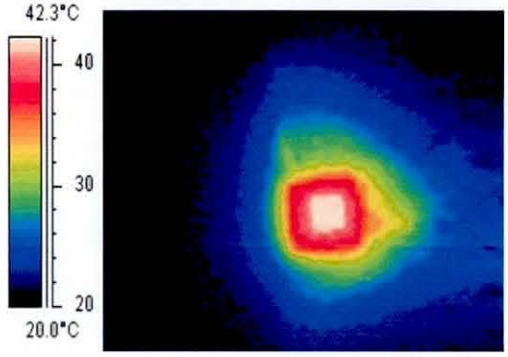
Low resolution: The low-resolution camera available was an **IRISYS®** IR11002 Infrared imager (multi-point radiometer) allowing for image capture of no more than 16x16 pixels for a given area. The camera came supplied with a Germanium lens that was to be used at a specified focal distance; in order to obtain a useful resolution from these experiments the manufacturer recommended that the camera be placed 100mm away from the sample. The camera lens had a 20° angle, such that the area covered was most of the board (an area of 35.2x35.2mm was obtained) therefore each pixel represented an area of $2.2x2.2mm^2$.



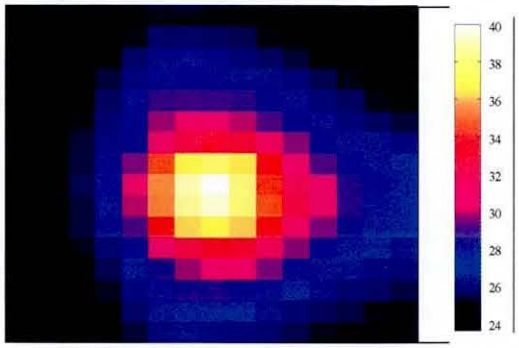
(a) Chip

Figure 4.2.3:
Examples of a high-resolution thermal plot against the low resolution equivalent. Also included is the original assembly showing the orientation

Airflow in this direction for all images



(b) High resolution



(c) Low resolution

On the high-resolution camera, the bodies of the heater and carrier chips could be clearly identified, as well as the immediate surrounding substrate area. Furthermore some of the copper tracks could be distinguished as these become slightly warmer due to heat conduction. When using the low-resolution camera, the MCM assembly could be distinguished from the surrounding substrate, though differentiating the heater and carrier chips was less obvious. The hottest pixel was assumed to be the heater chip, while adjacent chips were considered to represent the carrier chip. Despite the poor resolution (compared with that of the **ThermaCAM P40**) the IRISYS camera was considered useful as it had the facility to record transient thermal images. Furthermore, the camera was positioned first and then controlled by a PC so positioning errors were minimised.

It is clear that no single measuring technique could be used to capture all the thermal data required however, when used in conjunction with each other adequate temperature data could be obtained.

Two sets of thermal data were required; namely the steady state temperature distribution and the transient profiles. It was decided that the four- point resistance measurements could be used to obtain an absolute maximum temperature in the assembly, and the cameras could record the temperature distributions. The high-resolution camera was used to capture an image of the chip at steady state, while the low resolution could be used to record the transient behaviour of the assembly.

4.3: Thermal Data Capture from the MCMs

The MCM assemblies to be imaged were manufactured in accordance with the procedure derived in chapter 3. In addition, the assemblies were sprayed matt black to ensure uniform and high emissivity, such that good thermal images could be obtained from them. When the test vehicles were being profiled, they were secured to the viewing platform in designated positions by attaching high temperature tape to the corners of the substrate. It was decided to record images of different power levels while keeping the airflow rate constant. The results of the thermal profiles were to be compared with those obtained from the finite element models later.

Two sets of data were required from the test vehicles: **steady state** and **transient** data. steady state meaning the final temperature distribution of an assembly following a constant power supply while the transient data refers to the manner in which an assembly heats up and cools down. The steady state data was profiled at 5m/s due to use of the high-resolution thermal camera, while the transient data was recorded using an airflow rate of 10m/s.

4.3.1: Steady State Procedure

The MCMs were directly connected to the power source to provide constant power dissipation in the assemblies and the ohmmeter was also connected to the 4-point resistance track so the temperature could be measured. After the desired power level and air velocity had been correctly verified, the samples were left to reach steady state; steady state was determined when the ohmmeter detected a resistance change of less than 0.001Ω in one min. When the specimen had reached steady state, the camera was positioned and a thermal image of the sample was taken. Once captured, the next power level was set and the process was repeated.

4.3.1.1: Assemblies on FR4 Substrate

Apparatus and Setup

The **ThermaCAM** high-resolution camera was used to capture the thermal profile of the assemblies at steady state, as this camera offered the highest level of detail for any chip. In order to correctly use the **ThermaCAM** high-resolution camera, the central panel of the wind tunnel was first removed such that there were no physical obstructions between the camera and the test vehicle. An airflow rate of 5m/s was used to profile the test vehicles at steady state, as this was the maximum airflow rate obtainable in the wind tunnel without the central panel in place. Figure 4.3.1.1a shows the apparatus setup for data capture using the **ThermaCAM** camera. The camera was held manually at a working distance of approximately 500mm from the profiled specimen such that the correct focus was obtained. After data for all power levels had been captured, the results were then analysed to find the respective temperatures of the heater and carrier chips. Table 4.3.1.1 displays the recorded power levels against the 4-point resistance of the aluminium track.

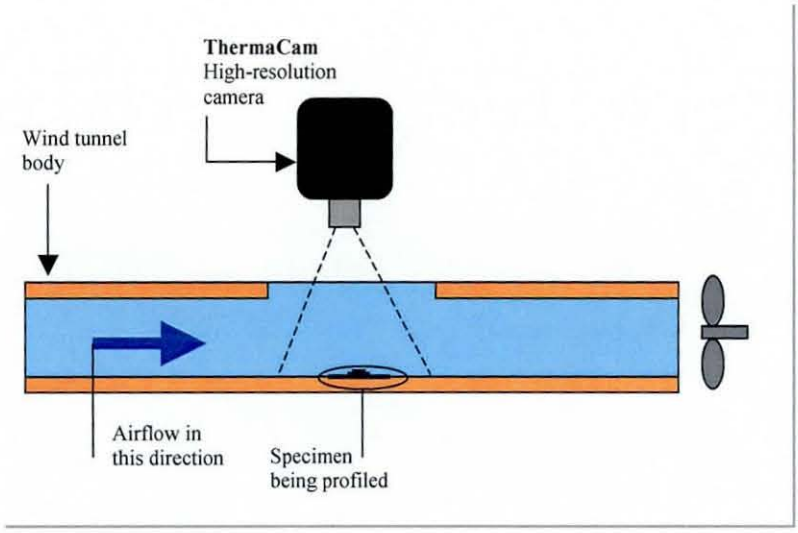


Figure 4.3.1.1.a: *Schematic of high-resolution camera set up as used for thermal profiling (not drawn to scale)*

Results

Test vehicle 003 was one of the samples used to obtain a thermal profile for the MCMs on FR4 substrate, therefore a direct temperature correspondence to the recorded 4-point resistance could be obtained. The observed resistance changes for chip 003 in section 4.2.2 were used to obtain the maximum temperature reached within the assemblies for a particular power level. The observed resistance was recorded and the following conversion was used.

$$T_r = (R_{obs} - R_{RT}) \times 44.91 + \boldsymbol{RT}$$ [1]

Key:
T_r = Predicted temperature
R_{obs} = Observed resistance
R_{RT} = Room temperature resistance (6.194Ω in this case)
44.91 = Observed temperature change for one Ω rise in resistance.
RT= Room temperature (in this case 26°C)

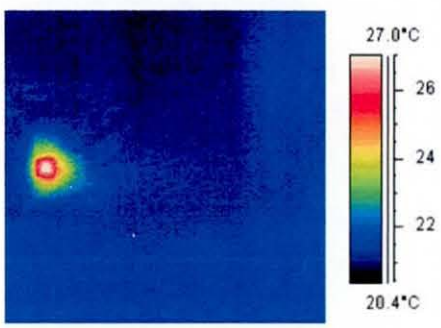
The thermal profiles of the MCM on FR4 were obtained at various power levels ranging from 0.065W to 1.4W and are displayed in table 4.3.1.1a. It can be seen that the maximum temperature of the assembly varied from 27.7°C through to 130.6°C depending on the power level used, with an ambient temperature of 26°C. The temperatures were obtained by noting the 4-point resistance value obtained for a specific power level and using the above formula to convert the resistance into temperature. The thermal profiles of the assembly at given power levels are also shown. For all the plots obtained, the heater and carrier chips are clearly identifiable as they each exhibit uniform and distinct temperatures according to the scale. For the substrate however, the warmer area is only that which is very close to the MCM while

[1] This formula was only valid when test vehicle 003 was profiled

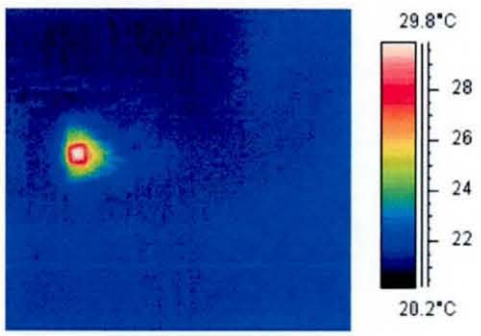
the remaining areas can be seen to be at room temperature. Table 4.3.1.1b shows the temperatures obtained form the thermal profiles, while figure 4.3.1.1b shows the temperatures of the bodies from the thermal images plotted against the temperatures obtained from the 4-point resistance measurements. From examining the 4-point resistance line a linear change with temperature and power levels can be seen. The deviations in the values were due to inconsistencies when positioning the camera (the camera was hand-held and therefore subject to human error i.e. distance, angle etc.) therefore a line of best fit was superimposed on the graph. While the absolute values obtained from the high-resolution camera may be subject to error, it was found that there was generally good agreement with the overall temperature distributions.

Image No	Input Voltage	Input Current (Amps)	Input Power (Watts)	4-point Resistance (Ω)	Corresponding Temperature / °C
1	1.31	0.049	0.0649	6.234	27.7
2 (a)	1.61	0.61	0.0974	6.289	30.2
3 (b)	1.74	0.065	0.113	6.321	31.7
4 (c)	2	0.075	0.15	6.378	34.3
5 (d)	2.25	0.085	0.19	6.453	37.6
6 (e)	2.5	0.094	0.234	6.527	40.9
7 (f)	2.75	0.103	0.282	6.612	44.8
8 (g)	3	0.113	0.337	6.705	48.9
9 (h)	3.25	0.121	0.395	6.813	53.8
10 (j)	3.5	0.13	0.455	6.924	58.8
11 (k)	3.75	0.139	0.519	7.027	63.4
12 (l)	4	0.148	0.59	7.154	69.1
13 (m)	4.25	0.157	0.665	7.275	74.5
14 (n)	4.5	0.165	0.743	7.407	80.4
15 (p)	4.75	0.174	0.827	7.558	87.2
16 (q)	5	0.182	0.91	7.702	93.7
17 (r)	5.25	0.191	1.003	7.850	100.3
18 (s)	5.5	0.199	1.095	8.011	107.6
19 (t)	5.75	0.208	1.193	8.167	114.6
20	6	0.216	1.293	8.352	122.9
21	6.25	0.224	1.4	8.525	130.6

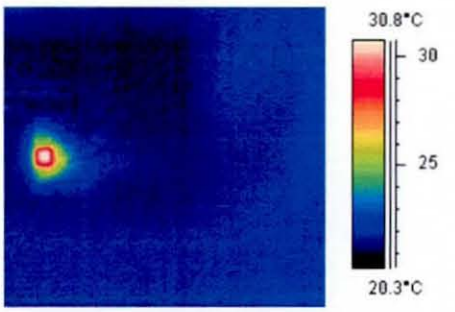
Table 4.3.1.1a
Power input and resultant4-point resistance measurements. Corresponding thermal imaging camera plots below. The letters in the left hand box refer to the images below and the grey boxes indicate that no image was available. Test vehicle 003 was used for the thermal profiling.



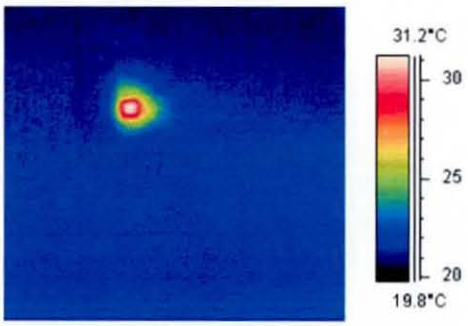
(a) 0.1 W



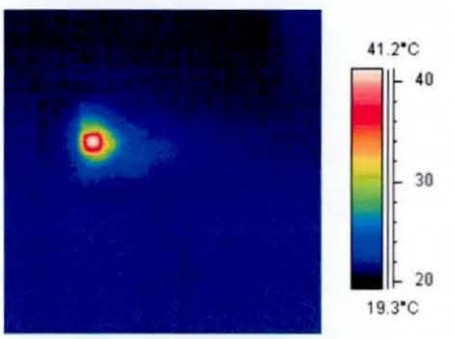
(b) 0.11 W



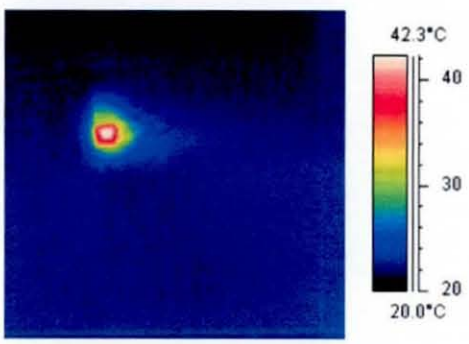
(c) 0.15 W



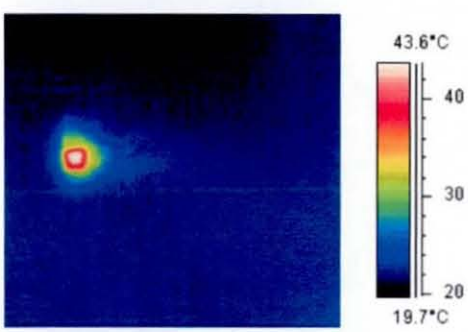
(d) 0.19 W



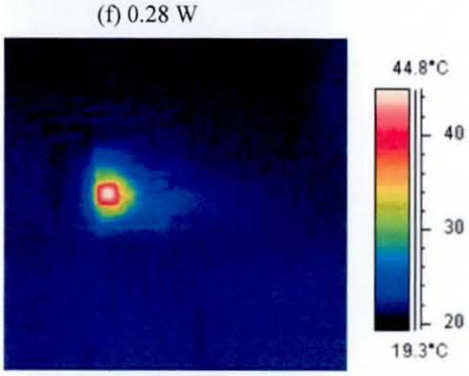
(e) 0.23 W



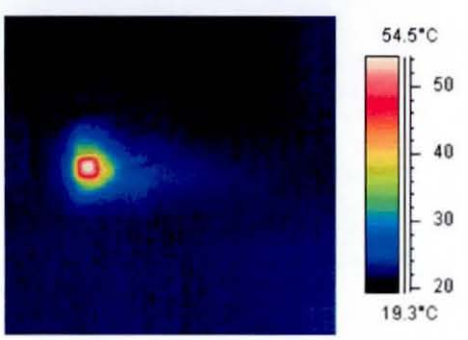
(f) 0.28 W



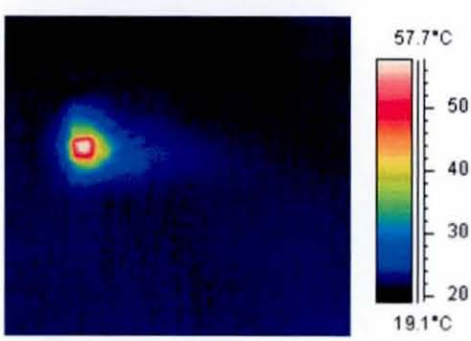
(g) 0.33 W



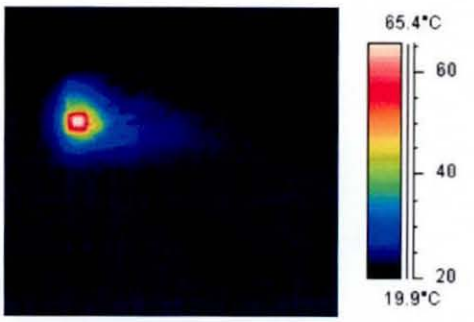
(h) 0.4 W



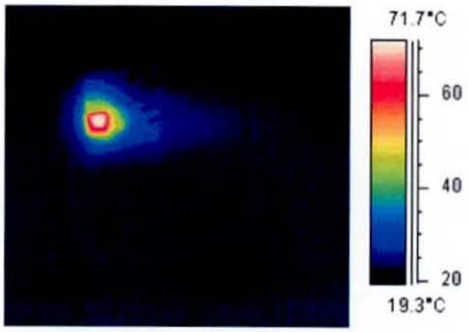
(j) 0.46 W



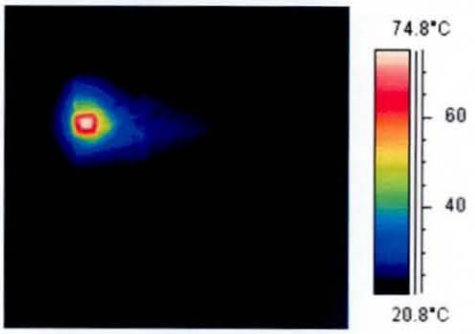
(k) 0.52 W



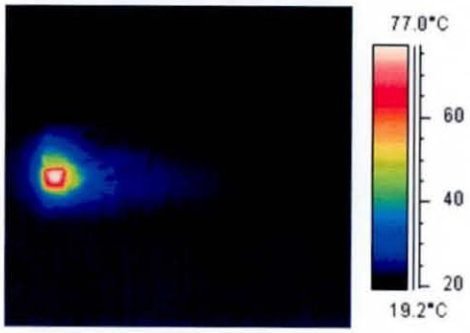
(l) 0.59 W



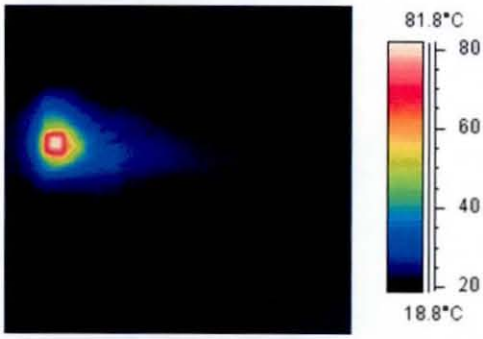
(m) 0.67 W



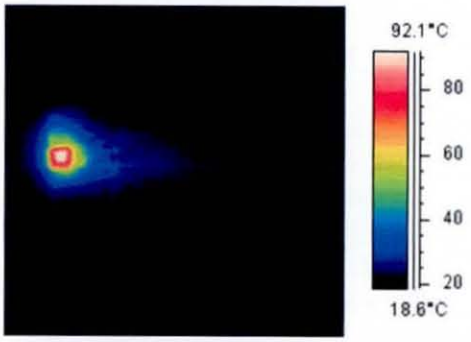
(n) 0.74W



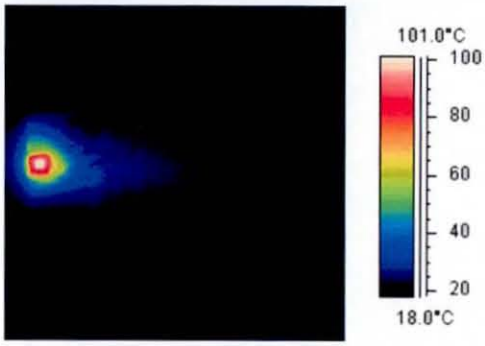
(p) 0.83 W



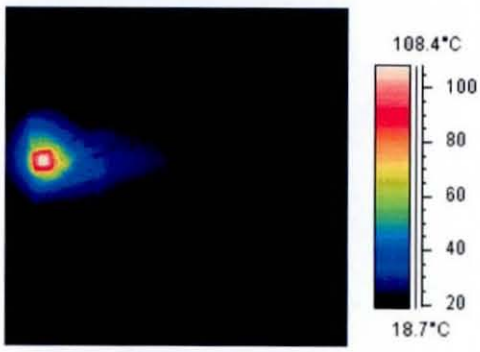
(q) 0.91 W



(r) 1.W



(s) 1.1 W



(t) 1.19 W

Power Level (W)	Heater Chip (°C)	Carrier Chip(°C)	Substrate temperature near MCM (°C)
0.1	27.0	26.0	24.2
0.11	29.8	28.5	25.0
0.15	30.5	29.2	25.2
0.19	31.2	29.7	25.4
0.23	41.2	38	31
0.28	42.3	39.	32
0.33	43.6	40.	33
0.4	44.8	41.	34
0.46	54.5	49.	38
0.52	57.7	52.	40
0.59	65.4	59.	42
0.67	71.7	65.	49
0.74	74.8	68.	50
0.83	77.0	70.	52
0.91	81.8	73.	55
1.00	92.1	81.	57
1.1	101.	91.	60
1.19	108.	100	63

Table 4.3.1.1b:
Temperatures of each body in the assembly at given power levels. The temperatures were obtained by matching the colours of the body with the corresponding scale. The heater and carrier chips both showed uniform temperature. The substrate temperature near the MCM is defined as the immediate area surrounding the MCM that can be seen as a hotter temperature than the remainder of the substrate.

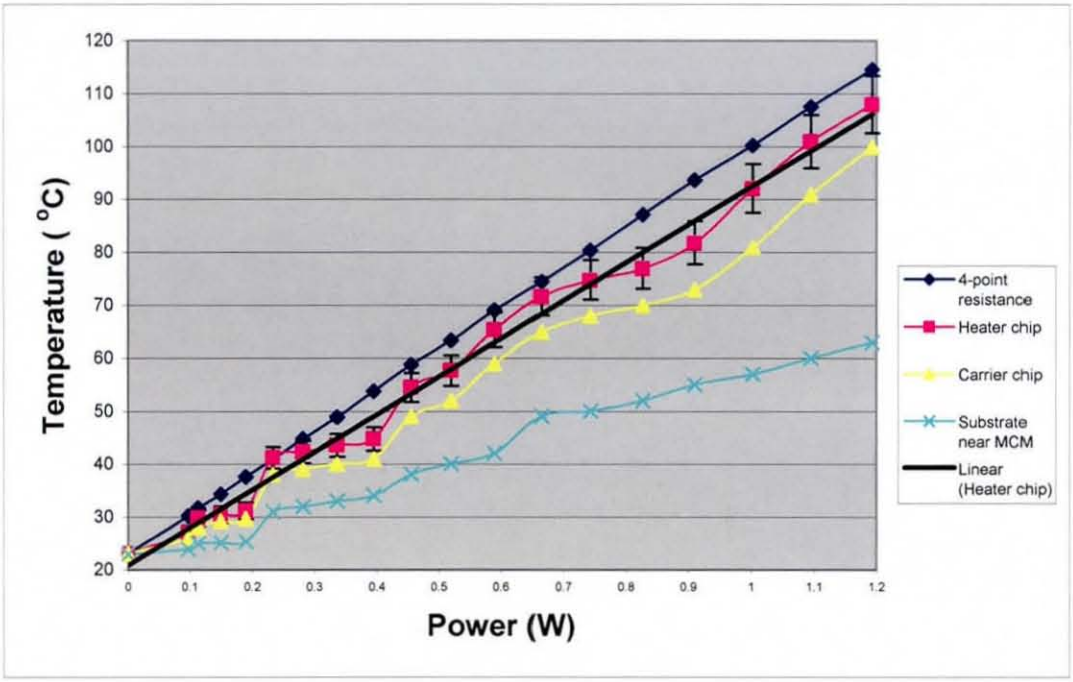


Figure 4.3.1.1b:
Graphical representation of the data in table 4.3.1.1b, also included is the 4 point resistance measurements and a "line of best fit" for the temperature of the heater chip (table data)

Table 4.3.1.1c shows the maximum temperatures obtained from the thermal imaging camera expressed as a percentage of the 4-point resistance measurements. It can be seen that for greater power levels that the temperature change recorded from the thermal imaging camera persistently predicts between 80 and 95% of the maximum temperature recorded from the 4-point measurements, the discrepancies were attributed to the thermal camera having no fixed or defined position (the camera was held by hand so the position used to capture the image may have varied somewhat).

Image No	*Input Power (Watts)*	*4-point Resistance (OHMS)*	*Corresponding Temperature Rise (°C)*	*Maximum temperature rise from image*	*%age maximum temp recorded on camera*
2 (a)	0.0974	6.289	10.2	7	68%
3 (b)	0.113	6.321	11.7	9.8	83%
4 (c)	0.15	6.378	14.3	10.5	73%
5 (d)	0.19	6.453	17.6	11.2	63%
6 (e)	0.234	6.527	20.9	21.2	99%
7 (f)	0.282	6.612	24.8	22.3	90%
8 (g)	0.337	6.705	28.9	23.6	81%
9 (h)	0.395	6.813	33.8	24.8	73%
10 (j)	0.455	6.924	38.8	34.5	88%
11 (k)	0.519	7.027	43.4	37.7	86%
12 (l)	0.59	7.154	49.1	45.4	92%
13 (m)	0.665	7.275	54.5	51.7	94%
14 (n)	0.743	7.407	60.4	54.8	90%
15 (p)	0.827	7.558	67.2	57	85%
16 (q)	0.91	7.702	73.7	61.8	83%
17 (r)	1.003	7.850	80.3	72.1	90%
18 (s)	1.095	8.011	87.6	81	92%
19 (t)	1.193	8.167	94.6	88	93%

Figure 4.3.1.1c:
Comparison of the maximum temperature obtained from the thermal image camera compared with the 4-point resistance measurement. The variation was due to the camera having no fixed position.

4.3.1.2: Assemblies on Copper Substrates

Apparatus and Setup

Unfortunately, the **ThermaCAM** high-resolution camera was unavailable to thermally profile the MCMs on copper boards so the IRISYS low-resolution camera had to be used. When using the low-resolution camera the arrangement shown in figure 4.3.1.2a was implemented; the manufacturers recommendation that the sample was positioned 100mm away from the lens was adhered to. When using the IRISYS camera, there was greater difficulty in distinguishing the heater chip from the carrier chip, due to the low resolution of the camera. However, the resolution of 2.2x2.2mm per pixel was considered to be sufficient to identify the major parts of the test vehicle. The heater chip was thought to be identifiable as the hottest pixel in any of the obtained images and it was anticipated that the copper board would form the majority of the image and it would be at uniform temperature. When the thermal images were captured, some pixels showed some inevitable overlap between the heater and carrier chip and careful interpretation was required. Figure 4.3.1.2b displays four possible pixel allocations using the IRISYS camera: **i** represents rotational misalignment between the camera and the MCM, in this case the display would bear little resemblance to the MCM; **ii** shows the heater chip covered evenly over 4 different pixels, if this were the case then there would be no clearly identifiable maximum temperature pixel; **iii** the pixel located at the dead-centre of the heater chip, in this case all surrounding pixels would display similar temperatures to each other; **iv** the pixel had fully located the heater chip but not on the centre, therefore the surrounding pixels would exhibit different temperatures i.e. some pixels would show greater temperature difference than others. The final scenario was considered the most likely situation and the best means of obtaining the temperature of the carrier chip, therefore all the pixels surrounding the maximum pixel in each image were compared to find the one that exhibited the greatest temperature difference from the maximum and this was assumed to represent the temperature of the carrier chip. The IRISYS camera could be operated remotely therefore once it was correctly positioned it was not moved until all the tests were completed.

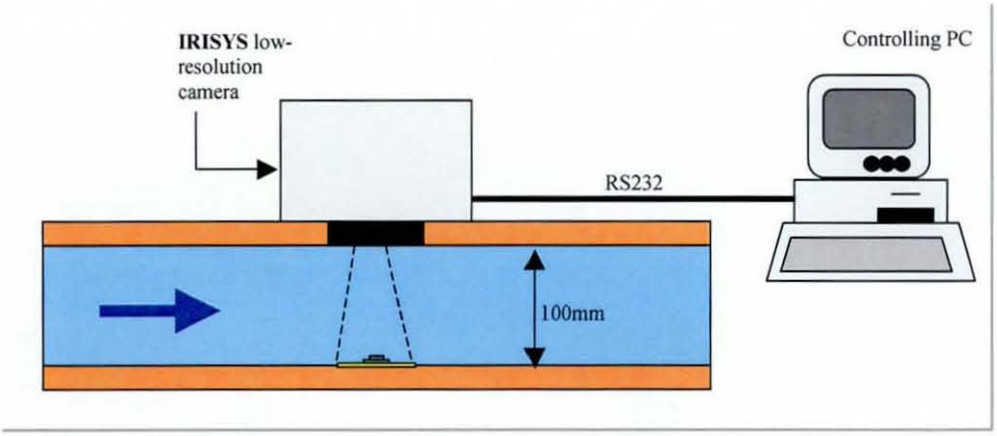


Figure 4.3.1.2a: *Schematic of low-resolution camera set up as used for thermal profiling (not drawn to scale)*

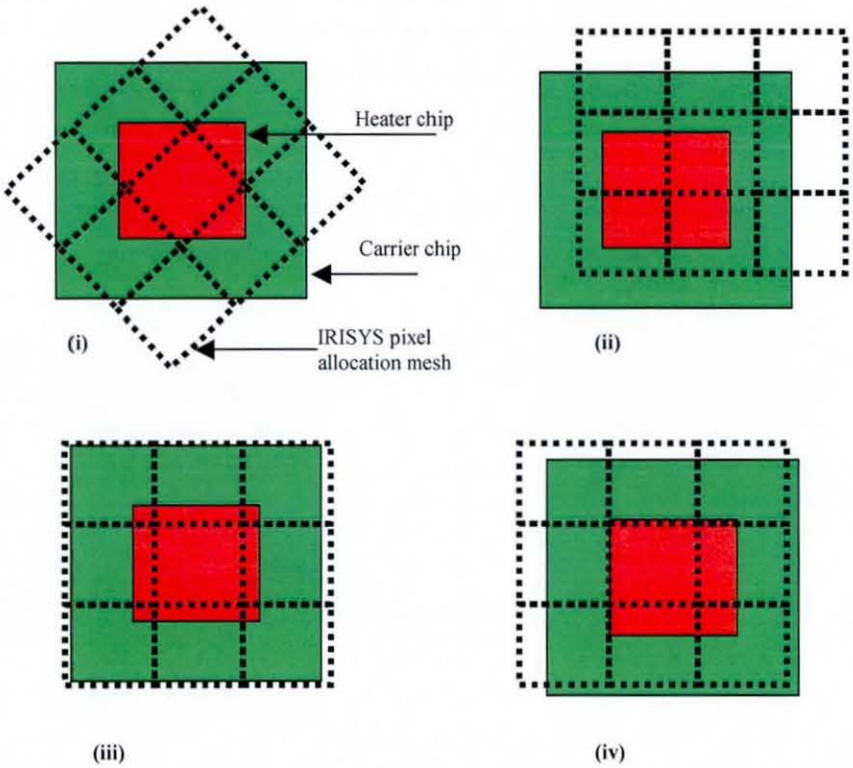


Figure 4.3.1.2b
Description of the 4 different potential pixel allocations, Scenario (i) was anticipated to be the most likely method of identifying the heater and carrier chip temperatures.

Results

Test vehicle A0 was used to test the performance of the MCM on copper substrate. Upon initial power-up of the device, it became apparent that the copper substrate test vehicles had a different temperature distribution to those on FR4, which allowed greater power levels to be implemented. Consequently the power ranges implemented were from 0.075 through to 3.3W as shown in table 4.3.1.2a. The formula below was based on the observed 4-point resistance changes for assembly A0 in section 4.2.2 and was used to convert the obtained resistance reading to the corresponding temperature.

$$T_r = (\Omega_{obs} - 5.825) * 48.25 + RT \quad [2]$$

Key:
T_r = Predicted temperature
Ω_{obs} = Observed resistance
5.825 = Resistance at room temperature
48.25 = Observed temperature change for 1 Ω rise in resistance.
RT = Ambient temperature (23°C)

To obtain the thermal profiles, the camera was positioned in a manner such that scenario D in figure 4.3.1.2b was apparent, potentially allowing the clearest temperature of the heater and carrier chips to be obtained. Once correctly positioned, the camera was not moved until all the thermal images were captured for different power levels. Figure 4.3.1.2c is an image capture of the assembly powered up and the pixels used to obtain the heater and carrier chip temperatures are labeled. From observing the plot, it can be seen that the substrate is at a uniform temperature, so some area away from the MCM was chosen for the substrate temperature to prevent potential overlap with the MCM. The thermal images corresponding to the power levels from table 4.3.1.2a are also shown and the temperatures obtained from the images are shown in table 4.3.1.2b, it can be seen that at high(er) power levels the areas of the assembly become clearer where the MCM can clearly be distinguished from the substrate. While the heater and carrier chips are identifiable, there is no indication whether the temperature across them is uniform as the allocated pixel area is a large percentage of the MCM dimension, so fine detail cannot be obtained.

Figure 4.3.1.2d shows the temperatures from the 4-point resistance readings plot with the temperatures obtained from the thermal imaging camera. It can be seen from both profiles that a linear relationship exists with the power level and the resultant temperature. Towards the lower power levels, it can be seen that the temperature of the heater chip (and the carrier chip and the substrate) are all greater than that indicated by

the 4-point resistance measurements, this error was attributed to *noise* contributing to the image. At these low power levels in the corresponding images, it was difficult to identify the heater chip as the camera has a low temperature range to operate in.

Image No	Input Voltage	Input Current (Amps)	Input Power (Watts)	4-point Resistance (Ω) (A0)	Corresponding Temperature (°C)
1 (a)	1.5	0.065	0.0975	5.903	26.7
2 (b)	2	0.084	0.168	5.939	28.5
3 (c)	2.5	0.104	0.259	5.995	31.2
4 (d)	3	0.122	0.366	6.060	34.3
5 (e)	3.5	0.141	0.492	6.141	38.2
6 (f)	4	0.159	0.636	6.231	42.6
7 (g)	4.5	0.177	0.796	6.328	47.3
8 (h)	5	0.195	0.975	6.432	52.3
9 (j)	5.5	0.212	1.166	6.552	58
10 (k)	6	0.230	1.38	6.691	64.8
11 (l)	6.5	0.248	1.612	6.833	71.6
12 (m)	7	0.265	1.855	6.974	78.4
13 (n)	7.5	0.282	2.115	7.132	86
14 (p)	8	0.299	2.392	7.309	94.6
15 (q)	8.5	0.316	2.686	7.483	103
16 (r)	9	0.332	2.988	7.659	111.5
17 (s)	9.5	0.348	3.306	7.853	120.8

Table 4.3.1.2a:
Table showing power input and resultant 4-point resistance measurements for the MCM assembly on a copper board. The temperatures were obtained from the 4-point resistance measurements and were converted to temperatures using the graph for A0 in figure 4.2.2

[2] Formula is only valid when test vehicle A0 is used

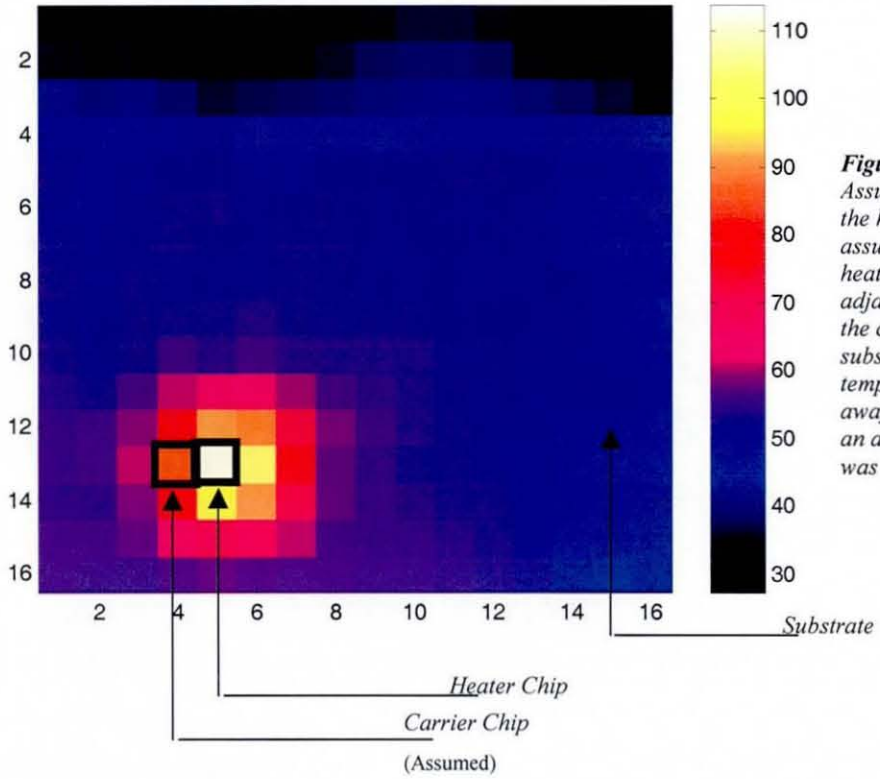
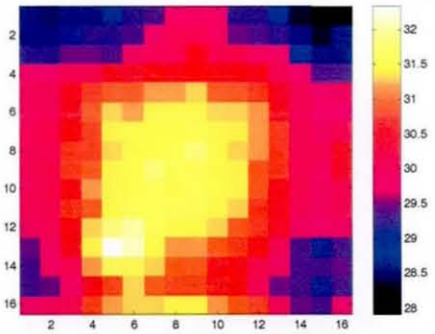
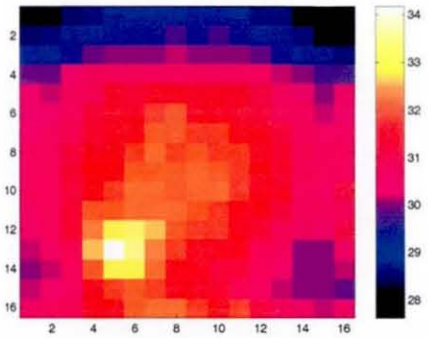


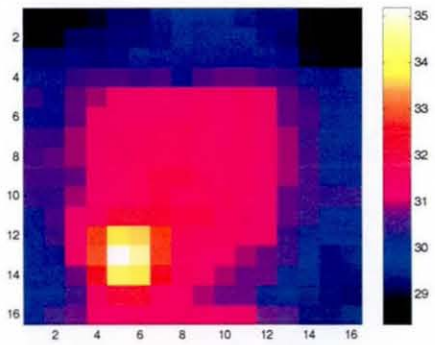
Figure 4.3.1.2c:
Assumed pixel allocation, the hottest pixel was assumed to represent the heater chip while the pixel adjacent was thought to be the carrier chip. As the substrate showed a uniform temperature, a pixel far away was chosen to be sure an adequate representation was obtained.



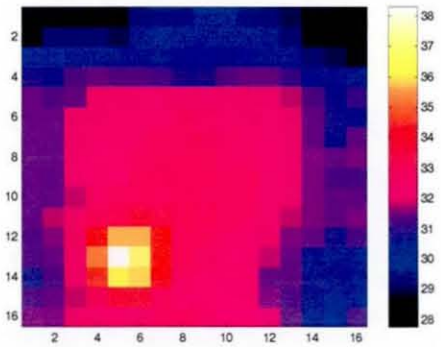
(a) Power =0.0975W



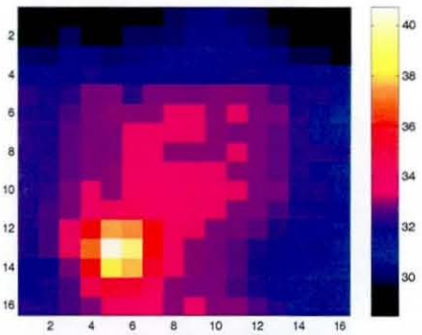
(b) Power =0.168W



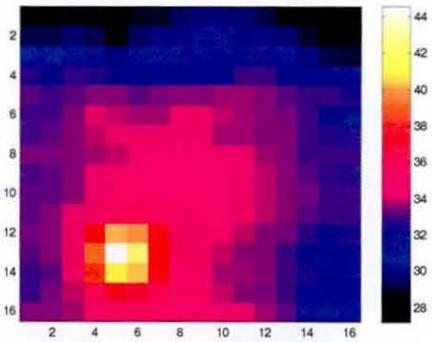
(c) Power =0.259W



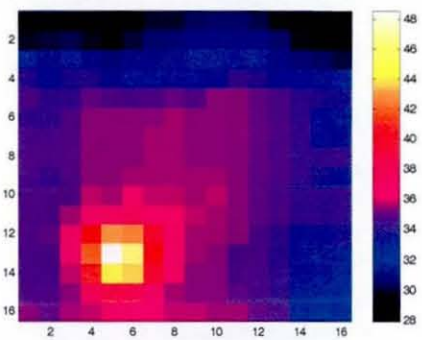
(d) Power =0.366W



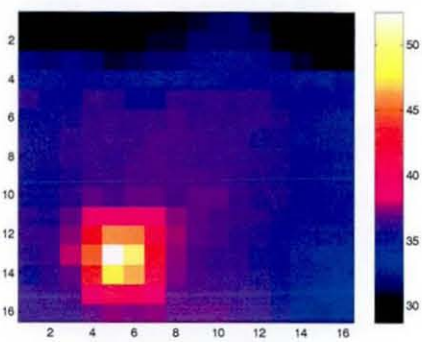
(e) Power =0.492W



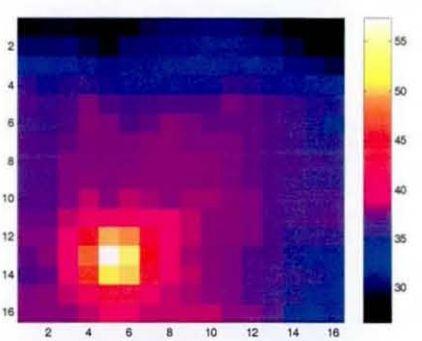
(f) Power =0.636W



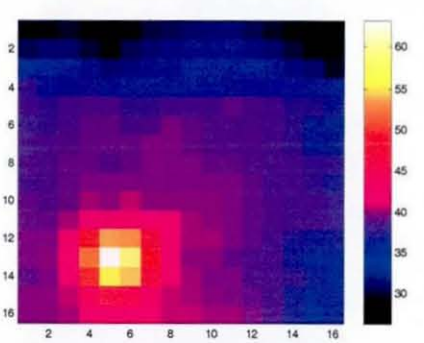
(g) Power =0.796W



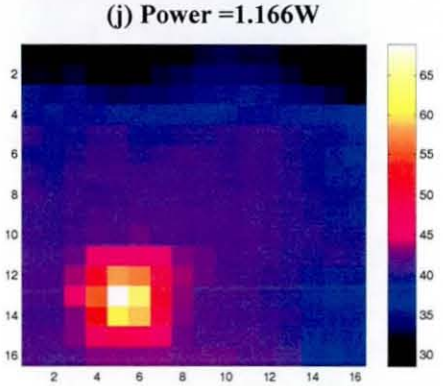
(h) Power =0.975W



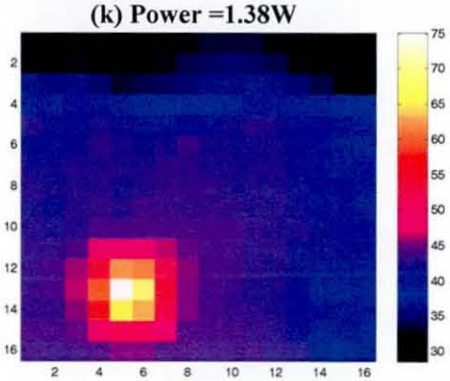
(j) Power =1.166W



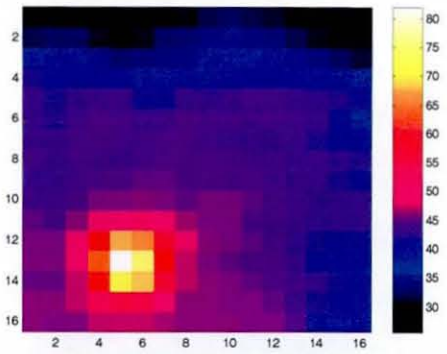
(k) Power =1.38W



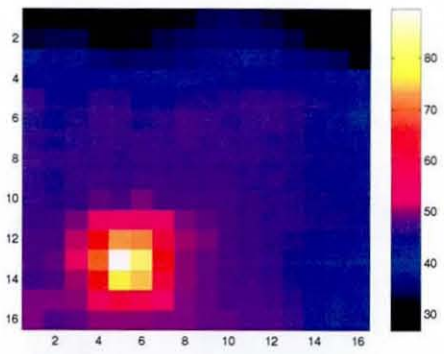
(l) Power =1.612W



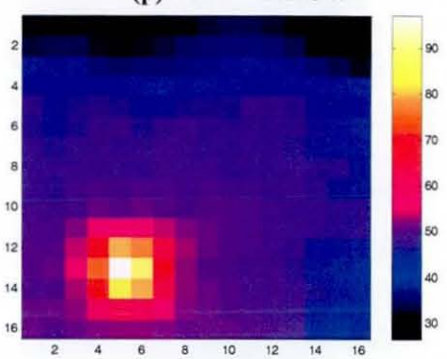
(m) Power =1.855W



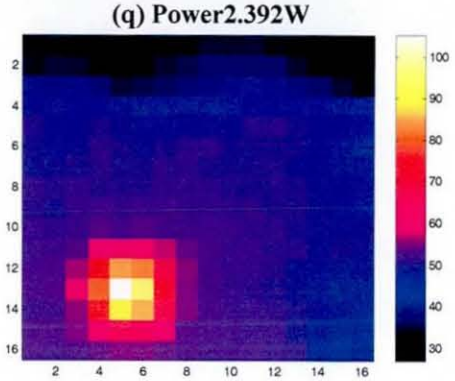
(p) Power =2.115W



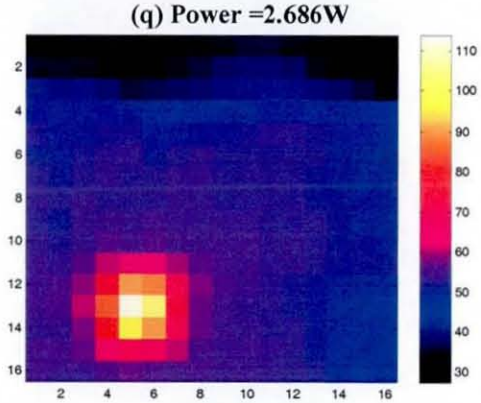
(q) Power2.392W



(q) Power =2.686W



(r) Power =2.988W



(s) Power =3.306W

Power Level (W)	Heater Chip (°C)	Carrier Chip (°C)	Substrate Temperature (°C)
0.0975	32	31	29
0.168	34	32.	30
0.259	36	33	31
0.366	38	35	32
0.492	41	37	33
0.636	44	40	34
0.796	48	43	35
0.975	53	46	36
1.166	57	49	37
1.38	63	53	40
1.612	69	56	42
1.855	75	60	43
2.115	82	65	45
2.392	90	70	47
2.686	97	75	49
2.988	105	80	51
3.306	113	86	53

Table 4.3.1.2b:
Table showing the temperature readings for each body, given the assumed pixel allocation described in figure 4.3.1.2b

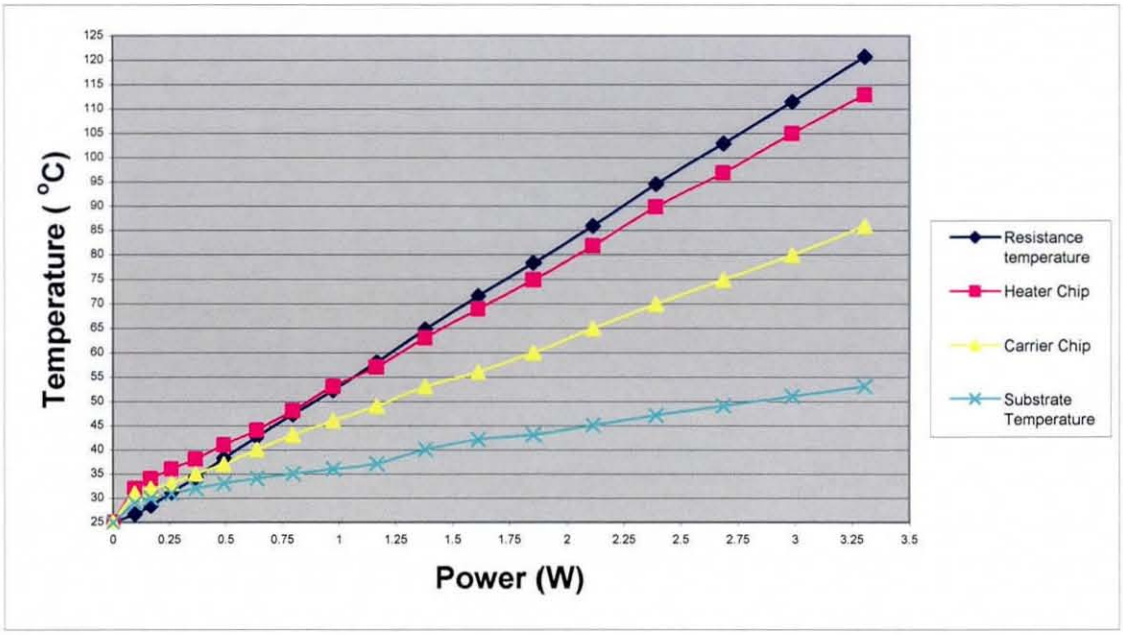


Figure 4.3.1.2d:
Plot of the body temperatures obtained from table 4.3.1.2b. Also included is the temperature obtained from the 4-point resistance measurements. Comparing with the results from the high resolution camera, it can be seen that the low-resolution camera provides a better agreement of data as it was secured in the same position for all the recordings, such that once it was set up, all the recordings were made with the camera in a consistent position. This was in contrast to when the high-resolution camera; where the position was held by hand, and had to be focused when each new power level was set; therefore the camera position is likely to have changed when different measurements were taken.

4.3.2: Transient Thermal Data Capture

The IRISYS low resolution thermal imaging camera was used to capture the transient thermal profiles experienced by the MCMs during power cycling, therefore the setup described in figure 4.3.1 was implemented. The assemblies were first connected to the power supply via the timer such that the ON/OFF periods could be controlled. As the tunnel was sealed (i.e. there were no large open gaps) this allowed for the maximum airflow rate of 10m/s to be utilised. When the air speed was set to the correct level the test vehicle was powered and allowed to cycle until it stabilised; this typically meant leaving it to complete a minimum of 4 cycles. Once ready, the camera was turned on and 3 or 4 complete power cycles were recorded. After the last cycle had completed a new power level was set and the process was repeated.

4.3.2.1: Transient Image Results (FR4 substrate)

The transient images were initially displayed as a dynamic thermal plot. Figure 4.3.2.1 shows one such plot at a maximum value, as well as the corresponding "real" device. In accordance with the procedure described in section 4.3.1b, the hottest pixel was initially located, and then pixels adjacent to it either horizontally or vertically were compared to see which exhibited the greatest temperature difference. Referring back to the 4 pixel allocation scenarios described in section 4.3.1c, it was clear that scenario d was the case with the obtained thermal plots as the *upside* and *right hand side* pixels demonstrated the greatest temperature difference. Figures 4.3.2.1b, c and d show the labeled axes, the pixel allocation for the *right hand side* and the *upside* respectively. Therefore the pixel values along the *upside* and *right hand side* axis were plotted in a time versus temperature graph to obtain the transient profile of the whole assembly along these pixels.

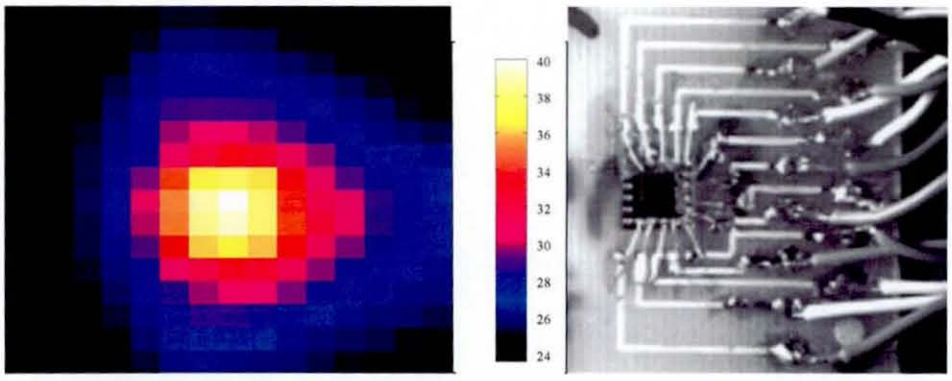
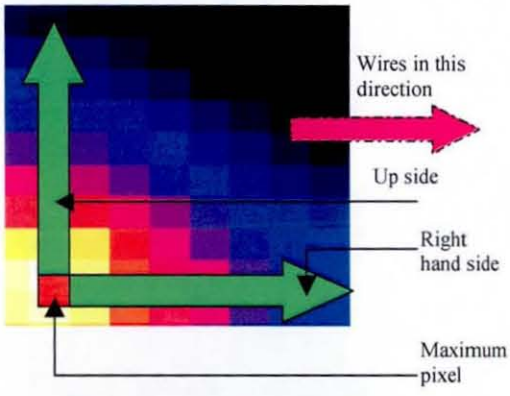
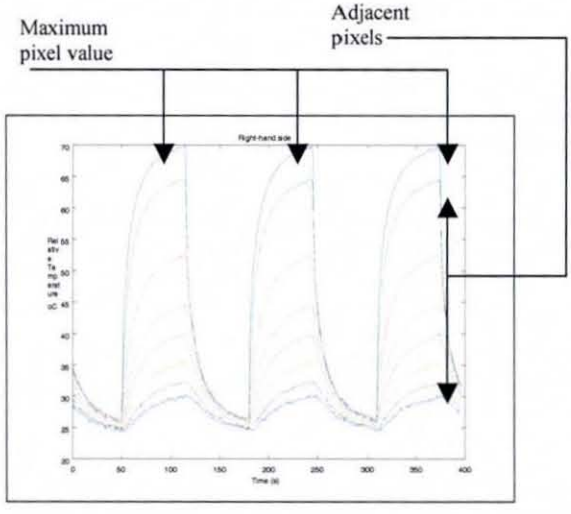


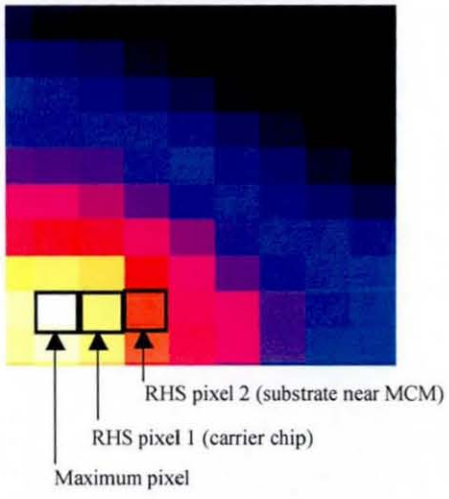
Figure 4.3.2.1a:
Examples of a low resolution thermal plot. Also included is the original chip showing the orientation

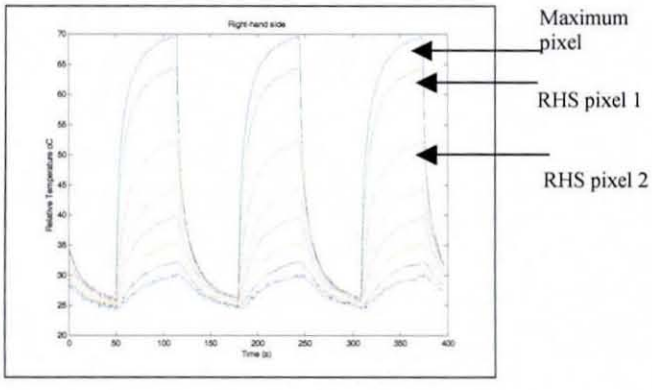
Figure 4.3.2.1b:
Demonstration of how pixel allocation relates to the graphs displayed, (i) shows a generic description, (ii) shows the right hand side allocation while (iii) shows the up-side allocation



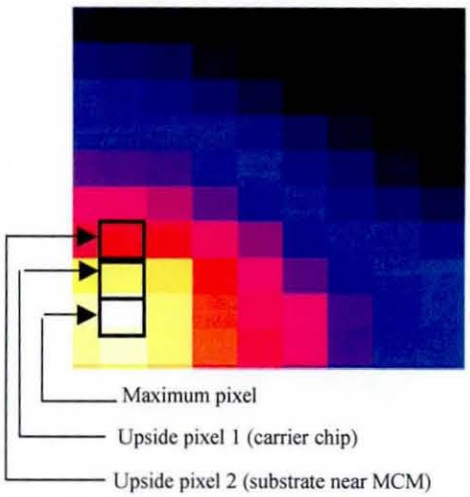


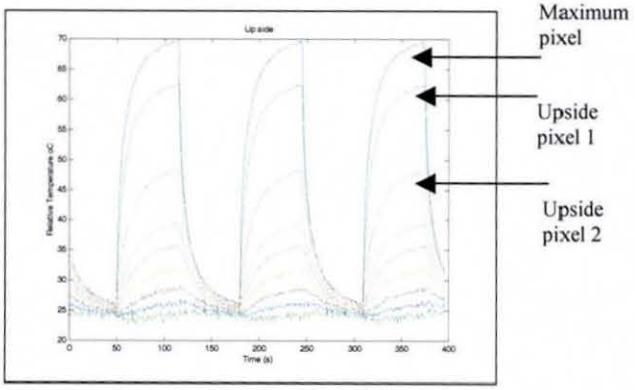
4.3.2.1b(i)





4.3.2.1b(ii)

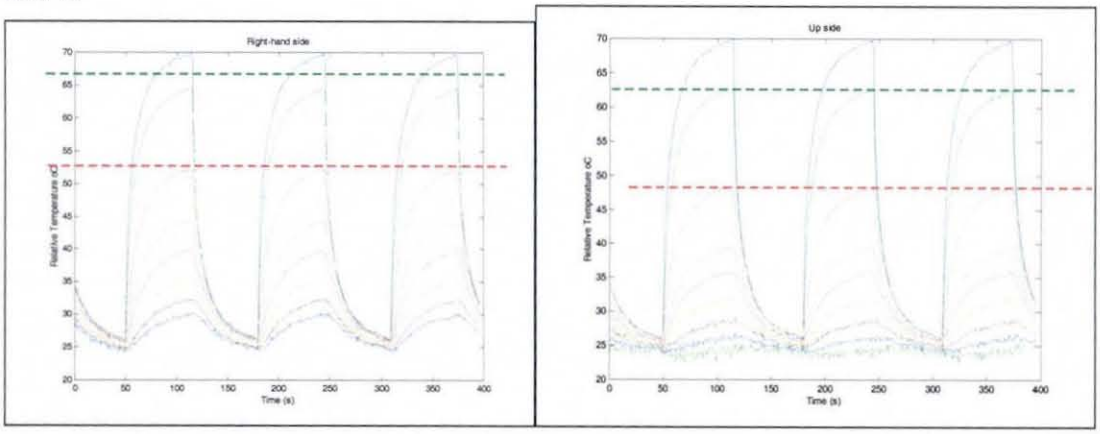




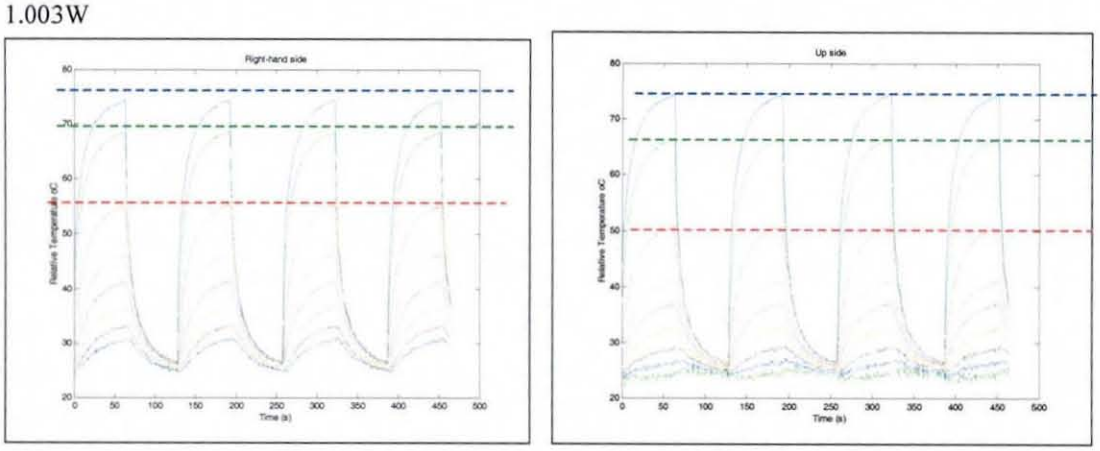
4.3.2.1b(iii)

The results of power cycling between 0.9 W and 1.4W are displayed. Superimposed on the plot are lines so the estimated temperatures of the heater and carrier chips can be obtained.

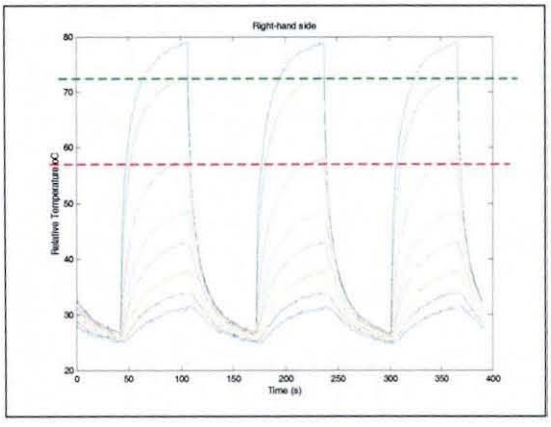
0.91 W

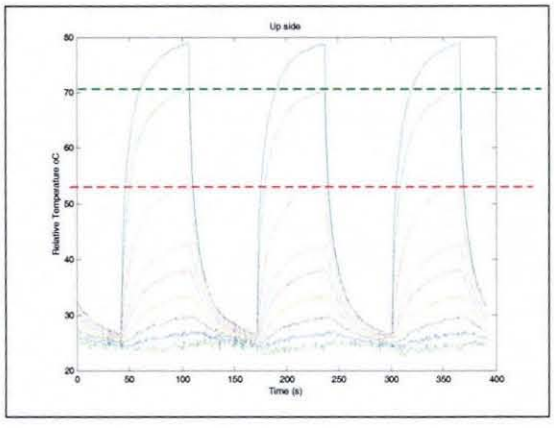


1.003W

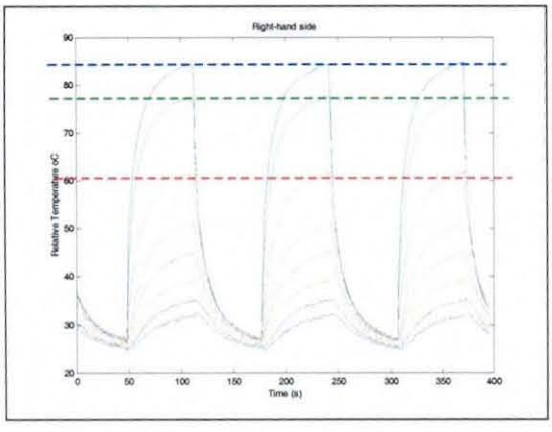


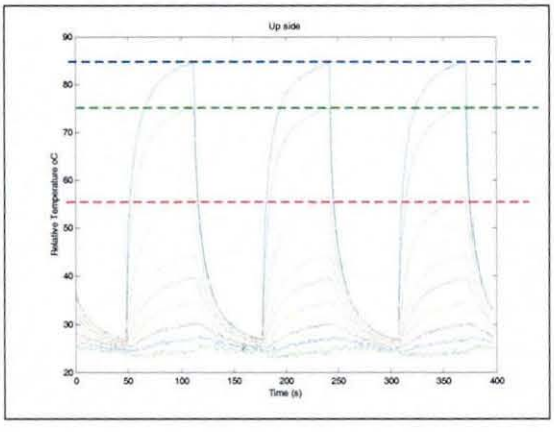
1.09W



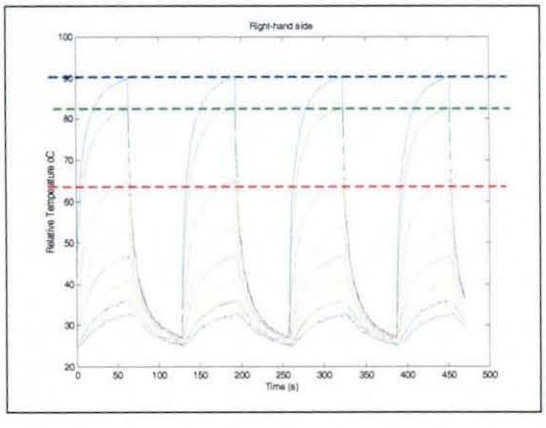


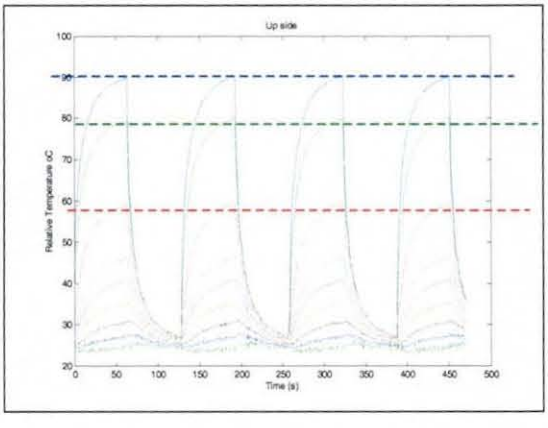
1.19W



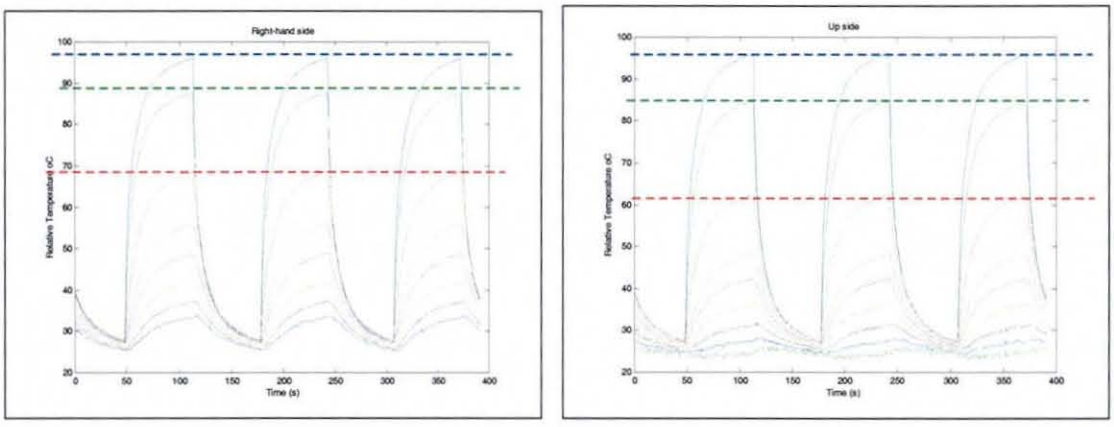


1.29W





1.4W



From figure 4.3.2.1b, there were a total of 9 pixels used for the *upside* of the chip profile and 8 pixels for the *right hand side* profiles. It could be seen that the shape of the graphs for the different power levels was consistent across the power range considered. All curves exhibited an initial steep temperature rise, which subsequently reduced the rate of change as the assembly approached steady state. From the graphs, it can be seen that at the peak temperatures (i.e. time = 64, 192 & 320 seconds) there was a greater degree of temperature difference in the *upside* graphs than on the *right hand side*. For the *right hand side* profile, it could be seen that all the connecting wires were to the right of the assembly. Also from the thermal plot in 4.3.2.1a it was clear that these wires demonstrated an increase in temperature as the power cycle was on. The heat was found to spread further out across the substrate, which was attributed to the connecting wires that assisted in the heat conduction. The small temperature rise observed in the peripheral pixels was attributed to the poor thermal conductivity of the FR4 substrates used. The implication of this was that the temperature changes within the FR4 substrate were limited to areas local to the MCM and therefore very little temperature change was noted on the board parts further away from the MCM.

4.3.2.2: Transient Thermal Data for MCMs on Copper Substrates

As with the FR4 substrate specimens, the results from the assemblies on copper substrates are displayed as a thermal plot in figure 4.3.2.2. The values recorded for each pixel were subsequently extrapolated and the graphs were plotted. From the plots obtained, it could be seen that the greatest temperature differences between the hottest pixel were the *upside* and the *right hand side* pixels. Therefore an average of these two

pixels was used as an estimate for the temperature of the carrier chip. It can be seen that the graphs lack the "smoothness" of the previous FR4 curves, which was attributed to background noise picked up by the camera due to the lower temperature range.

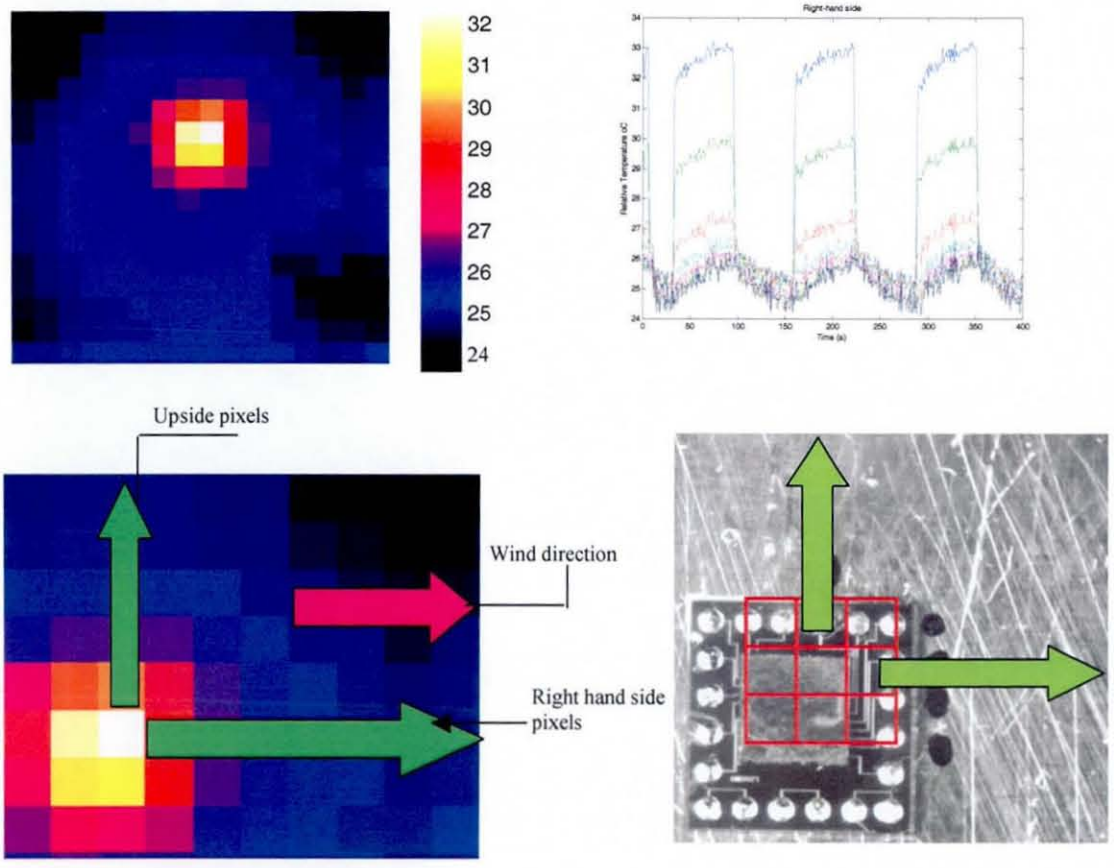
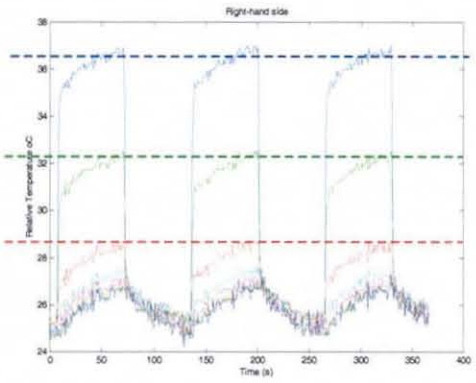


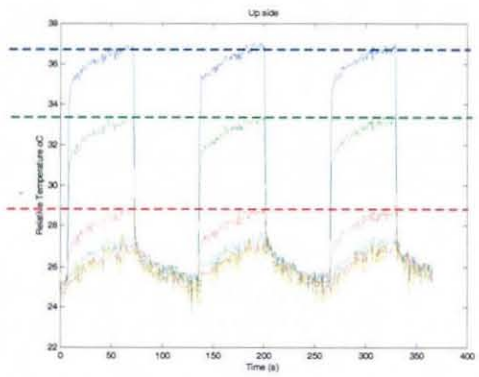
Figure 4.3.2.2a:
Examples of a low resolution thermal plot. Also included is the original chip showing the orientation. The darker regions in the corner of the plots are where tape was used to position/secure the chip in the wind tunnel.

Also demonstration of how the pixel allocation relates to the graphs

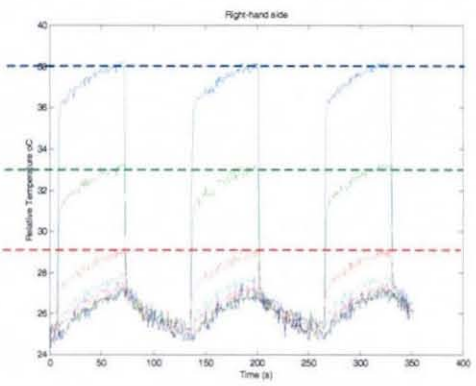
The results of power inputs from 0.92 W to 1.44W are displayed.

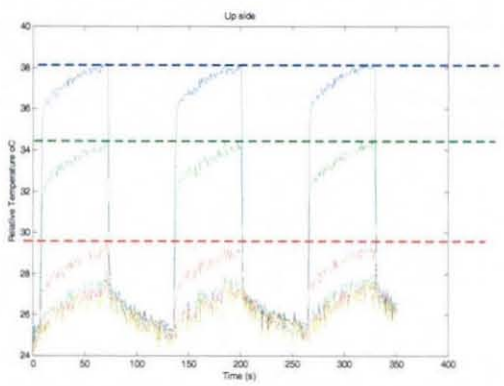
0.92W



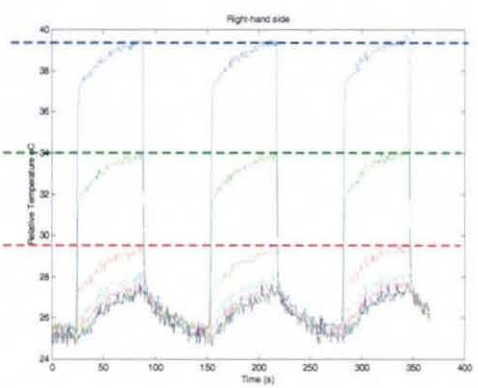


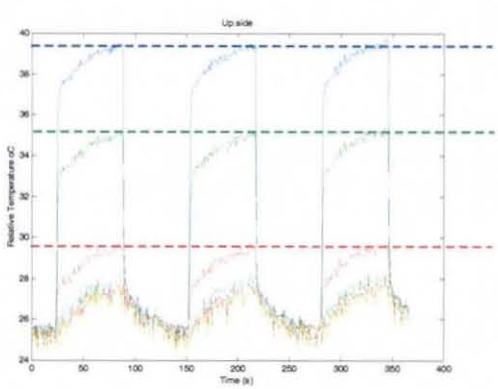
1.002W



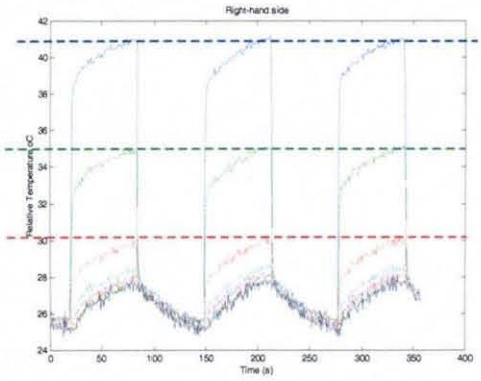


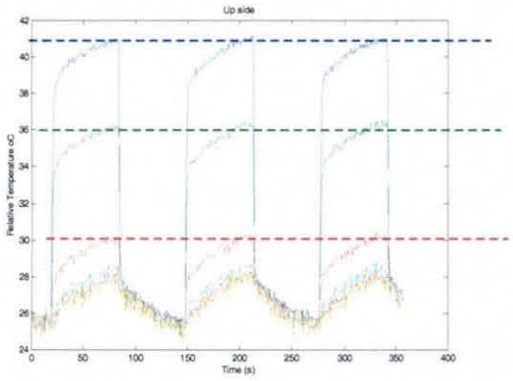
1.12W



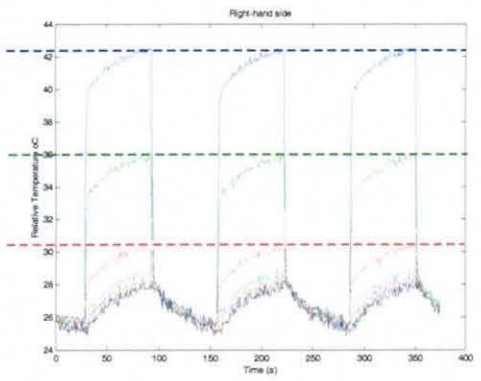


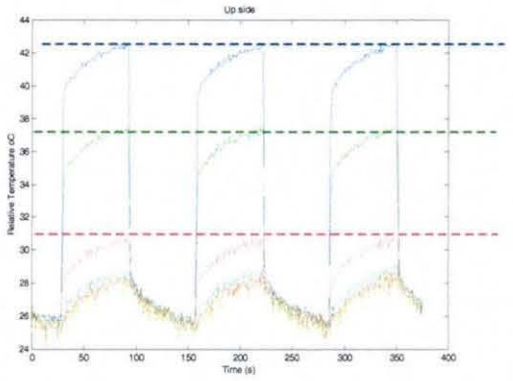
1.22W



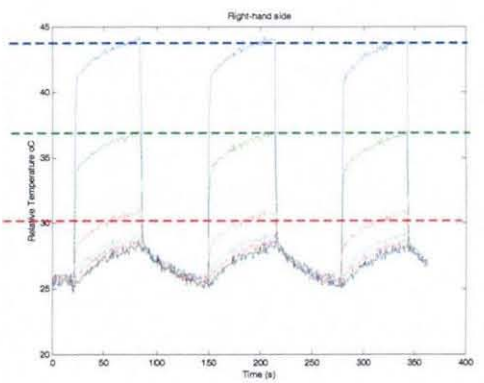


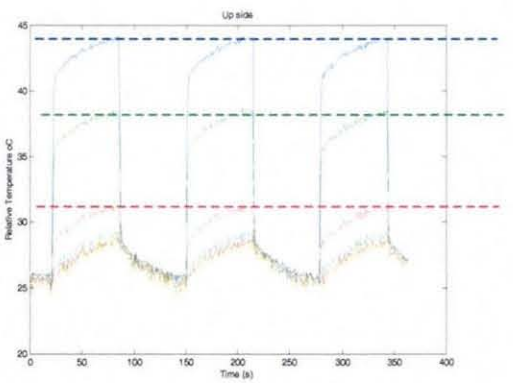
1.32W





1.44W





A total of 6 pixels were used for the *upside* profile, while 8 pixels were used in the right hand side profile. All the graphs showed a concentration of lines further away from the maximum pixel representing the copper substrate, implying that a uniform temperature existed across the substrate. The general profile of the curve is seen to have a very steep initial gradient that abruptly changes to a more gentle increase as the duration of the cycle continues and a similar trend is noted on the OFF cycle. The initial steep temperature rise is attributed to only the MCM assembly heating up at that

specific moment; that is to say the copper substrate has yet to be influenced by the heat generated within the chip. In fact it will be shown later that the behaviour/profile of the MCM at this moment is identical to that of the samples on the FR4 substrate. When the gradient of the graph changes, this was thought to be the influence of the copper substrate conducting the heat away from the MCM. It was noted that the temperature rise of the copper substrate was not apparent until the change in gradient of the MCM pixels. It was also noted that the temperature gradients of all pixels recorded then followed a similar temperature profile. As expected, it was found that the maximum temperature of the MCM during the ON cycle was still significantly greater than that of the copper substrate. Therefore, when the cycle was turned OFF the initial temperature change was great as both the copper substrate and the convected air both play important cooling modes in the MCM; the air played a great part due to the small thermal mass of the MCM and the copper board acted as a heat spreader. However, once the chip reached the temperature of the copper board, the rate of cooling slowed to match that of the rest of the assembly.

4.4: Discussion

4.4.1: Comparisons of Assemblies on FR4 and Copper

Steady State

The choice of substrate used in the MCM had a profound influence on the final temperature reached as well as the resultant temperature distribution of the substrate. Based on a maximum operating temperature it can be seen that the MCM on the copper board had the potential to accommodate three times the power input of the equivalent assembly mounted on FR4. In addition, the temperature of the substrate can be seen to be uniform across the copper substrate compared to the case with FR4 where the heat was largely centred around the area of the MCM.

The final temperatures of the carrier chip and the substrate from table 4.3.1.1b and 4.3.1.2b are shown as a percentage of the heater in tables 4.4.1a and b. The temperatures used were all obtained from the thermal images, therefore the

temperature ranges of the maximum/ minimum were used. It can be seen that for the assemblies on FR4, the carrier chip was always between 85 and 90% of the temperature of the heater chip, while the area of the substrate chosen was between 48 and 57% of the heater chip temperature. The inconsistencies of the substrate temperature were considered to be due to the large thermal gradient at the area chosen to record the temperature, resulting in difficulty in ensuring the exact same spot was chosen. For the MCMs mounted on copper board, the carrier chip was at 91% of the heater chip temperature for a low power level (0.0975W) while for 3.3W, the final temperature had decreased to 71% of that from the heater chip. This effect can be seen to a greater extent on the substrate where the temperature had decreased from 75% to 35% of the heater temperature. From examining the transient graphs of the copper assemblies (namely the gradient at the end of the ON cycle) it was apparent that the assemblies on copper substrate had not yet reached steady state, while the assemblies on FR4 board were close to, if not actually at steady state.

Transient

The MCM assemblies both exhibited different transient profiles depending on the substrate used. While the MCM on FR4 shows a smooth decline in the temperature increase as the ON cycle continues, the copper substrate is an almost "box-like" profile in the temperature increase. As the steady state and transient profiles were obtained at different airflow rates, a direct comparison between the two was not possible.

Power Level (W)	Heater Chip (°C)	Percentage temperature carrier chip	Percentage temperature (substrate)
0.1	27	85%	57%
0.11	29.8	88%	51%
0.15	30.5	88%	49%
0.19	31.2	86%	48%
0.23	41.2	85%	52%
0.28	42.3	85%	54%
0.33	43.6	85%	55%
0.4	44.8	85%	56%
0.46	54.5	85%	52%
0.52	57.7	85%	53%
0.59	65.4	86%	48%
0.67	71.7	87%	56%
0.74	74.8	88%	55%
0.83	77	88%	56%
0.91	81.8	86%	57%
1.00	92.1	85%	57%
1.1	101	87%	49%
1.19	108	90%	48%

Table 4.4.1a: *FR4 results Temperatures of carrier chip and substrate results from table 4.3.1.1b expressed as a %age of temperature reached by the heater chip. It can be seen that there is general good agreement with the %ages of the heater and carrier chip over the given temperature range. All the temperatures are obtained from the images (i.e. 4-point resistance temperature was neglected)*

Power Level (W)	Heater Chip (°C)	Percentage temperature carrier chip	Percentage temperature substrate
0.0975	32	91%	75%
0.168	34	85%	71%
0.259	36	81%	69%
0.366	38	83%	67%
0.492	41	80%	62%
0.636	44	83%	58%
0.796	48	82%	54%
0.975	53	79%	48%
1.166	57	78%	46%
1.38	63	76%	46%
1.612	69	73%	45%
1.855	75	73%	42%
2.115	82	72%	40%
2.392	90	71%	39%
2.686	97	71%	38%
2.988	105	71%	36%
3.306	113	71%	35%

Table 4.4.1b: *Copper results*
Table showing the temperature readings of the carrier chip and substrate location as a percentage of the heater temperature. Note how both the %ages decline rapidly as the power levels increase. . All the temperatures are obtained from the images (i.e. 4-point resistance temperature was neglected)

4.4.2: Comparison with Other Studies

The steady state results were in good agreement with those of Trigg and Corless (2) who showed that shear stress resulting from die/substrate temperature differences are apparent even when both materials are silicon. From the results, it can be seen that the heater chip reaches a higher temperature than the carrier chip when either substrate is used. As both the heater and carrier chips were silicon, this implied that a difference in thermal expansion that would lead to some shearing action occurring within the interconnection layer, that would not be expected to occur during conventional thermal cycling. Additional information obtained was the nature of the transient thermal profiles of the assembly and influences of the substrate were highlighted. This detail of information was not reported in the studies of Trigg & Corless.

Sur and Turlik (3) have monitored transient thermal behaviour and observed the rate at which the assemblies heat up and cool down. In their work, the assembly heated up fairly quickly (the die steady state was reached in 20 seconds) but the substrate was slower due to the higher thermal mass. In contrast however, the assembly took a comparatively long time to return to room temperature, for which the substrate was considered to govern the cooling rate of the die. In the results reported earlier, this was echoed with the case when the MCMs were mounted on copper substrate, where the substrate governs the rate in which the MCM heats up/cools down.

In the work of Sur and Turlik (3), the temperature profiles of the different bodies of the assembly did not appear to follow each other, while in this work, the bodies of the MCMs showed similar trends regardless of whether the assembly was heating up or cooling down. This was attributed to the nature of the convection cooling. The model by Sur and Turlik was cooled by natural convection and the heat transfer rates show a strong dependence on the ambient temperature (i.e. the smaller the temperature difference is the lower the heat transfer coefficient), therefore the heat transfer rates are increasing as the ON cycle continues but decreasing as the OFF cycle continues. The MCMs in this experiment were cooled by forced convection, where the cooling mode is governed by the air speed and not the temperature difference.

4.4.3: Hardware Merits and Flaws

Despite the setbacks and limitations of the hardware available, they were able to obtain some of the necessary thermal data for the assembly. The calibrated 4-point resistance tracks on the heater chip provided a simple, yet effective means of obtaining the maximum temperature within the assembly. In addition, physically observing the change in resistance allowed the user to determine how close the assembly was to its maximum temperature. It is only unfortunate that there were no such tracks on the carrier chip.

The **ThermaCAM** high-resolution camera was able to provide a detailed temperature distribution of the FR4 assemblies. It showed that the heater and carrier chips both had distinct and uniform temperature distributions, and captured how the temperature of the substrate a distance away from the MCM was largely unaffected by the heat from the MCM. However, as the camera was hand held, the positioning of the camera effectively limited the repeatability of the experiment and therefore compromised the accuracy of the temperatures obtained. From graph 4.3.1.1b these positioning effects are clearly apparent from the non-linearity of the plots compared with that of the 4-point resistance measurement. However, the results from table 4.4.1 show that the captured temperature distribution is the same, regardless of positioning inconsistencies. Therefore the **ThermaCAM** was useful in providing a thermal distribution; but could not be relied upon to give an exact maximum temperature. Further experiments should use a suitable clamp or holding mechanism to rectify this problem.

The **IRISYS** low-resolution camera proved invaluable as it could be secured in a position therefore eliminating the problems associated with the high-resolution camera and allowing for repeatability. The obtained steady state temperatures were in adequate agreement with those obtained from the 4-point resistance readings, and the camera captured the transient profiles of the assembly. There were two drawbacks noted when using the low-resolution camera: a) When the camera was used to obtain images when there was a large temperature difference, the obtained profiles were fairly sharp, however when the temperature change was small (as with the case for the

lower power levels used) the images were contaminated with noise, which was noted on both the steady state and transient profiles; b) the resolution of the camera was low and this resulted in difficulty when differentiating the heater from the carrier chip and the improvisation explained in section 4.3.1.2 had to be made. This also meant that the uniform temperatures of the heater and carrier chips observed on the MCM on FR4 could not be confirmed when the assembly was mounted on copper. The camera manufacturers offered to supply a lens that would be suitable for capturing finer detail of the MCM assembly by focusing on a smaller area, though this approach was proven to be cost prohibitive.

4.5: Conclusions.

The hardware has been used to capture the thermal characteristics of the MCM on different substrate materials. The thermal profiles have shown that the material properties of the substrate have a substantial effect on the MCM, for the same power levels. Namely the maximum temperature reached and the transient profiles of the two scenarios are different depending on if the substrate is a thermally conductive or insulating material.

References

1) **Farhad Sarvar, David C Whalley & Ming K Low** "IGBT Package Design for High Power Aircraft Electronic Systems" *Journal of Electronic Packaging ASME, Dec 2001, pages 338-43*

2) **A.D. Trigg & A.R. Corless** "Thermal Performance and Reliability Aspects of Silicon Hybrid Multi-Chip Modules" *40th Electronic Components and Technology Conference, 1990, pages 592-9*

3) **Biswajit Sur and Iwona Turlik** "Power Cycling and Stress Variation in a Multi-Chip Module" *IEEE Transactions on Components Packaging and Manufacturing Technology part B Volume 10 no 2, May 1995, pages 388-395*

Chapter 5: Finite Element Model Development

Within the past 50 years the role of computers in product design and evaluation has increased dramatically. Over this period, their use has evolved from merely correcting a fault once it has happened to anticipating where and when such a fault is likely to occur. Consequently manufacturers can now base their merchandise on cost effective designs rather than inefficient methods and materials resulting in high safety factors. Though the concept of finite element modelling was initially developed for general purposes, the analysis of electronics reliability has greatly benefited from this technique. The concept of **Finite Elements** or FE can be used to assess the impact such mechanical or thermal loads have on a given design of a product.

FE may be conveniently separated into two individual areas: the mathematical theory behind the technique otherwise known as the **Finite Element Method** (FEM) which is beyond the scope of this thesis, though a thorough review of the subject is covered in (1); and **Finite Element Analysis** (FEA) which refers to the practical application of this using relevant computer software.

Analysis Type

On a global scale, the analysis type is the most obvious factor likely to affect the results. According to Ogunjimi et al (2) and later Degryse and Bell (3), there are 4 different FEA modelling strategies that are commonly used, they are: 2D plane strain analysis, axi-symmetric analysis, 3D slice analysis and full 3D analysis. Following is a brief description of each of the essential characteristics of each meshing technique with examples of how they may be applied to a flip chip assembly. For detailed discussions on the individual meshing techniques interested readers are advised to read (1-3).

2D plane strain analysis is the simplest modelling technique to execute, whereby a simple 2-dimensional representation of the structure is presented as shown in figure 5.a It is also used for simpler analysis when 3-dimensional effects of an assembly are less apparent (4-14). As well as the simplicity of the model produced, the added advantage is that the FE model analysis time is typically short and the results files are

small compared to other FE modelling methods. However the elements are often given an assumed thickness and consequently the true 3-dimensional effects of the geometry are often neglected, potentially compromising the accuracy of the results. A simple way to address the issue of the 3D effects without increasing the number of elements used is to perform an ***axi-symmetric analysis*** This type of modelling is based loosely on the 2-dimensional plane strain analysis and can be seen in figure 5b, however the geometry is centred around one axis such that the elements that are a specific distance from the axis of symmetry are represented as a continuous circle. It is for this reason that axi-symmetric analysis is unsuitable for modelling flip chip assembly as the solder joints would be represented as a continuous *solder ring* hence would be a poor representative of the true flip chip solder joint geometry. A simple way to account for the geometry of the solder joints is to perform a ***3D "slice" model analysis*** (15,16); where the full geometry of the solder joints are considered in the model. The "slice" is typically chosen such that the corner joint is modelled, the corner joint is furthest away from the neutral point and is therefore expected to endure the most stress, hence the worst case scenario is accounted for as shown in figure 5c. However, there remains the omission of potential influences of neighbouring geometry such as adjacent joints or underfill. A way of rectifying this is to perform a ***full 3D analysis*** (16-21) where all geometry is considered and this modelling method naturally offers the best agreement with experimental methods; however, it is often expensive to produce and perform an analysis (often extensive time is needed to create the model, long computational time and large amount of memory is often required in comparison to other methods). Furthermore, locating the critical stress can also be complicated due to the large number of elements. However modelling only part of the geometry as opposed to performing a full 3D analysis can often reduce these impedances. For example when a flip chip assembly is modelled in FE, typically only ½, ¼ (16-20) or potentially only $^1/_8$ (21) of the geometry need be considered when the relevant laws of symmetry are applied. Figure 5d shows $^1/_8{}^{th}$ symmetry applied.

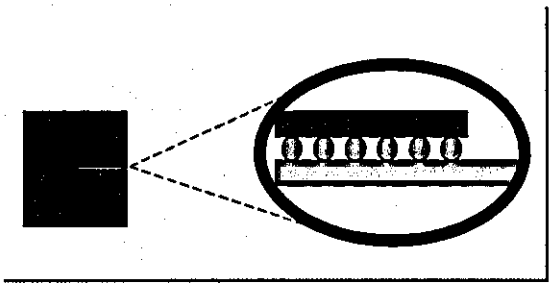


Figure 5a:
Example of a 2D plane strain analysis

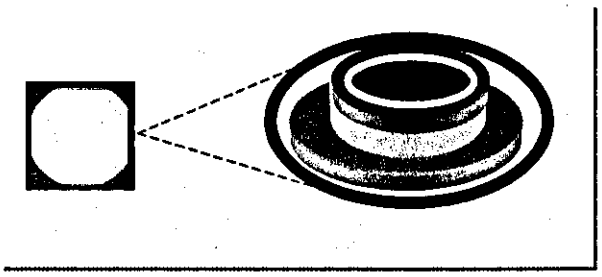


Figure 5.b:
Example of axi-symmetirc model geometry. The grey ring on top shows how the solder is typically represented

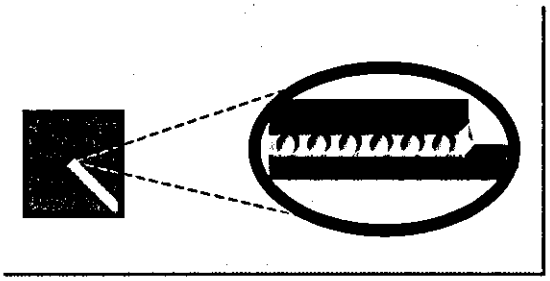


Figure 5.c:
Example of 3-D Slice model geometry

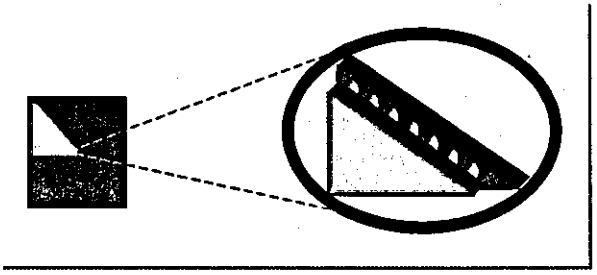


Figure 5d:
Example of 3-D full analysis. This example shows 1/8 of a complete model.

Ogunjimi et al (2) conducted an evaluation study comparing the relative merits and flaws of 2D, axi-symmetric and 3D analysis FE models utilising 926, 860 and 3038 elements respectively. They modelled a die attached to a substrate using a thermally conductive adhesive. The assembly was cooled from the adhesive curing temperature of 275°C to room temperature 25°C and stresses at various points in the model were conducted and compared with each other for analysis time memory requirements as well as results. As anticipated, 2D analysis was the quickest but gave poorest results while the 3D model offered the best results but took a long time to complete the analysis, furthermore a substantial time was utilised when creating the 3D geometry. The axi-symmetric model offered the best all round compromise as it offered results comparable with the 3D model in a time only marginally longer than that required for the 2D analysis. It is unfortunate that the very nature of the modelling method renders it unsuitable for modelling flip chip applications.

Due to the simplicity of the model and the speed in which results can be generated, 2D FE analysis is suited to situations when several models are required for comparison purposes. A typical example for flip chip is the comparison of underfills when an analyst is required to compare several and find out which is the best, but an

accurate value is not critical. 3D FEA on the other hand, should be used when accurate values are required such as when determining a maximum stress level is important, but several models are not needed.

In this study, **MARC- MENTAT** was used to create a FE model of the dimensions of a flip chip assembly; the geometry was such that the area of particular interest (the interconnection layer) was very small in comparison to the overall dimensions. Therefore issues such as mesh density and refinement became major factors when ensuring an acceptable model was created using an adequate number of elements. Furthermore, as the FE models were simulating a power cycle, the thermal boundary conditions also require careful consideration.

From the literature review, it can clearly be seen that FEA is a powerful tool to contemporary electronics analysis. It is also acknowledged that it is only a tool, one of many available to the thermal or stress analyst. As with all tools, FEA is only as effective as the knowledge of the implementing operator, hence it is important that the analyst is aware of the key criteria when implementing FEA. The number of elements used to represent specific parts of the model is crucial in ensuring that a good representation of the temperature/stress is obtained in a reasonable analysis time: if too few elements are used then the FE model is a poor representation of the true geometry in question; conversely, when a large number of elements are used then the model may not only take a long time to calculate but computer limitations (processing and memory) become significant factors. Furthermore, in scenarios such as the interconnection layer of a flip chip assembly, it may be impractical to create an entire model with elements at the density required with the disproportionate dimensions often involved. This critical issue is less apparent in conventional mechanical FE where generally objects are of reasonable size with each other. Therefore due to the computational expense of building a model with several elements, it is logical to consider local mesh refinement methods such that good results may be obtained within a reasonable calculation time and computer resources.

The aim of this part of the work was to investigate what impact the mesh density and element distribution had on the accuracy of the results as well as the time taken for the analysis to complete. Several different meshing techniques were investigated and both

the results and the analysis time were taken into consideration when each scenario was evaluated. It was also desirable to verify the results by means of an alternative solution (where possible) to ensure there was good agreement with the FE results. This chapter shows how an adequate element density for the model was established using **MARC MENTAT** FEA software; it starts with an overview of the **MARC MENTAT** package, then the mesh density was assessed from a thermal perspective. The results then influence the densities of the final FE models built.

5.1: MARC MENTAT Package Overview

The structure of **MARC MENTAT 2001** by **MSC Software** is typical of most self-contained FEA packages, where most are structured such that the analysis procedure is split into 3 stages: pre-processing, the model solution, and post-processing. MENTAT provides the user interface for all pre and post processing activities and communicates with the default FE solver MARC.

In the pre-processing stage, the geometry of the product must be specified. This is typically done by the analyst first defining the domain geometry, and then the necessary mesh is created. A mesh determines the relationship the nodes and elements have with each other, with respect to the represented domain. Within this stage, the mesh density, refinement and node repositioning are included. The boundary conditions (the external forces or physical restrictions imposed on the model) must also be defined and applied to the necessary geometry as well as the initial conditions (the state the model is in at the start of the analysis e.g. the initial temperature of the body). Also the material properties (such as the Young's modulus or thermal conductivity etc.), contact conditions such as friction (if relevant) and fatigue/fracture mechanisms are also created in the pre- processing stage.

The **MARC** solver requires the user to define a necessary **load-case** for the FE model. This includes specifying the analysis type (e.g. is it a mechanical or a thermal problem), the real time the analysis is simulated for and the number of increments the analysis is to perform, the types of elements and solver methods used and other relevant solver criteria. After the load-case has been completed, the user must then create a **job** to submit to **MARC**. As well as the load case, a **job** must include the

required output parameters (namely the element scalars and tensors) and the element types (dependent on the analysis type and the solver used). When this is complete the "**job**" is then submitted to the solver **MARC** where it is checked and then solved.

The post processing includes presenting the results in a "user-friendly" format so the failure analysis can be performed. The failure analysis typically includes identifying where and when a critical stress occurs, a maximum temperature and what the implications from the results are. **MENTAT** can display results at either a **nodal** or **global** level. To display the results at a **nodal** level, the user must manually select the node/nodes of interest by entering the node number, and then the data is collected such that a history plot (e.g. time-temperature graphs) may be displayed for that specific node. Conversely, at a **global** level the analyst can obtain a graphical plot of the thermal stress profile of the whole body. The different scalar/tensor regimes are often displayed as contour-bands or by some similar method of discretisation. A scale defining each range is given with the plot. **MENTAT** may also display each increment sequentially such that the analyst can observe the transient behaviour of the model (e.g. the model heating up or buckling etc) as a movie.

Element Classification

As well as the mesh density, the type of element used to represent the geometry is also known to influence the results significantly. When performing 3D analysis, MARC allows for either one of two element shapes to be used: **hexahedral** or **tetrahedral.** Hexahedral shape elements are generally preferred as they are known to offer marginally more accurate results due to a greater presence of near right angles; in addition when modelling in MENTAT hexahedral elements are simpler to manipulate. Conversely tetrahedral elements may offer a more accurate representation of the actual model geometry and furthermore, for some 3D analysis $^{1}/_{8}^{th}$ of a model can commonly be represented without difficulty using tetrahedral elements whereas the use of hexahedral elements effectively restricts the FE representation to ¼ of the object. The nature of the symmetry of the MCMs in question for this experiment dictated that they may be simplified to a minimum of ¼ of the model, therefore there was no apparent benefit from using tetrahedral elements. As a result, all the flip chip models built utilised hexahedral elements.

In addition to the criteria mentioned previously, the number of nodes per element is also a consideration as there are several different types of elements available for the analyser to use. Taking 2D Quadrilateral elements as an example, MENTAT offers a choice between 4 noded and 8 noded elements. 4 noded elements use 1 integration point within the element and 8 noded elements use 4 as shown in figure 5.1.3a. For 3D analysis there are 8 and 20 noded hexahedral elements utilising one and eight integration points respectively. As a rule of thumb, the more nodes per element, the longer the equations take to solve, though in some cases offer a faster convergence rate.

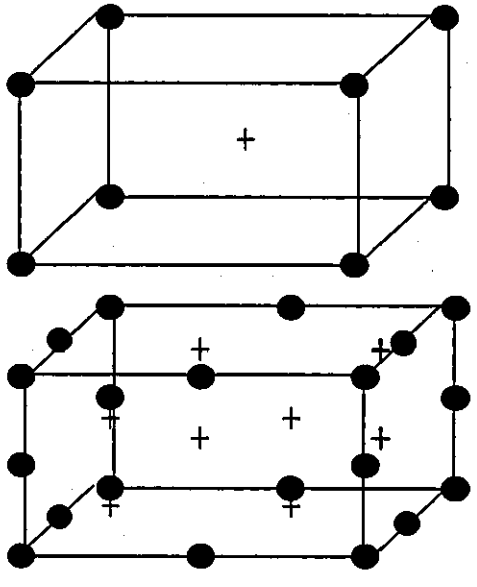


Figure 5.1.3.a:
Examples of both Hex 8 & Hex 20 elements.

Details show the nodes (black circles) along with the integration points (black crosses)

If either type of element is to be effective, then the aspect ratios should be kept as low as possible, therefore care should be practiced to ensure that a similar ratio of length/ width and height is kept whenever possible. For the case of hexahedral elements it is desirable that the elements are as close as possible to a perfect cube whereas for tetrahedral elements, an equilateral tetrahedron is sought after. It is acknowledged that it is seldom the case that it is possible for all elements to be represented this way, therefore some distortion in the element shapes is permitted by most FE packages. Most FE packages have their own tolerances when determining what extent the aspect ratios can be away from the ideal and elements can often be verified they are within

the package tolerances by performing an **element check.** Figure 5.1.3b shows examples of both good and bad elements for quadrilateral and triangular analysis.

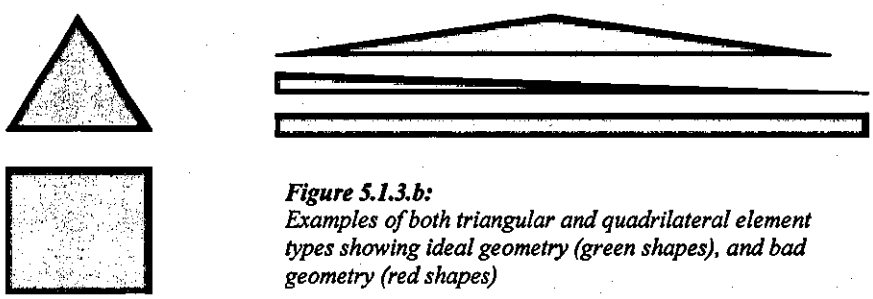


Figure 5.1.3.b:
Examples of both triangular and quadrilateral element types showing ideal geometry (green shapes), and bad geometry (red shapes)

5.2: Model Construction for Thermal Analysis

To ensure that an accurate model of the MCM was built it was necessary to ensure that the correct elements were used. The aim of this work was to examine the impact the element type, size and mesh refinement method would have on both the analysis time and the results accuracy. This was performed by a FE simulation of a Silicon block subjected to a heat flux at one face and a convection heat loss implemented in the remaining faces. Two separate issues were evaluated in this section: the mesh density and the mesh refinement methods used. These issues were evaluated separately i.e. when the mesh density was being investigated, there was no mesh refinement employed at all. The model results were subsequently compared with an alternative solution to verify that they were correct. The following sections describe the geometry and boundary conditions used in these tests.

Test Model Description

The geometry modelled was a silicon block with dimensions 6x6x39mm.The block height was given an oblong shape so a clear temperature gradient could be observed over its length without difficulty. The heat flux of 4.8 J/s was supplied at one end and an arbitrary convection coefficient of 444W/m^2 was applied to all of the remaining faces as depicted in figure 5.2.b. For the case with no mesh refinement the geometry

of the block was modelled with either 1, 8, 64, 512 or 4096 elements; the required mesh density was achieved by subdividing the existing model by 2x2x2 in the x, y and z-axes. The subsequent refinement methods utilised either 7 or 341 elements. As the mesh refinement was typically biased, (e.g. one end very finely meshed and the other is coarse) the end to which the heat flux was applied to was also varied.

A load case was created where the block was subjected to a transient thermal analysis for twenty seconds with twenty increments. From figure 5.2.3 later on it can be seen that this was sufficient time for the model to reach over 99% steady state. The time taken for each analysis to complete was recorded and the maximum and minimum temperatures of each analysis were compared with models of the alternative mesh densities and/or refinements as well as the alternative numerical solution.

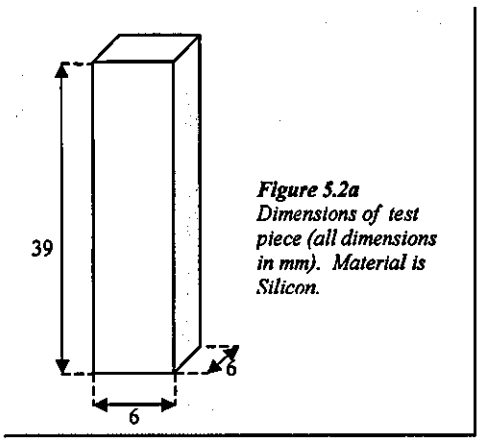


Figure 5.2a
Dimensions of test piece (all dimensions in mm). Material is Silicon.

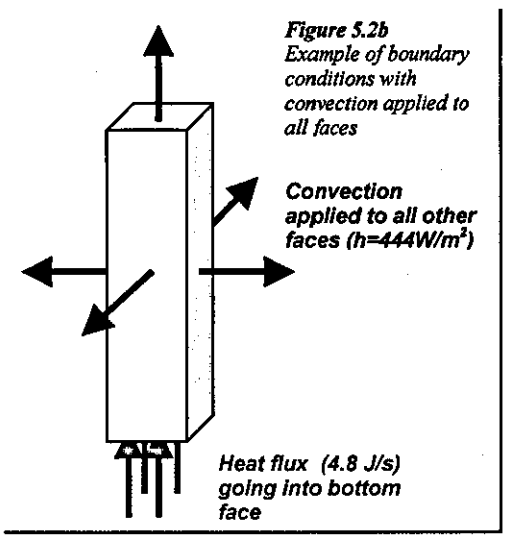


Figure 5.2b
Example of boundary conditions with convection applied to all faces

Results

The results obtained for each scenario are shown in their respective sections. Table 5.2a shows the recorded criteria for each model. Each model was given a unique identifier: for the case when hex-8 elements were used this identifier was a number whereas when hex-20 elements were used it was a letter. The manipulated variables were either the number of elements used or the mesh refinement method implemented. The measured variables were the time taken for the analysis to complete and the acquired temperature distribution. The maximum and minimum temperatures are displayed in the tables while the temperature distributions are shown as thermal plots.

Model number	Identifier for model analysis
Number of elements	Number of elements used in the model
Time taken	Real time taken for the solver to complete the analysis
Maximum temperature	Greatest temperature recorded in the model (typically from a node at the heat flux end)
Minimum temperature	Smallest temperature recorded in the model (typically at the end furthest from the heat flux)

Table 5.2a: *Description of criteria recorded from the FE models*

5.2.1: Models Evaluated with respect to Mesh Density

Initially the block was modelled without any mesh refinement to act as a benchmark for the other methods to be compared against and to evaluate what mesh density would offer good results within a reasonable time. This was an appropriate control condition, as there were no areas of discontinuity or 'cracks' within the model. In addition all the elements were of the same size and were of identical proportion to the original silicon block with internal right angles. The model was initially created as a one-element model using an 8 noded hexahedral element and after each simulation a copy of the model was made and subdivided by 2x2x2 to obtain the next level of refinement resulting in the following mesh densities:

1) One-element model
2) 8 element model (2x2x2)
3) 64 element model (4x4x4)
4) 512 element (8x8x8) model
5) 4096-element (16x16x16) model

The simulations were repeated with the same mesh densities, however the 8 noded elements were replaced with 20 noded hexahedral elements.

Table 5.2.1a shows the results obtained from the case when 8 noded elements were used while table 5.2.1b shows the results when 20 noded elements were used. From the maximum and minimum temperatures obtained and the analysis time it could be seen that when 8 noded hexahedral elements were used, the model had to be subdivided 3 times (64 element model) before convergence of the temperature was observed (i.e. temperature change was 0.2°C or less). When the 20 noded hexahedral

elements were used, an equivalent convergence was obtained after only two subdivisions. For the case when 20 noded elements were used, it could be seen that convergence at lower mesh density occurred though this benefit was compromised by the extra time required for the solver to complete the analysis.

Taking models **5** and **C** into consideration, the same temperature values were obtained, however when the 20 node elements were used the result was reached in $1/20^{th}$ of the time taken for a model with the 8 noded elements. Model 3 used 4 elements across to represent a particular feature whereas model 5 utilised 16 elements. Based on these results a minimum mesh density of 4 elements in the length width and height in the x, y, and z-axes utilising 20 noded hexahedral elements should be used.

Test	Number of elements	Time taken (Seconds)	Maximum temp. (°C)	Minimum temp (°C)
1	1	11.29	44.21	28.08
2	8	11.69	46.79	30.26
3	64	16.94	47.46	30.77
4	512	70	47.64	30.90
5	4096	763.39	47.69	30.93

***Table 5.2.1a:** Results for 8 noded hexahedral elements with no refinement*

Test	Number of elements	Time taken (Seconds)	Maximum temp. (°C)	Minimum temp (°C)
A	1	11.66	47.22	30.85
B	8	15.04	47.64	30.90
C	64	39.32	47.69	30.93
D	512	246.68	47.70	30.94
E	4096	2973.4	47.70	30.94

***Table 5.2.1b:** Results for 20 noded hexahedral elements with no refinement*

5.2.2: Mesh Refinement Evaluation

5.2.2.1: Refinement Model 1:

It is well known that FE models with a large number of elements are undesirable as they often require a long calculation time. In an attempt to minimise unnecessary elements, a model of the silicon block was created to investigate the impact of some crude meshing or nodal discontinuity caused by poor refinement techniques. Though poor accuracy was anticipated, its use may have been justified if a solution could be obtained in a quick analysis time and the degree of error was investigated and found to be within an acceptable limit. Therefore this crude mesh refinement method may allow for a flip chip design to be implemented rapidly and a prediction of the final temperature distribution obtained. To ensure element distortions did not influence the results all the elements used were kept in the same proportion (i.e. length, width and height ratio) as the original block dimensions.

A so-called refinement level of 1 was used to create the FE geometry where the refinement level 1 refers to the level of discontinuity as depicted in figure 5.2.2.1. In this example, the model refinement was as follows (1^2-2^2-4^2-8^2-16^2) elements. Two cases were considered, one where a coarse mesh was used at the end where the heat flux was applied and the other where a fine mesh was applied at the heat flux end.

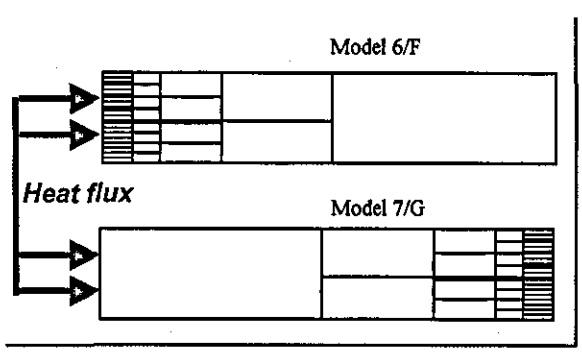


Figure 5.2.2.1 *Diagrams showing mesh refinement models used*

Results

Model	Characteristics
6/F	Refinement level 1, heat flux applied at fine mesh end
7/G	Refinement level 1, heat flux applied at coarse mesh end

Table 5.2.2.1a: *Refinement method and model identifier*

Test	Number of elements	Time taken (Seconds)	Maximum temp. (°C)	Minimum temp (°C)
6	341	47.36	49.11	30.13
7	341	52.55	47.09	30.06

***Table 5.2.2.1b:** Results of models when 8 noded elements were used*

Test	Number of elements	Time taken (Seconds)	Maximum temp. (°C)	Minimum temp (°C)
F	341	160.87	48.84	30.48
G	341	165.94	47.83	30.56

Table 5.2.2.1c: Attributes of models when 20 noded elements were used

Figure 5.2.2.1 shows the end the heat flux was applied to (either the fine end or the coarse end) while table 5.2.2.1a shows the corresponding model number. The results are shown in tables 5.2.2.1 b & c for when 8 noded and 20 noded hex elements were used respectively. It can be seen that when the refinement was applied at the heat flux end, the temperatures of this face were significantly greater than the case when there was no refinement, while the models with the heat flux applied to the coarse end appear to underestimate the predictions of the values with no mesh refinement. These discrepancies were attributed to the discontinuity for the various nodal paths. It was also noted that the times taken for both the FE analyses to complete were still lengthy, so both refinement methods were neglected, as there was no benefit of time or elements saved. Even when considering the case when 20 noded elements were used (i.e. F, & G) the models could be seen to over estimate the maximum temperatures reached by an unacceptable degree, therefore these results confirm that this mesh refinement method is unreliable.

A possible way to overcome the limitations of these models is to employ tying equations. This involves modifying the equations at the nodes at the refinement interface, such that the equations account for the path discontinuity that exists within the model. However, this process requires manual input for each node where the discontinuity exists and is therefore a very labour intensive procedure. If this method is to be implemented then an evaluation study that investigates whether the extra time spent selecting and amending the nodes would benefit the subsequent analysis.

5.2.2.2: Refinement Model 2

The previous experiment had established the importance of nodal path continuity when refining the model. An alternative refinement method was investigated that involved ensuring continuity between all nodes within the model (i.e. no cracks) although this led to skewness of internal angles as a result of the necessary element distortion. Using the same overall model geometry, two different variations of this were implemented both using 7 elements; two elements had no distortion, while the remaining 5 exhibited internal acute angles. Two models were created: one with the heat flux applied at the refined end and the other with the heat flux applied at the coarse end.

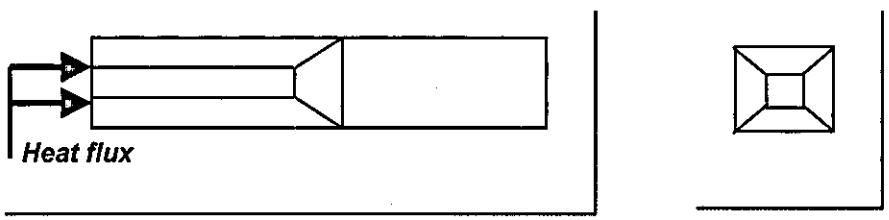


Figure 5.2.2.2a:
Diagrams showing mesh refinement models used from side view and plan view

Results

Table 5.2.2.2a explains which end the mesh refinement was applied to while tables 5.2.2.2a and b show the recorded results from the analysis compared with the models 2 and b. This particular experiment was to investigate how element distortion influenced the analysis time and the accuracy of the results. The models were compared with those that used a similar number of elements from section 5.2.1 (i.e. without mesh refinement). Therefore models 8 and 9 were compared with model 2 while model H and J were compared with model B. The results showed an excellent agreement with respect to the final values reached, particularly for the case when the heat flux was applied at the densely meshed end as shown in table 5.2.2.2b. Thereby showing that a moderate level of element distortion does not compromise the model results.

Figure 5.2.2.2.b demonstrates how such a refinement method can be used to ensure continuity from a small body to one that is semi-infinite as is the case for solder joints in flip chip assemblies. This technique should be used to model the solder joints and

ensure continuity. The correct mesh density can then be obtained by subdividing the entire model as performed in 5.2.1. Furthermore, figure 5.2.2.2c shows how such a meshing technique can be used to model an entire solder interconnection layer, so less elements are used in the surrounding bodies.

Model	Characteristics
8/H	Refinement model 2, Heat flux applied at fine mesh end
9	Refinement model 2, Heat flux applied at coarse mesh end

Table 5.2.2.2a: *Model parameter identifiers*

Test	Number of elements	Time taken (Seconds)	Maximum temp. (°C)	Minimum temp (°C)
8	7	11.95	46.78	30.28
9	7	11.89	46.64	30.24
2	8	11.69	46.79	30.26

Table 5.2.2.2b: *Results of models with 8 noded elements*

Experimental results for 20 noded 3D elements

Test	Number of elements	Time taken (Seconds)	Maximum temp. (°C)	Minimum temp (°C)
H	7	13.68	47.64	30.91
B	8	15.04	47.64	30.90

Table 5.2.2.2c: *Results method for models with 20 noded elements*

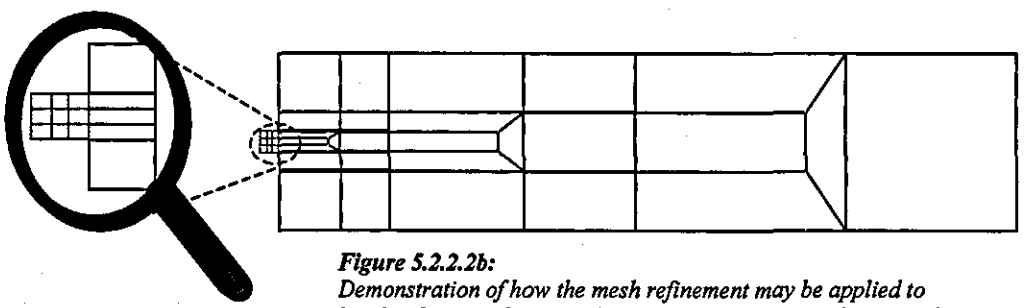


Figure 5.2.2.2b:
Demonstration of how the mesh refinement may be applied to localised areas of interest (orange mesh represents the area of interest (e.g. solder joint, small in comparison to other geometry)

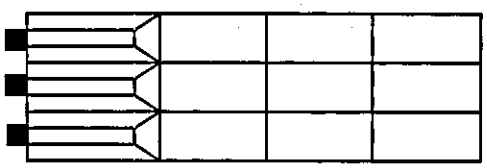


Figure 5.2.2.2c
Demonstration of how interconnection joints can be modelled utilising the mesh refinement method.

5.2.3: Alternative Solution (Numerical)

From table 5.2.1 a and b (when no refinement method was used) it can be seen that as the number of elements is increased, that the models appeared to converge to what could only be assumed were the correct values. If these are the correct temperatures then these values can be verified by some alternative method. Due to the difficulty in obtaining a viable analytical solution for this particular problem, it was decided to calculate the heat transfer by taking the temperature gradients and working backwards to obtain the power required to create such a temperature profile. Model 4 was used in the experiment as it showed a reasonable level of convergence without an excessive amount of elements used where it can be seen that there is a linear temperature gradient across the vertical surfaces of the block and a uniform heat loss from the top face. The time/temperature graph for the heated face (where the heat flux was applied) and the far end (where the coolest temperature was detected) is shown in figure 5.2.3 It can be seen that the model appears to have reached steady state, so Newton's Law of cooling may be used to evaluate the steady state heat transfer.

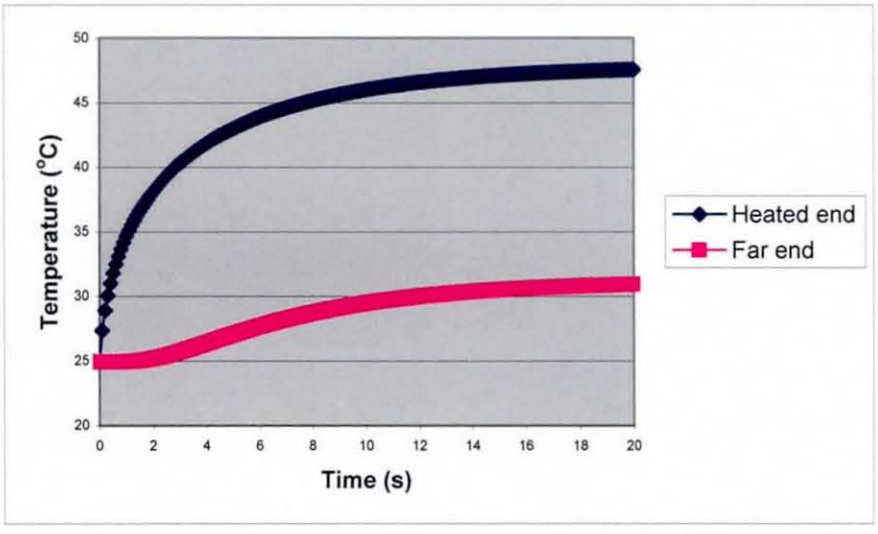


Figure 5.2.3a: *Transient profile of two nodes taken from either end of the model, from where the heat flux was applied and the far end.*

To perform the calculation, the block was divided up into separate faces based on the vertical node distance as shown in figure 5.2.3b. The temperatures for each node surrounding the face were taken and an average obtained which was assumed to be uniform across the surface.

For the top surface the heat flux was calculated using: -

$$\frac{dQ}{dt} = Ah(T - T_\infty)$$

$$=(6x10^{-3})^2 \times 444(30.94-25)$$

$$=0.0949j/s$$

Using the temperature of 47.7°C obtained from the converged FE model (no 4) and applying Newton's law of cooling the heat loss from the top face was found to be 0.0949J/s. The same procedure was applied to the external faces of the other elements as shown in table 5.2.3 which gave the total heat loss through the face of each model.

$$\frac{dQ}{dt} = Ah(T - T_\infty)$$

Where
A= area of surface
h=444
T= Surface temperature
T_∞=Ambient temperature

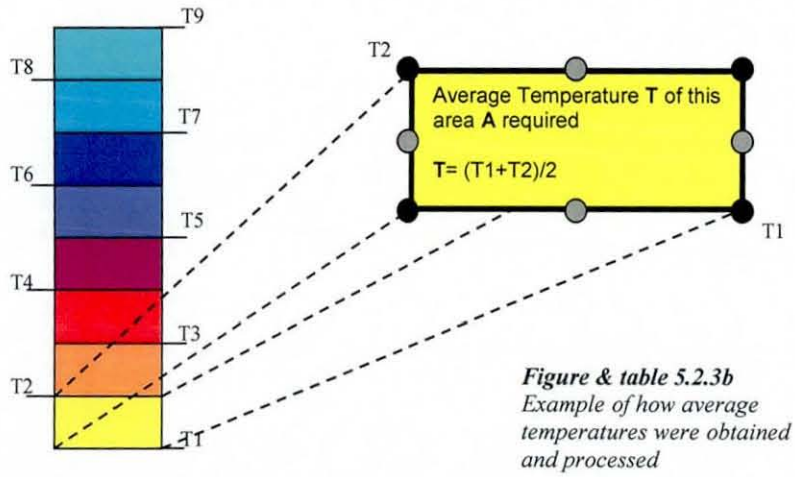


Figure & table 5.2.3b
Example of how average temperatures were obtained and processed

Actual Nodal Temp/ ℃	***Average temp between nodes /℃***	***dQ/dt (J/s)***
47.69 (T1)	45.355	0.2643504
43.02 (T2)	41.265	0.2112336
39.51 (T3)	38.135	0.1705842
36.76 (T4)	35.72	0.1392206
34.68 (T5)	33.92	0.1158440
33.16 (T6)	32.63	0.0990908
32.10 (T7)	31.775	0.0879869
31.45 (T8)	31.305	0.0818830
31.16 (T9)		
	Total	1.1701935

Total energy lost through one face is 1.1701935 J/s

Multiplying by 4 to get loss from all sides is 4×1.1701935 = 4.680774 J/s
Now adding loss from top face is 4.680774+0.0949= 4.77 J/s
The value obtained of 4.77 J/s was in good agreement with the original heat flux used in the model of 4.8 J/s

5.3: Heat Transfer by Convection

5.3.1: Introduction

Convection is defined as the heat transfer between a solid surface and a surrounding fluid. In comparison with conduction, convection is distinctive as fluid motion contributes significantly to the resultant heat transfer. Convectional heat transfer to and from the environment plays a central role when either thermal or power cycling an electronic assembly: if it is thermal cycled then the assembly is both heated and cooled by means of convection; however convection only contributes to the cooling of the assembly if it is power cycled. In either case, the rate of convection heat transfer can be determined by Newton's law of cooling or:

$$Q=hA(T_s-T_\infty)$$

Key:
Q= heat energy
h=Convection heat transfer coefficient
A=Area of convection surface
T_s= Temperature of surface
T_∞= Natural temperature of fluid.

Once the value of the convection heat transfer coefficient **h** has been successfully determined, Newton's law is directly applicable to all but the most complex convection scenarios (22-24). Nonetheless, obtaining the correct h value can be challenging and problematic as this coefficient is dependent on several individual parameters and conditions, thus determining the correct h coefficient is an intricate task.

When an electronic assembly such as a flip chip is subjected to thermal cycling, the typical slow ramp times and long dwell times involved permit the assumption that the temperature of the component is the same as that of the surrounding fluid. Furthermore, the experimenter often rigorously controls these temperatures at any given time, so that only a basic knowledge of convection heat transfer is necessary when thermal cycling a component. Power cycling on the other hand, requires a more profound knowledge of the relevant convection heat transfer mechanics. For example when an electronic component is subject to an internal heat flux, the convection heat transfer rate ultimately determines:

- The maximum temperatures the various assembly bodies reach when the heat flux is applied
- The time taken for the whole assembly to reach this maximum temperature or steady state
- The rates the bodies cool down after the heat flux is removed.

Though a convection heat transfer value has been briefly discussed and implemented in the preceding subsection, the number was used without revealing how such a value was or should be obtained. Though the h coefficient is dependent on several parameters, the fundamental criterion in determining the h value for any surface to or from a fluid is the type of convection heat transfer the assembly is subjected to. Convection is typically classified as either forced or natural (22-24).

Forced Convection

Forced convection occurs when a fluid is forced to flow over or through a body. Typically some mechanical apparatus such as a fan or pump initiates the fluid motion, however the fluid motion may also be induced by natural mechanisms such as wind or similar fluid current surges. Forced convection may be further sub-categorised to either external or internal convection. External convection is used to describe fluid flow over a surface, whereas internal forced convection indicates that fluid is flowing through a body by means of a pipe or duct. With specific reference to electronics cooling, a component is said to be cooled by external convection if a cooling fan forces air across the component surface, conversely it is cooled by internal convection if the assembly has cooling fluid passed through internal vias or channels. Apart from the physical properties of the cooling fluid, the fluid velocity principally governs how effective the coolant will be.

One of the assemblies modelled in this study was a MCM that was mounted onto a much larger substrate; the whole assembly was subjected to an environment such that airflow was at a constant velocity across the body. Therefore the electronic assembly in question was subjected to external forced convection and for the purposes of modelling the thermal behaviour, it was necessary to determine appropriate values for h, which will be discussed later.

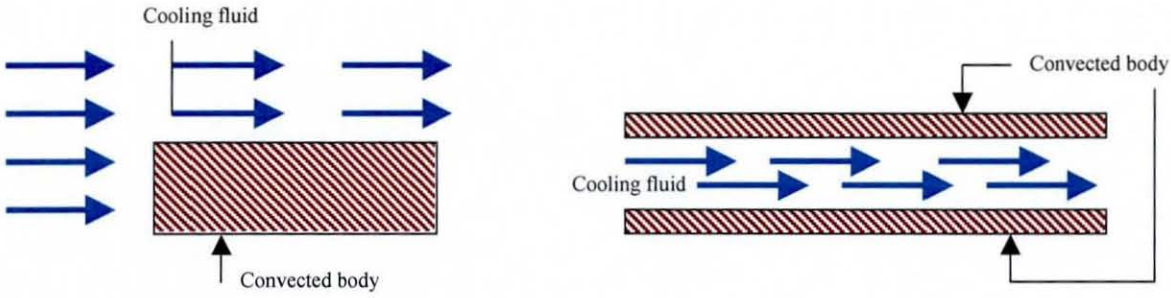


Figure 5.3.1a &b:
Illustration of a body cooled by forced convection. The body is cooled by external convection (a) and internal convection (b)

Natural Convection

The fluid motion in natural convection models occurs as a consequence of the buoyancy phenomenon. It is well known an area of warm fluid is of less density than the cooler surrounding fluid. Consequently the warmer fluid molecules rise until the thermal energy is diffused. The larger the temperature difference between the surface and the ambient fluid, the more significant the influence gravity has on the cooler fluid molecules, hence the greater the fluid velocity at the component surface and the larger the resultant heat transfer rates.

Generally the heat transfer rates are lower for a component cooled by natural convection than those subjected to forced convection. However, natural convection heat transfer rates are normally greater for vertically aligned surfaces than the horizontal equivalents. In the case of surfaces that are vertically aligned, the convection current is parallel to the surface as shown in figure 5.3.1c. For horizontal surfaces, the heat transfer rates are strongly dependent on the direction of the face and whether it is warmer or cooler than the ambient fluid as depicted in figure 5.3.1d. It can be seen that cases a and b allow higher heat transfer rates to and from the environment as the surfaces are aligned with the buoyancy effect. In scenarios c and d the surface orientation impedes the natural convection flow as the fluid must flow across the surface before it can continue.

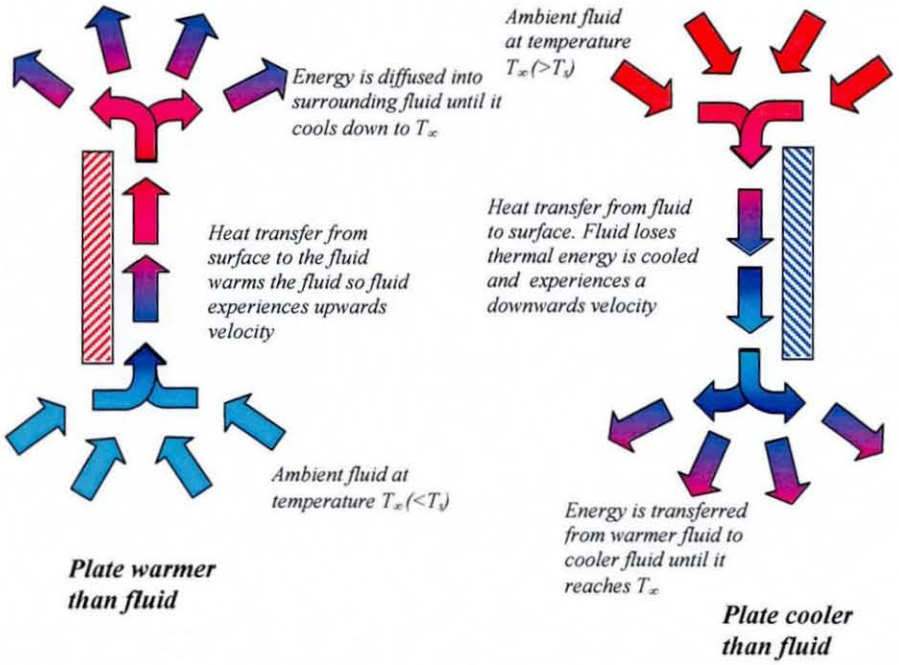


Figure 5.3.1c: *Schematic showing the vertical natural convection currents depending on if the fluid is cooler or warmer than the surface*

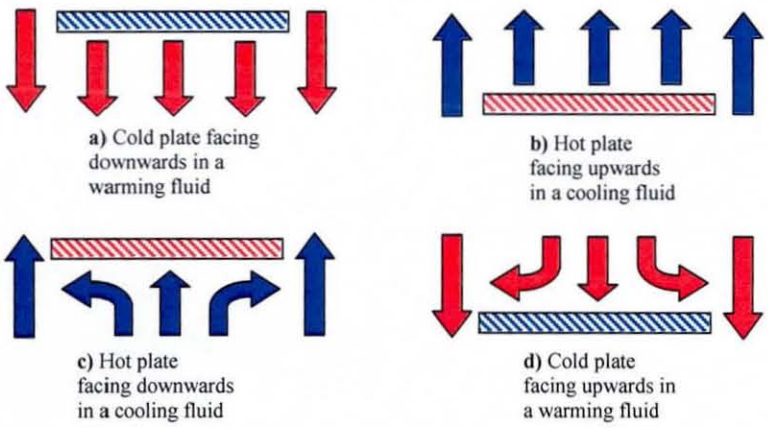


Figure 5.3.1d: *Schematic showing the horizontal convection characteristics of different plate configurations*

5.3.2: Fundamental Framework for Convection

Once a basic understanding of fluid flow concepts has been established, then it is possible to correctly obtain an adequate convection heat transfer value. This short section serves to define some fluid flow characteristics that are common to both natural and forced convection, namely the **velocity** and **thermal boundary layers**.

5.3.2.1: Velocity Boundary Layer

Forced Convection

Due to bonding properties of fluid molecules, the speed that a fluid flows over a surface or body varies with the distance from the surface. Fluid molecules in direct contact with the surface may adhere to it, and the motion of the layer of molecules immediately above these will be restricted so they move very slowly in the direction of the fluid flow. Molecular layers further from the surface flow at greater velocity and thus a velocity gradient exists within the fluid flow from directly above the surface until the fluid molecules are far enough away from the surface for the adhesion effects to be negligible. If U is defined as the fluid velocity within the region between the body surface and some point vertically above, then U_∞ will represent the velocity of the fluid when it is far enough from the surface for free fluid flow to occur. Thus the values of U range from 0 at the surface of the object to U_∞ and this region is defined as the velocity boundary layer. The velocity boundary layer initiates when the fluid flows over the surface leading edge, and its thickness δ increases with the distance from the edge as shown in figure 5.3.2.1a. The velocity boundary layer is defined as the region where $\mathbf{U/U_\infty < 0.99}$

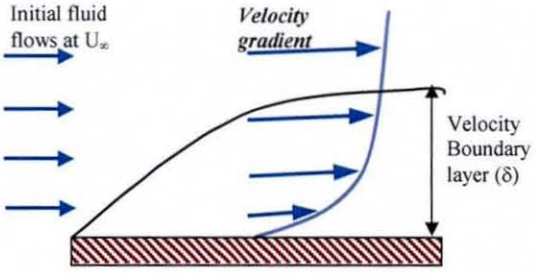


Figure 5.3.2.1a: *Schematic of drag/adhesion effects as a fluid flows over a surface*

The nature of fluid motion may be classified as either **laminar** or **turbulent**. When the fluid flows over the body in a smooth and orderly manner the fluid flow is said to be laminar, whereas turbulent fluid flow implies that it is hectic and random in nature. For forced convection, the fluid motion is initially laminar until some critical distance over the surface is reached (x_c) then a transition occurs between laminar and turbulent flow. This transition is characterised by minor fluctuations or 'ripples' and some distance after this, the fluid flow becomes fully turbulent as shown in figure 5.3.2.1b. The positions of these critical points are dependent on the fluid properties, the velocity and the surface geometry.

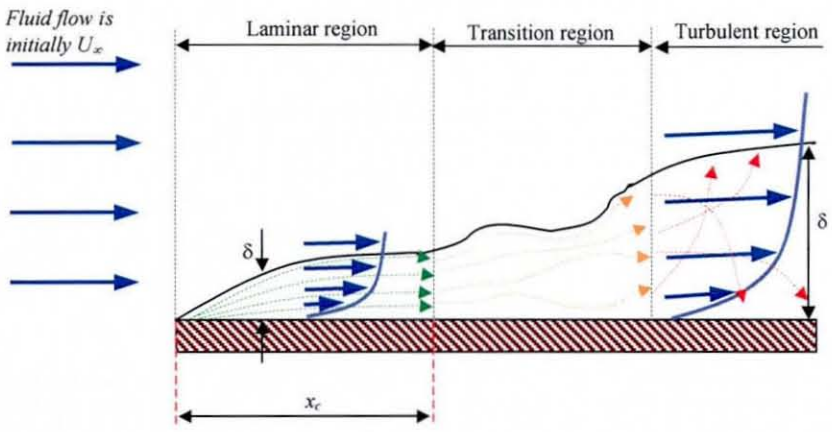


Figure 5.3.2.1b: *Fluid flow evolving from laminar, through transition to turbulent*

Natural Convection

In natural convection, fluid and body surface adhesion still occurs and some velocity gradient exists within the local space very close to the surface. However, the velocity of the fluid reaches a maximum value some small distance away from the surface and after this point the fluid velocity diminishes as the thermal energy is diffused. For natural convection the velocity boundary layer is specified to be the distance where there exists fluid motion as a result of the temperature difference between the surface and the ambient fluid.

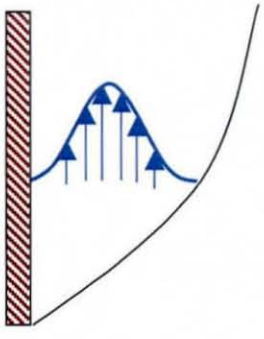


Figure 5.3.2.1c: *Illustration of natural convection where velocity at the surface is 0,which reaches a maximum in the boundary layer before diminishing as it reaches the edge.*

As with the case for forced convection, the fluid flow within the velocity boundary layer is initially laminar until some critical distance is reached and turbulence is introduced into the fluid flow. However, unlike the case for forced convection where a significant transition region exists, the change is characterised as instantaneous.

5.3.2.2: Thermal Boundary Layer

As there is heat transfer between the surface and the fluid, this indicates that some thermal gradient exists within the fluid; hence a thermal boundary layer similar to that of the velocity boundary layer exists.

At the surface of a component, the fluid velocity is zero; therefore the heat transfer at the surface is initially by conduction only. These fluid particles then emit thermal energy to the moving adjacent layer and so on. It should be noted that as the heat is transferred to above layers the fluid motion is playing a role of greater and greater significance, consequently the transfer of thermal energy diminishes across the thermal boundary as demonstrated in figure 5.3.2.2 and these thermal characteristics are the same for both forced and natural convection. If the temperature at the body surface is defined as T_s and the temperature significantly far away from the body is T_∞, then let T be the fluid temperature between these two. In this case the thickness of the thermal boundary layer θ is defined as $(T_s - T)/(T_s-T_\infty)<0.99$

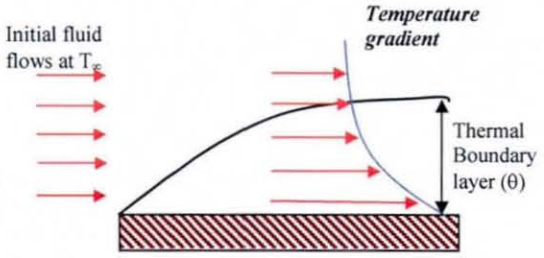


Figure 5.3.2.2: *Schematic diagram showing propagation of the thermal boundary layer. Example is for a warm plate in a cooling fluid*

It is often assumed that the thermal boundary layer always starts at the leading edge of a surface. While this is indeed the case for a uniform heated surface, for a non-uniformly heated surface (i.e. leading edge is not heated) the thermal boundary layer may propagate after the velocity boundary has formed.

5.3.3: Dimensionless Parameters

As the convection coefficient h is dependent on many variables, several dimensionless numbers are defined that describe the relevant characteristics of either the heated body or the surrounding fluid (24). These numbers significantly simplify the subsequent calculation of the h coefficient and this section first introduces these numbers and later describes the relationship the numbers have with each other.

5.3.3.1: Prandtl Number

The Prandtl number fundamentally describes the generic properties of the fluid that are not specific to the fluid velocity or surface geometry. Namely it is a ratio of the fluid kinematic viscosity and the thermal diffusivity of the fluid. That is to say, the Prandtl number describes the growth of the thermal boundary layer with respect to the velocity boundary layer and vice versa. The Prandtl number may be defined simply as: -

$$\mathbf{Pr}=\nu/\alpha \text{ or } \mathbf{Pr}= \mu c_p/k$$

Key
ν = *Kinematic viscosity of fluid*
α = *Thermal diffusivity of fluid*
μ = *Dynamic viscosity of fluid*
c_p = *specific heat of fluid (const pressure)*
k = *thermal conductivity of fluid*

A Prandtl number less than 1 implies that the thermal boundary layer is growing faster than the velocity boundary layer, conversely a Prandtl number greater than one indicates the velocity boundary layer is growing faster as shown in figure 5.3.2.1a and b respectively. Prandtl numbers can be grouped according to their value, and range from very low for liquid metals, to very high for some industrial oils.

Air (in either a forced or natural convection model) is a common cooling fluid for electronic assemblies and was used in the experimental and modelling studies presented here. The Prandtl values for air show a very small change in value (<0.001) within the necessary temperature regions considered, therefore the Prandtl value implemented in the models was assumed to be constant and temperature independent between 25 and 100°C and was 0.714 (24).

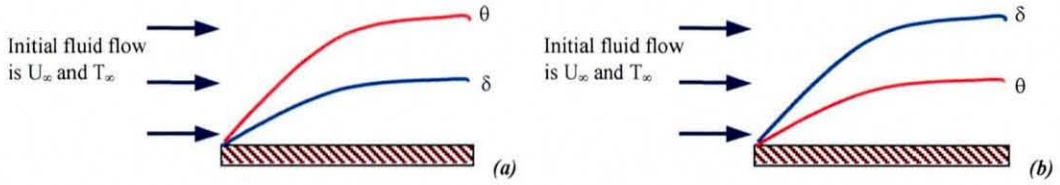


Figure 5.3.2.1a &b:
Graphs depicting the scenarios (a) Pr is less than 1 (i.e. thermal boundary grows faster, and (b) when the velocity layer grows faster

5.3.3.2: Reynolds Number

The Reynolds number is a dimensionless parameter used to describe fluid velocity properties driving forced convection. The Reynolds number is defined as the ratio of the inertia force against the viscous force within the boundary layer. A large Reynolds number implies that inertia forces are dominant, while a small Reynolds numbers indicates that viscous forces are dominant. Over a flat surface the Reynolds number may be specified as:

Key
U_∞ = ambient fluid velocity
x =length of surface
v= Kinematic viscosity of fluid

$$Re = U_\infty x/v$$

All the relevant data regarding the fluid flow and the geometry are expressed as one number. Therefore the Reynolds number for air flowing at 5m/s over a 2m long surface has a value equivalent to that of air flowing at 10m/s over a 1m long surface. Furthermore, the Reynolds number can be used to determine whether the airflow is laminar, turbulent or in the transition region (24). A Reynolds value up to and including $5x10^5$ designates that the airflow is laminar, Reynolds values between this and $1x\ 10^6$ denote the transition region, while greater values indicate that the airflow is fully turbulent, though specific surface or fluid criteria may alter these values.

Velocity Boundary Layer Thickness

The maximum boundary thickness δ can be determined from the Reynolds number also. For the laminar region of the fluid flow, the boundary may be determined as:

$$\delta/x = 5/\sqrt{Re}$$

5.3.3.3: Grashof and Rayleigh Numbers

While the Reynolds number is used to describe the flow conditions in a forced convection environment, fluid motion for natural convection may be summarised using an equivalent number. The Grashof number describes the ratio of the buoyancy force and the viscous force within the boundary layer and is expressed as:

$$Gr_x=[g\beta(T_s-T_\infty)x^3]/v^2$$

Key
β = the buoyancy force ($\beta=1/T$ for an ideal gas),
*x - the characteristic length**
g= gravity
T_s = Temperature at the surface
T_∞ = Ambient temperature
v= kinematic viscosity of the fluid

**x*= length for vertical plate and
x= Area/Perimeter for horizontal plate

Another dimensionless number can be used to determine if turbulence exists along the surface. The Rayleigh number is defined as the product of the Grashof and Prandtl number such that:

$$\mathbf{Ra_x=Gr_x\,Pr}$$

As there is no transitional region between laminar and turbulent airflow in natural convection, there is one Rayleigh value identified that is considered to correctly represent the change from laminar to turbulent airflow: below 10^9 the flow is laminar and above this the flow is turbulent (23,24)

5.3.3.4: Nusselt Number

Whereas the Reynolds and Grashof numbers simplify the fluid motion characteristics for forced and natural convection respectively, the Nusselt number (**Nu**) is used to simplify the thermal properties of the fluid. In particular it provides a measure of the convective heat transfer at the surface. The greater the Nusselt value, the larger the temperature gradient at the surface, hence a large heat transfer coefficient. The Nusselt number is principally defined as:

$$\mathbf{Nu_x}= hx/k$$

Where
h= Convection heat transfer coefficient
x= surface length in fluid flow direction
k= Thermal conductivity of the fluid

5.3.3.5: Relationships Between the Dimensionless Parameters

The Nusselt number can also be expressed as a function of either the Reynolds number or Grashof number depending on the type of convection used. For forced convection, the Nusselt number is given by:

$$\mathbf{Nu}_x=C\mathbf{Re}_x^{\,m}\mathbf{Pr}^n$$

Where C, m, and n are constants whose values depend on the Prandtl number of the convection fluid and the type of fluid flow (i.e. laminar or turbulent). For a heated surface subject to being cooled by forced air the full expression is $0.332\mathbf{Re_x}^{1/2}\mathbf{Pr}^{1/3}$ (equation 5.3.3.5 a_i)for laminar airflow and $0.0296\mathbf{Re_x}^{4/5}\mathbf{Pr}^{1/3}$ (equation 5.3.3.5a_{ii}) when the airflow is turbulent (24)

The Nusselt number may also be obtained as a function of Rayleigh and Prandtl numbers for natural convection. Table 5.3.3.5 describes the various expressions for the Nusselt numbers for any fluid depending on the conditions.

Nusselt number expression	***Rayleigh number***	***Plate conditions***
$0.68+(0.670Ra_l^{1/4})/[1+(0.492/Pr)^{9/16}]^{4/9}$	$Ra < 10^9$	Vertical
$(0.825+(0.387Ra_l^{1/6})/[1+(0.492/Pr)^{9/16}]^{8/27})^2$	$10^{-1}<Ra_l<10^{12}$ *	Vertical
$0.96Ra_l^{1/6}$	$200<Ra_l<10^4$	Horizontal (upwards heated)
$0.59Ra_l^{1/4}$	$200<Ra_l<10^4$	Horizontal (upwards heated)
$0.54Ra_l^{1/4}$	$10^4<Ra_l<10^7$	Horizontal (upwards heated)
$0.15Ra_l^{1/3}$	$10^7<Ra_l<10^{11}$	Horizontal (upwards heated)
$0.27Ra_l^{1/4}$	$10^5<Ra_l<10^{10}$	Horizontal (upwards cooled)

Table 5.3.3.5: *Summary of the different natural convection parameters used*

Once the nature of the fluid flow has been established, a value for the Nusselt number can be obtained by measuring the relevant parameters and collecting published data. Once the Nusselt number has been determined equation 5.3.3.4 may be rearranged to obtain a specific value for h. This expression can then be multiplied by 2 to obtain the average convection heat transfer coefficient for the whole surface, such that:

$$h= 2(\mathbf{Nu_x} \bullet k)/x \qquad \text{(equation 5.3.3.5.b)}$$

Non-heated Starting Length

It is often assumed that the velocity and boundary layer initiate from the same starting point. While this is the case for a uniform heated surface, this is an invalid assumption for the case where the starting length is not heated and the thermal boundary layer is instigated some time after the velocity boundary layer. In this case then the following formula is used:

$$\mathbf{Nu}_x = 0.332\mathbf{Re}_x^{1/2}\mathbf{Pr}^{1/3}\left(1-(x_0/x)^{3/4}\right)^{-1/3} \qquad \text{(equation 5.3.3.5c)}$$

Where
x_0 = the unheated starting length
x =is the entire length
& the Nusselt, Reynolds and Prandtl numbers as defined previously

5.3.4: Heat Transfer Coefficient for the MCM Models

From equation 5.3.3.5b, it can be seen that the heat transfer coefficient is strongly dependent on the surface length in the fluid flow direction. Therefore the value may vary significantly depending on how the surface geometry is considered. The purpose of this section is to show how the convection coefficient was obtained, the method of deriving its calculation and the approximations of the model geometry that would ultimately result in the calculation being simplified.

All models were created utilising the mesh generation technique described at the end of this section. In addition, particular care was taken to ensure that the cross-sectional area of the solder joints on the FE model were equivalent to those on the actual assembly. The centre point of each solder joint was measured in Cartesian form from what was considered to be the centre of the chip using an optical microscope, and these reference points were used to generate node locations. The remainder of the model was constructed with respect to these positions.

The following is a discussion of the different ways in which the experimental geometry was interpreted to influence the convection heat transfer coefficient h. The coefficient was evaluated for the MCMs based on the environment they were to be operated in. For each convection scenario, the h values were obtained by first determining the Reynolds number, subsequently deriving the Nusselt numbers and finally obtaining the average h coefficient. The equations for the dimensionless parameters specified in section 5.3.3 were utilised for this work, while the model used for the verification work was the MCM assembly mounted on FR4 substrate, powered up with 1W and a steady state analysis. The results were compared with the results

from the thermal imaging camera of the MCM powered up with the same conditions from table 4.3.1.1b

a) Leading edge x starts at the entrance of the wind tunnel (1500mm)

To begin with, the leading edge (entrance) of the wind tunnel was considered as the point from which the velocity boundary layer would propagate from. The position of the MCM assembly was 1.5 metres away from the wind tunnel entrance. However as this was unheated, this required the use of equation 5.3.3.5c to evaluate h at the point of the assembly.

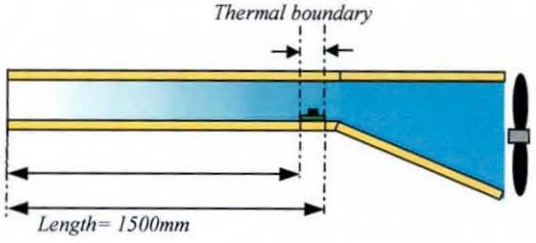


Figure 5.3.4.a
Leading edge considered to be wind tunnel platform.

i)Determine the Reynolds number (Re_x) for the wind tunnel entrance:

$$Re_x = U_\infty \frac{x}{v}$$

$$\therefore Re_x = 471994.965$$

$U\infty$= 5m/s
x=1.5m
v=15.89x10^{-6} m^2/s

ii) Determining Nusselt number (Nu_x)

Leading edge of the tunnel is not heated, so equation 5.3.3.5c should be used:

$$Nu_x = 0.332\left(Re_x^{\frac{1}{2}}\right)\left(Pr^{\frac{1}{3}}\right)\left[1-\left(\frac{x_0}{x}\right)^{\frac{3}{4}}\right]^{\frac{-1}{3}}$$

$$\therefore Nu_x = 0.332\left(471994.965^{\frac{1}{2}}\right)\left(0.714^{\frac{1}{3}}\right)\left[1-\left(\frac{1.46}{1.5}\right)^{\frac{3}{4}}\right]^{\frac{-1}{3}}$$

$$\therefore Nu_x = 0.332(687.019)(0.894)[3.68]$$

$$\therefore Nu_x = 750.198$$

Re_x= 471994.965
Pr =0.714
x_0 =1.46m
x =1.5m

iii) Determining the maximum convection coefficient (h_x)

$$h_x = \frac{Nu_x \bullet k}{x}$$

$$h_x = \frac{750.198 \times 0.0263}{1.5} = 13.153$$

Nu_x= 750.198
K=0.0263W/m^2
x =1.5m

But to obtain the average heat transfer coefficient, the maximum value must be multiplied by 2

$$\overline{h} = 2 \times h_x = 2 \times 13.153 = 26.307 \text{ W/m•K}$$

b) Leading edge x is the dimension of substrate (40mm)

The subsequent analysis involved taking the assembly into isolation (i.e. neglecting the characteristics of the leading edge of the wind tunnel) it was noted that the length of the substrate dominates the overall chip geometry; therefore the calculation of the h value only considered the length of the substrate and neglected the geometry of the assembly. The thermal effects of the chip were considered to start in the region of the chip and the model was assumed to have an unheated starting length.

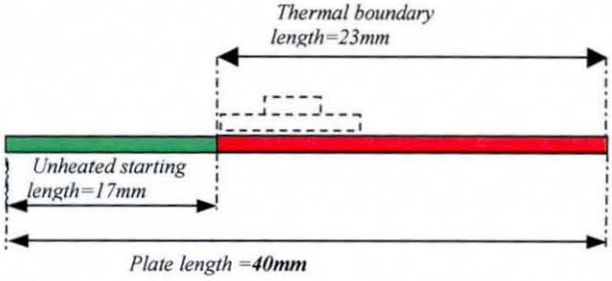


Figure 5.3.4.b
Schematic of unheated starting length of the substrate. MCM is placed after 17mm on the board.

i)Determine Reynolds number

$$\mathrm{Re}_x = U_\infty \frac{x}{v}$$

$$\therefore \mathrm{Re}_x = 12586.532$$

$U\infty = 5m/s$
$x=0.04m$
$v=15.89\times10^{-6}m^2/s$

ii) Determining Nusselt number:
Leading edge not heated so equation 5.3.3.5c should be used

$$Nu_x = 0.332\left(\mathrm{Re}_x^{\frac{1}{2}}\right)\left(\mathrm{Pr}^{\frac{1}{3}}\right)\left[1-\left(\frac{x_0}{x}\right)^{\frac{3}{4}}\right]^{\frac{-1}{3}}$$

$$\therefore Nu_x = 0.332\left(12586.532^{\frac{1}{2}}\right)\left(0.714^{\frac{1}{3}}\right)\left[1-\left(\frac{0.017}{0.04}\right)^{\frac{3}{4}}\right]^{\frac{-1}{3}}$$

$$\therefore Nu_x = 0.332(112.19)(0.894)[1.283]$$

$$\therefore Nu_x = 42.708$$

$Re_x = 12586.532$
$Pr = 0.714$
$x_0 = 0.017m$
$x = 0.04m$

iii) Determining the Convection coefficient

$$h_x = \frac{Nu_x \bullet k}{x}$$

$$h_x = \frac{42.708 \times 0.0263}{0.04} = 28.081$$

$$\therefore \overline{h} = 2 \times h_x = 2 \times 28.081 = 56.161$$

$Nu_x = 42.708$
$k=0.0263W/m^2$
$x = 0.04m$

c) Heater and carrier chips, and the substrate considered separately

The final approach was to consider each body of the assembly individually, i.e. to consider independent h values for the heater chip, carrier chip and the substrate. This was thought to be a better approach due to the widely different temperatures of the parts of the assembly that made an average, or overall, h coefficient inappropriate. Most of the heat loss from the assembly was expected to be through the heater and carrier chips, as opposed to a longer route through the substrate due to the distance the substrate is from the heat source (the heater chip), the thermal impedances encountered and the comparatively poor thermal conductivity of FR4 compared to silicon.

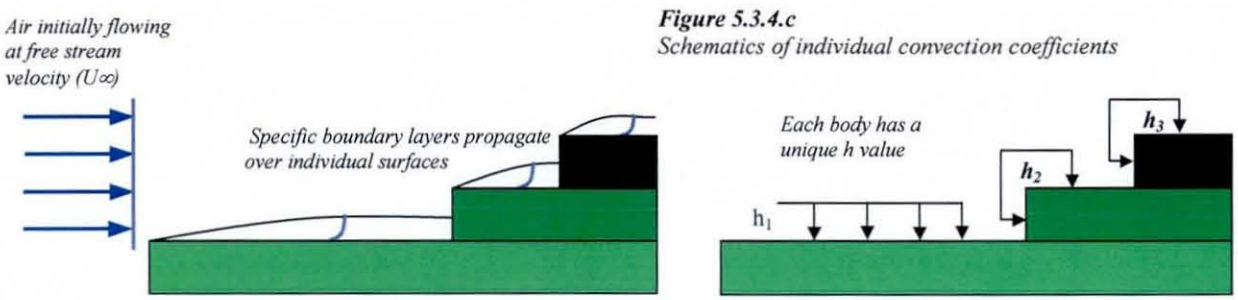


Figure 5.3.4.c
Schematics of individual convection coefficients

Heater chip:
i)Determine Reynolds number

$$\mathrm{Re}_x = U_\infty \frac{x}{v}$$
$$\therefore \mathrm{Re}_x = 943.99$$

U_∞= 5m/s
x=3x10^{-3}m
v=15.89x10^{-6}m^2/s

ii) Determining Nusselt number:
No unheated starting length so equation 5.3.3.5ai should be used

$$Nu_x = 0.332\left(\mathrm{Re}_x^{\frac{1}{2}}\right)\left(\mathrm{Pr}^{\frac{1}{3}}\right)$$
$$\therefore Nu_x = 0.332\left(943.99^{\frac{1}{2}}\right)\left(0.714^{\frac{1}{3}}\right)$$
$$\therefore Nu_x = 0.332(30.724)(0.894)$$
$$\therefore Nu_x = 9.117$$

Re_x= 943.99
Pr =0.714
x =3x10^{-3}m

iii) Determining the Convection coefficient

$$h_x = \frac{Nu_x \bullet k}{x}$$
$$h_x = \frac{9.117 \times 0.0263}{3 \times 10^{-3}\, m} = 79.926$$
$$\therefore \overline{h} = 2 \times h_x = 2 \times 79.926 = 159.852$$

Nu_x= 9.117
k=0.0263W/m^2
x =3x10^{-3}m

Carrier chip:

i)Determine Reynolds number

$$\mathrm{Re}_x = U_\infty \frac{x}{v}$$

$$\therefore \mathrm{Re}_x = 1887.98$$

U_∞= 5m/s
x=6x10^{-3}m
v=15.89x10^{-6}m^2/s

ii) Determining Nusselt number:

No unheated starting length so equation 5.3.3.5ai should be used

$$Nu_x = 0.332\left(\mathrm{Re}_x{}^{\frac{1}{2}}\right)\left(\mathrm{Pr}^{\frac{1}{3}}\right)$$

$$\therefore Nu_x = 0.332\left(1887.98^{\frac{1}{2}}\right)\left(0.714^{\frac{1}{3}}\right)$$

$$\therefore Nu_x = 0.332(43.451)(0.894)$$

$$\therefore Nu_x = 12.893$$

Re_x= 1887.98
Pr =0.714
x =6x10^{-3}m

iii) Determining the Convection coefficient

$$h_x = \frac{Nu_x \bullet k}{x}$$

$$h_x = \frac{12.893 \times 0.0263}{6 \times 10^{-3}\, m} = 56.516$$

$$\therefore \overline{h} = 2 \times h_x = 2 \times 56.516 = 113.033$$

Nu_x= 12.893
k =0.0263W/m^2
x =6x10^{-3}m

Substrate:

i)Determine Reynolds number

$$\mathrm{Re}_x = U_\infty \frac{x}{v}$$

$$\therefore \mathrm{Re}_x = 12586.532$$

U_∞= 5m/s
x=0.04m
v=15.89x10^{-6}m^2/s

ii) Determining Nusselt number:

No unheated starting length so equation 5.3.3.5ai should be used

$$Nu_x = 0.332\left(\mathrm{Re}_x{}^{\frac{1}{2}}\right)\left(\mathrm{Pr}^{\frac{1}{3}}\right)$$

$$\therefore Nu_x = 0.332\left(12586.532^{\frac{1}{2}}\right)\left(0.714^{\frac{1}{3}}\right)$$

$$\therefore Nu_x = 0.332(112.19)(0.894)$$

$$\therefore Nu_x = 33.291$$

Re_x= 12586.532
Pr =0.714
x =0.04m

iii) Determining the Convection coefficient

$$h_x = \frac{Nu_x \bullet k}{x}$$

$$h_x = \frac{33.291 \times 0.0263}{0.04m} = 21.888$$

$$\therefore \overline{h} = 2 \times h_x = 2 \times 21.888 = 43.777$$

Nu_x= 33.291
k=0.0263W/m^2
x =0.04m

Effect of "h" Value on FE Model Results

A summary of the relevant convection criteria is shown in table 5.3.4. a. The results of using the different convection heat transfer coefficients were evaluated against an actual assembly. The MCM assembly mounted on FR4 substrate where the assembly was powered by 1W and left to run to steady state with airflow of 5m/s as described in section 4.3. The results of the temperature of each body were assumed to be an adequate benchmark for evaluating the heat transfer coefficients and are displayed in table 5.3.4b. The temperature values used were taken from the relevant thermal image in section 4.3.

Modelled parameters	***Coefficient value (W/m^2)***
a) Unheated starting length (1.5m)	13.15
b) Unheated starting length(0.04m)	56
c) Separate h coefficients	159 (heater), 113 (carrier), 43 (substrate)

***Table 5.3.4.a**: Summary of Convection values used*

Scenario	***Heater chip (°C)***	***Carrier chip (°C)***	***Substrate near MCM (°C)***
Control exp	90	82	57
a) D=1.5m	166.6	160.3	109.1
b) D=0.04m	112	105.4	64.9
c) Specific h values	87	81.4	54

***Table 5.3.4b:** Summary of values obtained*

It can be seen that scenario c (where each body had specific coefficients applied) provided the best agreement with the practical trial, with only a marginal discrepancy. Average coefficients for the entire assembly as used in sections a and b greatly overestimated the temperature reached, therefore they were considered inadequate ways of obtaining the convection heat transfer coefficient.

The size of each body part was observed to be the dominant factor when obtaining the convection coefficient for the MCM. In the case of obtaining an average convection coefficient for the whole assembly, the h value was observed to vary considerably depending on whether the velocity boundary layer was assumed to propagate from the

wind tunnel entrance or the substrate. To investigate this further, two MCMs were secured to two different size FR4 substrates: - measuring 40 and 90mm in the airflow direction respectively. The assemblies were run to steady state concurrently and resultant thermal characteristics were recorded. A negligible difference was observed in terms of both the maximum temperature reached and the overall temperature distributions. However, when the different scenarios were modelled using method "b", the h value changed from 57 (40mm substrate) to 38.5 (90mm substrate). According to the FE models the maximum temperature increased from 112°C to 132°C. Clearly this was unrealistic and an average coefficient for the whole assembly is inadequate for this example.

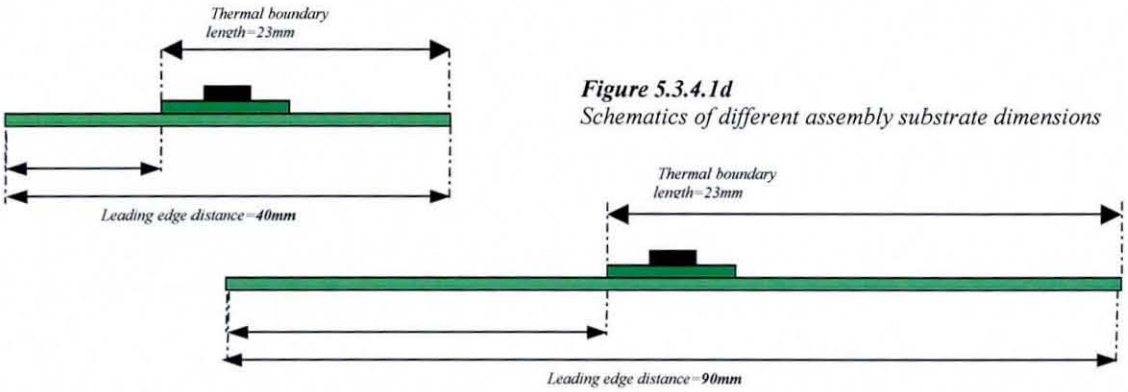


Figure 5.3.4.1d
Schematics of different assembly substrate dimensions

The influence of surface radiation: The radiation effects were considered and found to be negligible within this modelling work, as the high airflow rates were seen to dominate the heat transfer mode. Two thermal analyses of the model were ran, one with 0% emissivity and the other with 100% emissivity, and a negligible difference in the final thermal profiles was observed.

5.4: Material Properties

All of the material properties used for the FE modelling are summarised in table 5.4 and apart from the solder, all were assumed to be linear elastic, isotropic. Apart from the LTCC data which was supplied by Dupont and the adhesive data supplied by Chomerics, all material properties were initially obtained from **MatWeb** group (25), which is a large searchable database of material properties where manufacturers voluntarily submit material data. Afterwards the data was checked with other values used in previous modelling work (8-11,14,18,20,21) and they were found to be in general good agreement.

FR4 material properties: It was decided that, as the FR4 board was not in direct contact with the solder joints, any stresses present in it and board bending were of little interest (the boards were there to provide the necessary thermal characteristics and a rigid surface to enable connections). Therefore to simplify the model it was decided to model the FR4 board with isotropic material properties. In addition, this also permitted fewer elements to be used when modelling the board hence further simplifying the model.

Material	Young's Modulus (GPa)	Poisson's Ratio	CTE (parts per million/°C)
Silicon (Die)	160	0.28	2.5
Nickel (UBM)	207	0.31	13.1
Aluminium (Bond pad)	68	0.33	24
Adhesive	0.069	0.34	100
FR4 (substrate(MCM))	18.2	0.25	15
Copper (substrate (MCM))	110	0.343	16.4
Alumina (Substrate (DCA))	370	0.22	7.4
LTCC (substrate (DCA))	142	0.17	5.8

***Table 5.4a:** Linear mechanical properties of the materials used*

Material	Thermal Conductivity (W/m-K)	Heat Capacity (J/Kg-°C)	Density (g/m^3)
Silicon	124	702	2330
Nickel (UBM)	60.7	460	8880
Aluminium	210	900	2699
Adhesive	0.37	109	1420
FR4	1.3	810	2520
Copper	385	385	8960
Alumina	30	850	3960
LTCC	3.3	100	3100
63/37SnPb solder	50	173	8600

Table 5.4b: *Thermal properties of the materials used*

The 63/37 SnPb eutectic solder thermal properties were assumed to be constant however the mechanical material properties required careful consideration as the values of the Young's Modulus and the yield strength are known to be heavily dependent on the nature and temperature the initial tensile tests were recorded at. The value obtained initially from the MatWeb database was 33 GPa which was in good agreement with the value used by Pang et al (10), though Stoyanov et al (20,21) model the Young's Modulus with a much lower value of 10 GPa presumably accounting for the lower values at high temperatures. Hong and Burrell (8,9), Lau (11), Lee (18) and Yang et al(14) all modelled their solder properties as temperature dependent. The final temperature dependent solder values used are shown in table 5.4c based on the results of Lau and Lee. According to Lau and Lee, after the applied stress has exceeded the yield strength, eutectic solder behaves super-plastically at high temperatures.

Temperature (°C)	Young's Modulus (GPa)	Poisson's Ratio	Yield Strength (MPa)
-33	39.435	0.36	38
7	33.367	0.37	32
47	27.299	0.37	25
87	21.231	0.38	18
127	15.163	0.38	8

Table 5.4c: *Temperature dependent properties of 63/37SnPb Eutectic solder used*

5.5: Model Generation

This experiment has shown that for identical models with consistent boundary conditions, that depending on the meshing techniques employed by the operator, very different results may be obtained. Though the mesh density and refinement level were shown to be significant factors, the dominant factor proved to be the element type used in the analysis. Using 20 noded hex elements instead of 8 noded ones proved to increase the accuracy of the results. It does this to such an extent that according to the results, an 8 noded mesh would have to be 4 times the density to achieve the equivalent results. However it should be noted that calculation time and computer memory also consequently increase for equivalent results.

Therefore it is recommended that the user initially meshes their models using 8 noded brick elements for a pilot study; to first obtain a general idea of the expected results, but then the 20 noded elements should be used to confirm these results (predictor/ corrector scenario).

It was intended to use the same base-model for both the thermal and stress analysis with the thermal model being run first and the stress analysis performed later. As the model could only be simplified to ¼ symmetry hexahedral elements proved to be the most economical.

For the thermal work the most suitable type of elements offered were eight noded, isoparametric three-dimensional brick elements with tri-linear interpolation (MARC element no 43) with one degree of freedom (thermal). The primary objective of the thermal work was to obtain an overall temperature distribution and for each of the "important" parts of the model there were an adequate number of elements such that there was little benefit in employing 20 noded elements.

For the stress modelling, it was intended to convert the models from 8 noded to 20 noded elements such that a better representation of the interconnection layer would be obtained. However, in attempting to run the model, there were insufficient computer resources available to run such a model so 8 noded distortable elements equivalent to

those from the thermal analysis were used (MARC element no 7). It is acknowledged that this may compromise the accuracy of the stress results, though unfortunately with the available resources there was little that could be done to rectify this issue.

This section first describes how the model was constructed and then how the boundary conditions were applied. In addition the models would have an initial condition applied to the model such that the ambient temperature was room temperature and the solder at a stress free initial state. All the thermal and mechanical analyses were performed using an explicit solver.

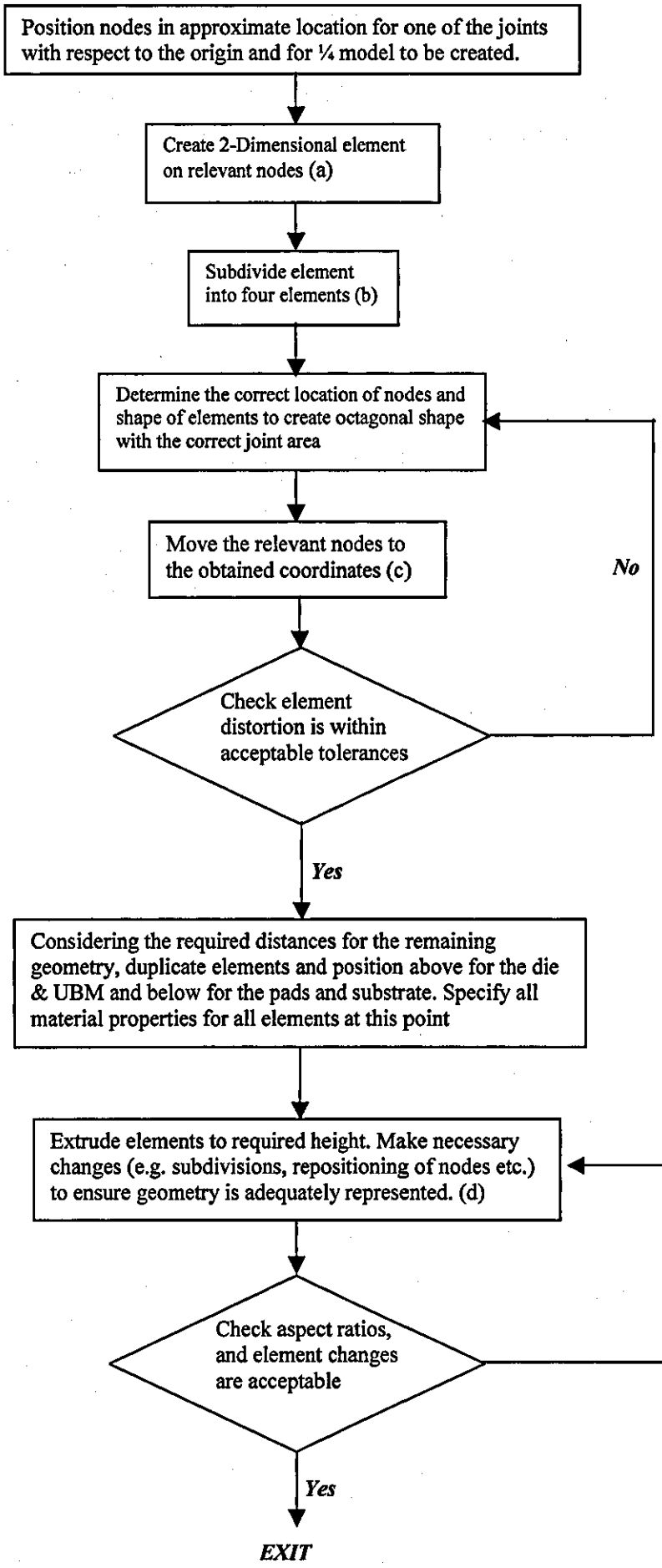
5.5.1: Mesh Generation Procedure

This section describes a generic procedure for creating the FE model geometry for the flip chip assembly. The aim is to describe the overall process how models were created. It neither describes any specific model nor addresses complications that arose when such models were created. The model development was divided into three stages:

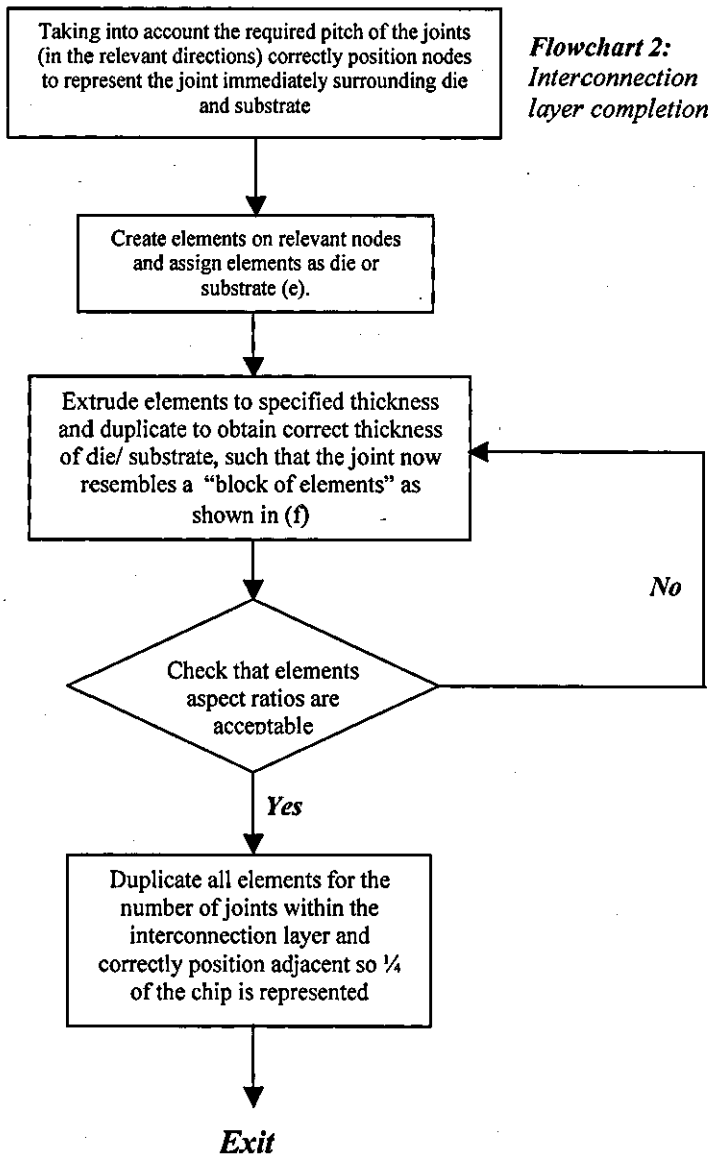
1) Joint creation
2) Interconnection layer completion
3) Remainder of the model geometry

For each of the stages, a flowchart is shown detailing the procedure used. The flowcharts also make reference to some graphical representations of the stages which are also presented.

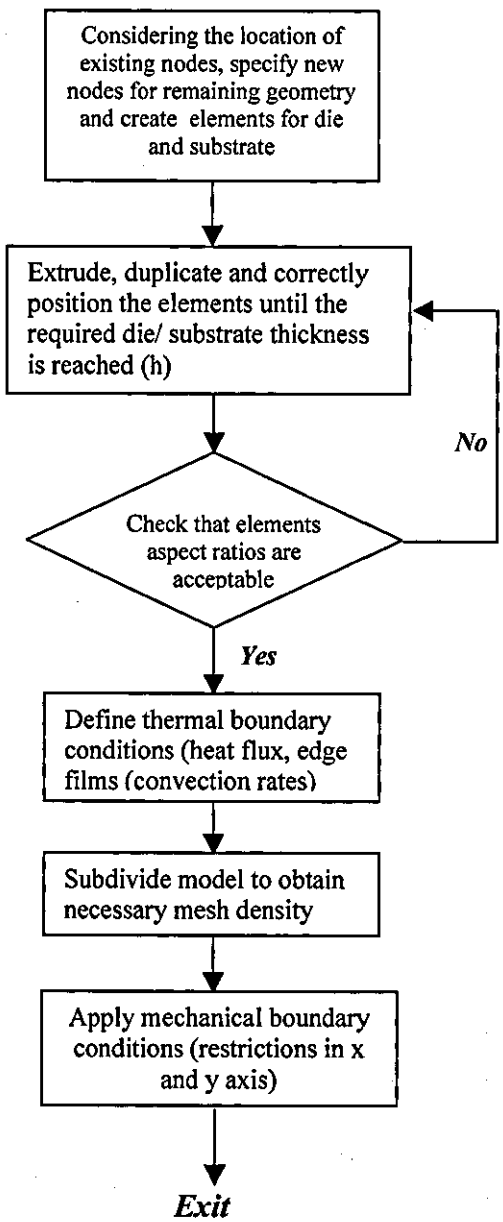
The example used was a generic model that could potentially be modelled utilising $1/8^{th}$ symmetry. However, the models to be considered had different numbers of solder connections depending on which side of the die was of interest (i.e. the sides with the daisy chain had 8 connections while the sides with the 4-point resistance and power measurements had 10 joints) They were also of different pitch so the model could not be modelled using the $1/8^{th}$ symmetry and ¼ was the most it could be viably reduced to.



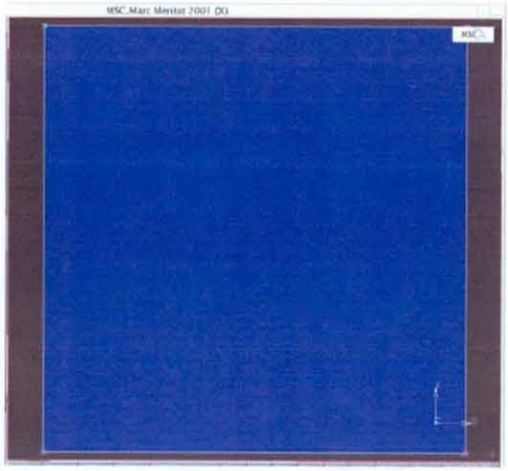
Flowchart 1:
Joint creation



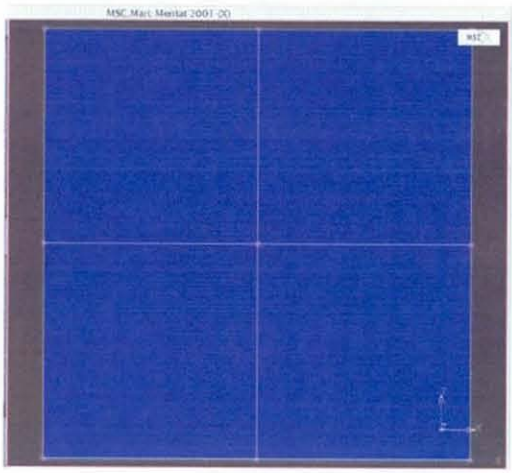
Flowchart 2: *Interconnection layer completion*



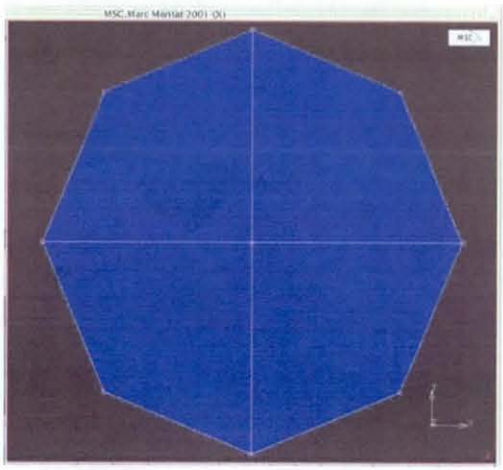
Flowchart 3: Remainder of the geometry



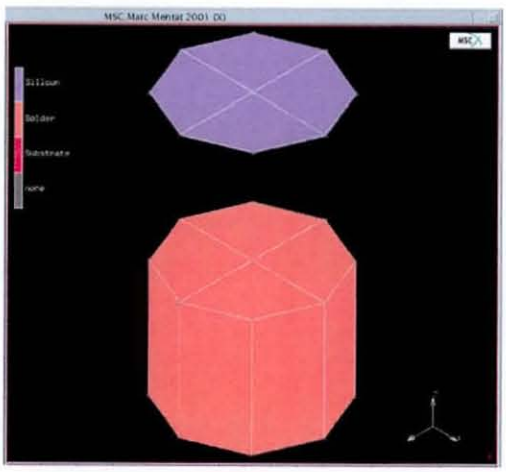
(a) Nodes correctly positioned and element created



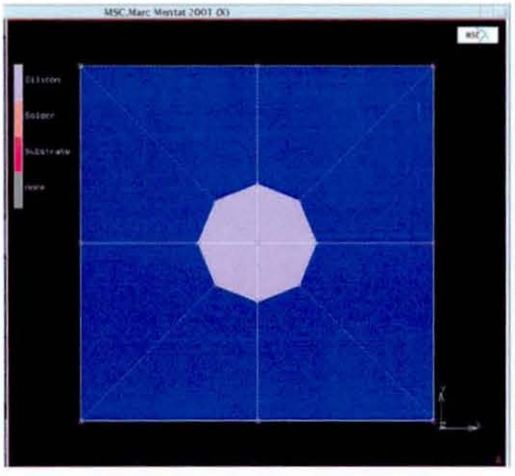
(b) Element was subdivided to create 4 new elements



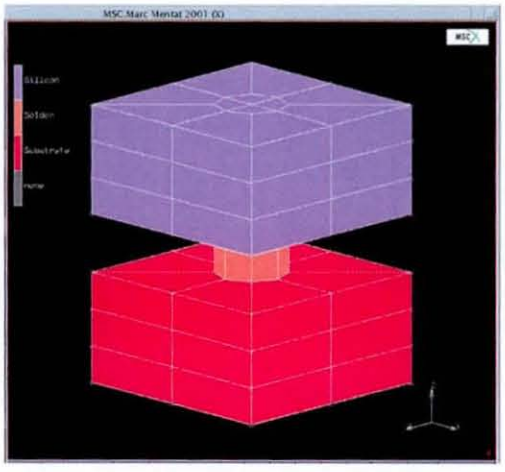
(c) Nodes repositioned using trigonometry to obtain octagonal shape



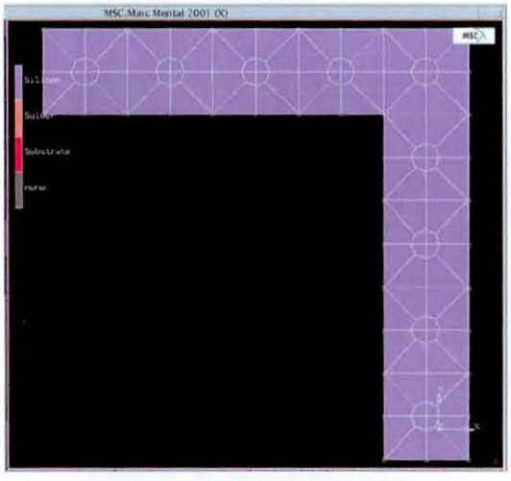
(d) Duplicate original elements created. The original element assigned solder material properties. Other elements assigned silicon and substrate properties

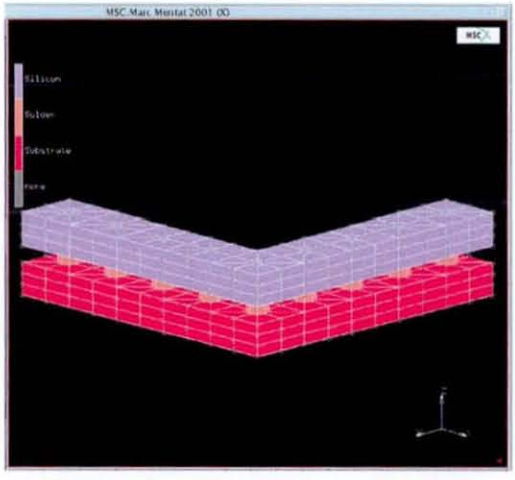


(e) Adjacent nodes placed around silicon and FR4 and elements defined with the correct properties. Finally elements are extruded to correct length

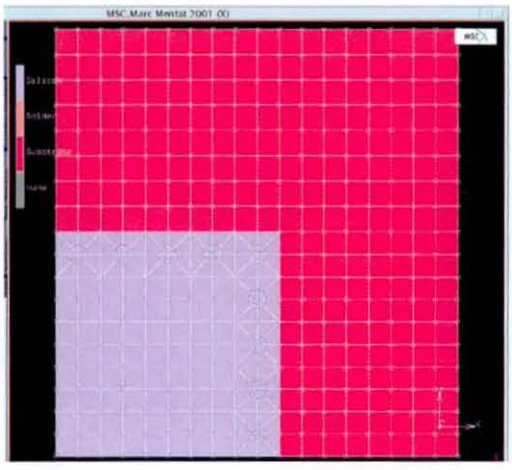


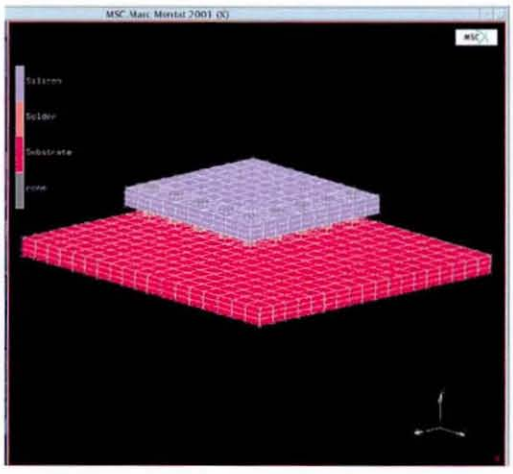
(f) Duplicate elements and place on top of existing elements until desired height was reached, so that a complete joint is created.



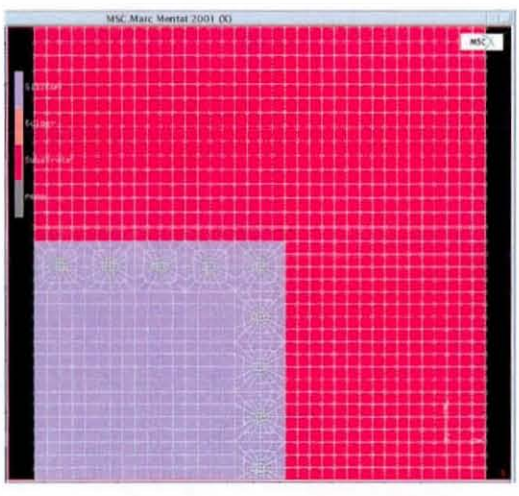


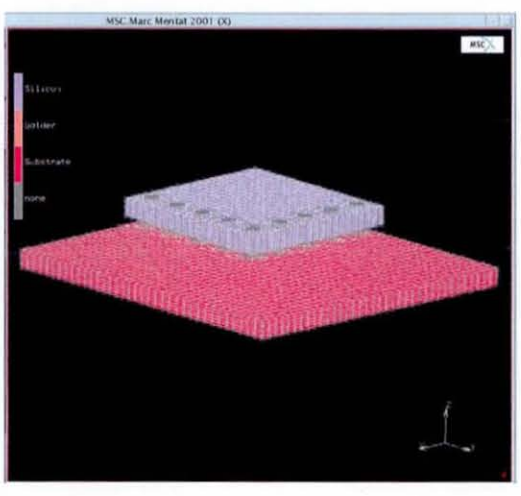
(g) All existing elements were duplicated to create ¼ of all the solder joints in the device





(h) Remaining elements to complete the die and substrate created





(a) All elements subdivided to obtain better mesh density

5.5.2: Boundary Conditions

Thermal Boundary Conditions

Heat flux: The heat source was modelled by applying a heat flux to the area representing the resistive element on the heater chip as shown below. The user specifies the value of the heat flux for the designated area in units of W/m^2.

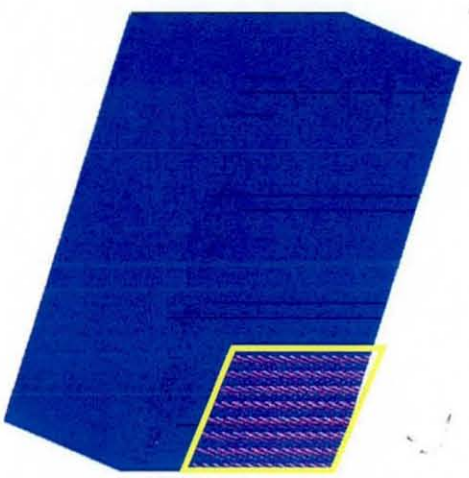
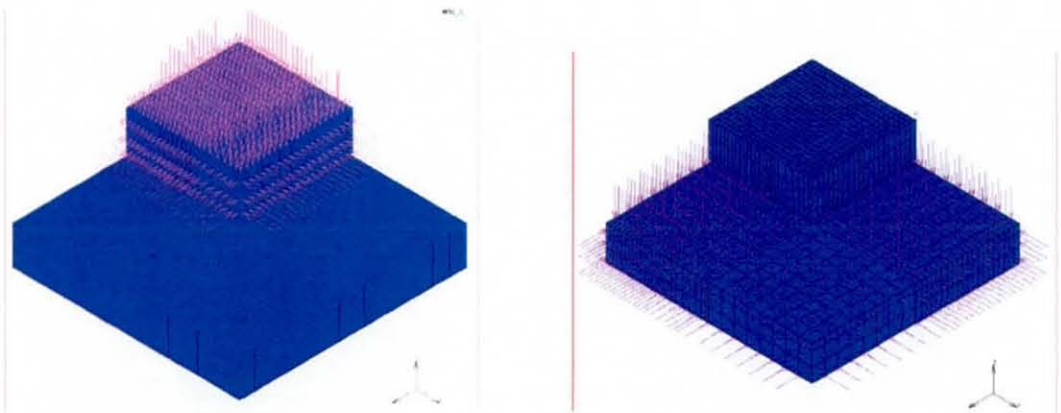


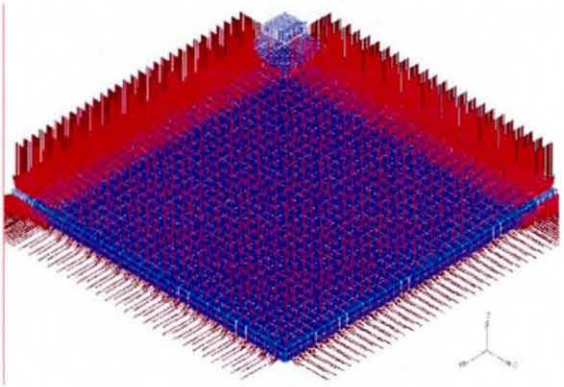
Figure 5.5.2a: *Heater chip viewed from the bottom, the highlighted are is where the heat flux was aplied*

Convection coefficient: The FE model had to simulate the cooling conditions the test-pieces were subject to in the trials. This involved a forced convection scenario where air is forced to flow over the top of the assembly at a constant rate. The convection mechanism was applied utilising the procedure described in section 5.3.4 (case c) where each of the bodies had their own specific coefficient. The temperature of the cooling air was always specified as x˚C.



Heater Chip Carrier Chip

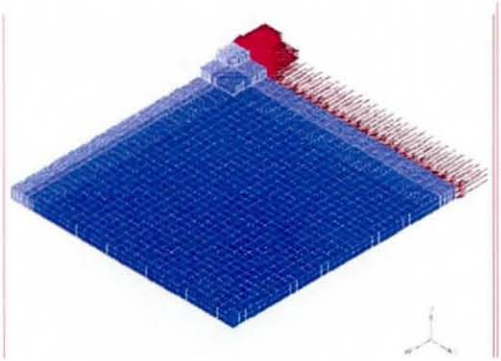
Figure 5.5.2b: *Surfaces where the convection coefficient was applied on the heater and carrier chips*



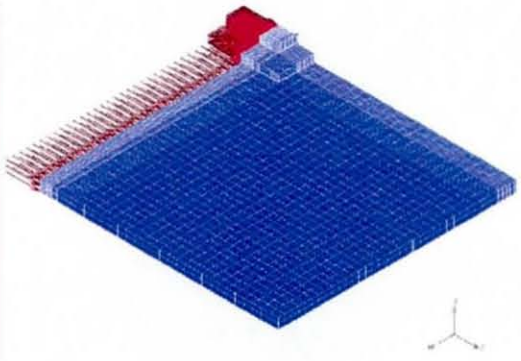
Substrate

Mechanical Boundary Conditions

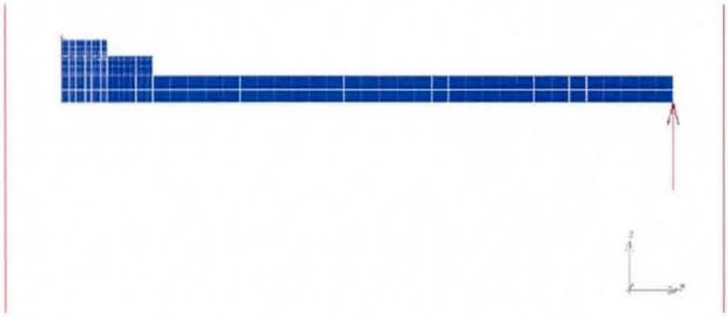
The model was a ¼ representation of the assembly so there was symmetry assumed about the x and y axes; these symmetric boundary conditions were applied by fixing all the nodes along the relevant axis such that they could not move during the analysis. A further restriction in the z axis was also applied. In addition, the thermal characteristics at this edge were perfectly insulated (¼ symmetry)



Symmetry about the x axis



Symmetry about the y axis



z axis restriction

Figure 5.5.2c: *Symmetry lines about the x and y axes to ensure ¼ model is represented. Restriction of one node in the z axis to maintain stability of the model*

5.6: Conclusion

A method for creating the FE models has been derived. The required mesh density was evaluated using a thermal simulation of a block heated with a flux and observing when suitable convergence had occurred. Various refinement methods were also evaluated to investigate how areas of interest could be represented adequately without compromising the overall analysis time and memory usage; a suitable method was found and will be employed in subsequent FE models.

The necessary convection coefficient was obtained from first principles, and an endeavour was made to ensure that it was temperature independent. This meant assuming that cooling is solely by forced convection.

Combining the two allowed a model development procedure to be derived utilising the mesh density, refinement and boundary conditions used. However creating the model was inevitably labour intensive, and in addition the overall geometry is semi-infinite compared to the interconnection layer. Therefore there is still likely to be a large number of elements in the subsequent models, despite the best efforts to reduce the number of elements. Therefore adequate time and resources must be provided to allow the models to complete.

References

1) **O.C. Zienkiewicz and R.L. Taylor** "The Finite Element Method (Volume 1)" *McGraw Hill, 2000 (5th edition)*

2) **A.O. Ogunjimi, D.C. Whalley, and D.J. Williams** "A Comparison of Modelling Methods for Electronic Interconnect Studies" *IEEE Conference paper, 1993, pages 871-6*

3) **Dominiek Degryse, Alcatel Bell** "Simulation of Solder Joint Reliability for CBGA packages" *IMAPS 13th European Microelectronics and Packaging Conference and Exhibition, 2001, pages 360-5*

4) **E. Madenci, S Shkaeayev and R Mahajan,** "Potential Failure Sites in a Flip Chip Package With and Without Underfill" *Journal of Electronic Packaging Volume 120, December 1998, pages 336-341*

5) **H. Doi, K Kawano, A Yasukawa, T Sato** "Reliability of Underfill Encapsulated Flip Chips with Heat Spreaders" *Journal of Electronic Packaging Volume 120, Dec 1998, pages 322-8*

6) **Tyan-Min Niu, Bahgat G Sammakia (fellow IEEE) Sanjeev Sathe** "Void Effect Modelling of Flip Chip Encapsulation on Ceramic Substrate" *IEEE Transactions on Components and Packaging Technology Volume 22 No 4, December 1999, pages 484-7*

7) **Stylianos Micaelides, Suresh Sitaraman** "Die Cracking and Reliable Die Design for Flip Chip Assemblies" *IEEE Transactions on Advanced Packaging Volume 22 No 4, November 1999, pages 602-13*

8) **Bor Zen Hong, Lloyd G Burrell** "Non linear Finite Element Simulation of Thermo Viscoplastic Deformation of C4 Solder Joints in High Density Packaging under Thermal Cycling" *Thermal Phenomena in Electronic Systems 1994 I-THERM IV 'Concurrent Engineering and Thermal Phenomena' InterSociety Conference,4-7 May 1994, pages 117 - 25*

9) **Bor Zen Hong & Lloyd G Burrell** "Modelling Thermally Induced Viscoplastic Deformation and Low Cycle Fatigue of CBGA Solder Joints in a Surface Mount Package" *IEEE Transactions on Components Packaging and Manufacturing Technology Part A Volume 20 No 3 Sept 97, pages 280-5*

10) **John HL Pang, Tze Ing Tan, Suresh K Sitaraman** "Thermo-Mechanical Analysis of Solder Joint Fatigue and Creep in a Flip Chip On Board Package Subject to Temperature Cyclic Loading" *48th IEEE 1998 Electronics Components and Technology Conference, 25-28 May 1998, pages 878 - 83*

11) **John Lau** "Solder Joint Reliability of a Low Cost Chip Size Package" (Introductory Invited Paper) *Microelectronics Reliability 38, 1998, pages 1519-29*

12) **Qizhou Yao, Jianmin Qu, Shean X Wu** "Solder Fatigue in Two Chip Scale Packages" *1999 International Symposium on Microelectronics, Proceedings of the Society of Photo-Optical Instrumental Engineers (SPIE) 3906, 1999, pages 563-70*

13) **Cheng Bo, Wang Li, Zhang Qun, Gao Xia, Xie Xiaoming, Wolfgang Kempe** "Flip Chip Solder Joint Reliability under Harsh Environment" *Solder and Surface Mount Technology, 15th March 2003, pages 15-20*

14) **D.G. Yang, G.Q. Zhang, Leo j Enst, Cornelis, Van't Hof, J Caers, H. J Bresser , J H Janssen** "Investigation of Flip Chip Solder Joint Fatigue with Cure Dependent Underfill Properties" *IEEE Transactions on Components and Packages Technology Volume 26 No 2, June 2003*

15) **Carlton E Hanna, Suresh K Sitaraman** "Role of Underfill Materials and Thermal Cycling on Die Stresses" *EEP-Volume 26-1 Advances in Electronic Packaging Volume 1 ASME, 1999, pages 795-801*

16) **Robert Darveaux** "Effect of Simulation Methodology on Solder Crack Growth" *Electronic Components and Technology Conference, 2000 Proceedings. 50th, 21-24 May 2000, pages 1048 - 58*

17) **Schubert, R Dudek, R Leutenbauer, R Doring, H Opperman, B Michel, H Reichi, D Baldwin, J Qu, S Sitaraman, M Swaminathan, C.P. Wong, R Tummala** "Do Chip Size Limits Exist for DCA?" *1999 International Symposium on Advanced Package Materials, 1999, pages 150-57*

18) **Chin C Lee** "Analysis and Study of Stress in the Structure of High Power Laser Chip Bond on a Metallic Joint" *Final Report for MICRO 99-069, 1999 pages 1-8*

19) **Charlie j Zhai, Sidharth, Richard Blish II** "Board Level Solder Reliability Vs Ramp rate and Dwell Time Temperature Cycling" *IEEE transactions on Device and Materials Reliability Vol 3 No 4, Dec 2003, pages 207-12*

20) **S. Stoyanov, C Bailey, H Lu M Cross** "Integrated Computational Mechanics and Optimisation for Design of Electronic Components" *3rd Optimisation in Industry Conference Tuscany Italy, 17-22nd June 2001, pages 57-70*

21) **S. Stoyanov, C Bailey, H Lu M Cross** "Solder Joint Reliability Optimisation" *APACK 2001 Conference on Advances in Packaging, Singapore, 2001*

22) **Jerry E Sergent & Al Krum** "Thermal Management Handbook for Electronic Assemblies" *McGraw Hill, 1998*

23) **Louis C Burmeister** "Convective Heat Transfer (second edition)" *John Wiley and Sons, 1993*

24) **Kirk D. Hagen** "Heat Transfer with Applications" *Prentice Hall, 1999*

25) http://www.matls.com

Chapter 6: Finite Element Model Results (Thermal)

6.1: Introduction

The model was built with the pad and chip dimensions as specified in chapter 3. Having constructed an FE model of the MCM assembly with the necessary refinement and boundary conditions considered, it was then necessary to validate the model predictions compared with the thermal profiles obtained from the "real" devices. This was carried out for both steady state and transient results over the same range of power levels.

Model Construction and Boundary Conditions

Principally, it was required to check how the temperature of each body within the FE model compared to that of the experimental devices under the same test conditions. The power levels and the cooling airflow rates were the two variables that were manipulated in the FE model, the power levels were represented as a heat flux applied to the surface of the underside of the heater chip over the same area as the heater on the "real" device. The cooling airflow on each of the bodies was represented as a face film applied to each of the surfaces as described earlier in section 5.5.1. The analysis was performed both for FR4 and copper substrates. The fact that it is possible to amend the material properties with relative ease demonstrates the principal advantage of creating the FE model as opposed to manufacturing a physical test vehicle (all the manufacturing processes need not be considered, and the inevitable hurdles faced when implementing a manufacturing method are also bypassed).

The model results were extracted utilising the post processing features of **MENTAT**; as described in chapter 5, the results were presented at a global level as a thermal plot using contour-bands, as well as at a nodal level (as a graph). The global thermal plots enabled a direct comparison with the equivalent thermal images obtained from the steady state experimental analysis, while the graphs of the selected nodes could be compared with the graphs of the transient experimental work. If a record of the selected node was kept then the same node could be examined in subsequent analysis,

therefore allowing for consistency and repeatability as only a change in the desired variable could have resulted in the new profile. In total 5 nodes were selected from the model such that each of the critical areas of the model were represented. Table 6.1 shows the node reference numbers and their location on the assembly and their locations on the FE model are shown in figures 6.1 a and b. The node on the centre-top of the heater chip (a) was selected to allow for a basic temperature comparison with that obtained from the experimental MCM. Nodes on the side (b) and the corner of the carrier chip (c) were selected to observe the temperature gradient over the carrier chip. Two further nodes were chosen on the substrate: one close to the MCM (d) and the other much further away (e) again to observe the temperature distribution.

Location	Comments
6992 (a)	Heater chip: (located centrally on the top surface)
8339 (b)	Carrier chip: (on the side of the chip
9679 (c)	Carrier chip: (located on the corner)
16619 (d)	Substrate: Located near the MCM
20850 (e)	Substrate: Located at the far corner

Table 6.1: *Location of nodes and description of representation*

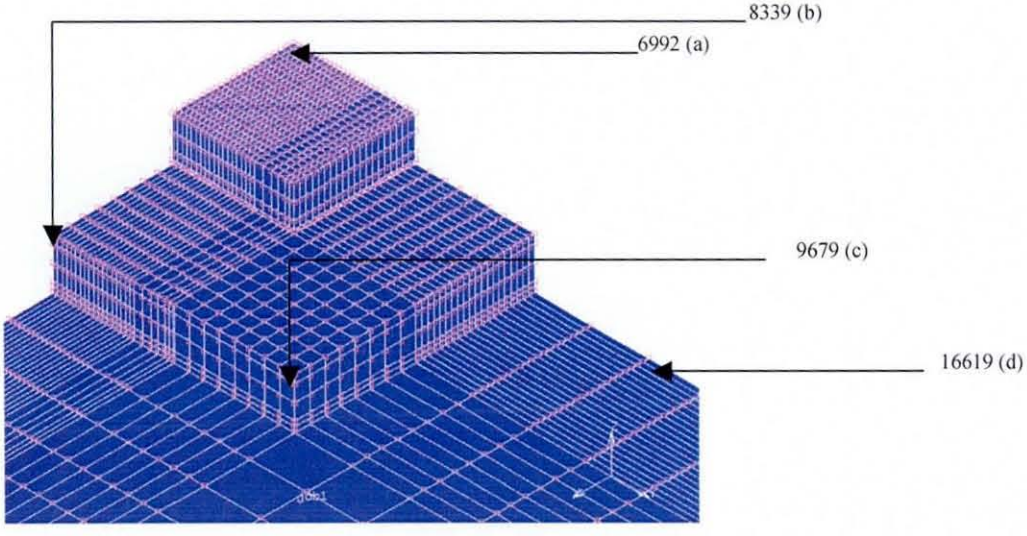
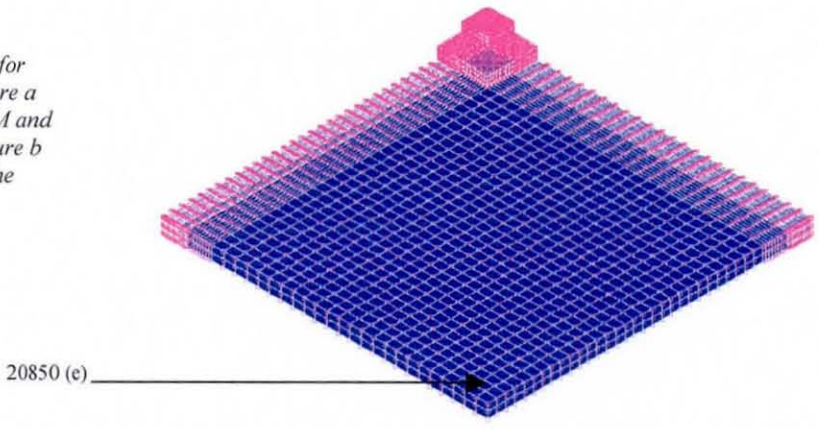


Figure 6.1a & b:
Location of nodes chosen for the thermal analysis. Figure a displays those on the MCM and those close to it, while figure b shows the node far from the MCM.



This chapter presents the results for the thermal FE analysis, first for steady state and then transient conditions. Following this is a discussion on how the different conditions compare with each other and finally a comparison with the FE results and the "real" experiments is carried out.

6.2: Steady State Finite Element Results

A heat transfer steady state analysis was performed on the FE models when FR4 and copper substrates were used with an airflow rate of 5m/s for consistency with the previous thermal profiling work, power ranges between 0.1 to 1.4W in increments of 0.1W were modeled. Within the model, the air temperature was set to 20°C, however in the experimental work air temperatures of 26 and 23°C were recorded during the tests of samples with FR4 and copper substrates respectively. The temperature change within the model was therefore used for subsequent analysis and comparison with the "real" test devices.

6.2.1: FR4 Results

Figure 6.2.1a &b show the temperature distribution obtained for an MCM mounted on FR4 with a power level of 0.2 & 1.2W respectively. It can be seen that while the temperatures reached may be different, the overall thermal profiles are very similar; i.e. both the heater and carrier chips show a fairly uniform temperature across their top surfaces, with the heater chip exhibiting the greater temperature, though a slight gradient can be seen with a finer temperature range in the cross-sections of figures

6.2.1c. Figures 6.2.1c & d also show the interconnection layer with the corresponding temperature scale standardised to highlight differences over the MCM. The interconnection can be seen to be at a temperature in between the heater and the carrier chip. The figures also show the temperature gradient across the heater chip. In figure 6.2.1d, the area immediately below the solder joint on the right hand side is marginally warmer than the rest of the chip where there is a direct link to the hotter heater chip. Figure 6.2.1e & f focus on the temperature distribution on the substrate. The area beneath and local to the MCM is very close to the MCM temperature but it can be seen that the temperature rapidly diminishes as the distance from the substrate increases.

As expected, the heater chip showed the highest temperature and most of the heat is conducted through the solder joints to the carrier chip, which is at a marginally lower temperature, and finally through the adhesive layer and to the substrate. As also anticipated, there is an even temperature distribution across the heater and carrier chips but on the substrate the temperature rise is very local to the MCM.

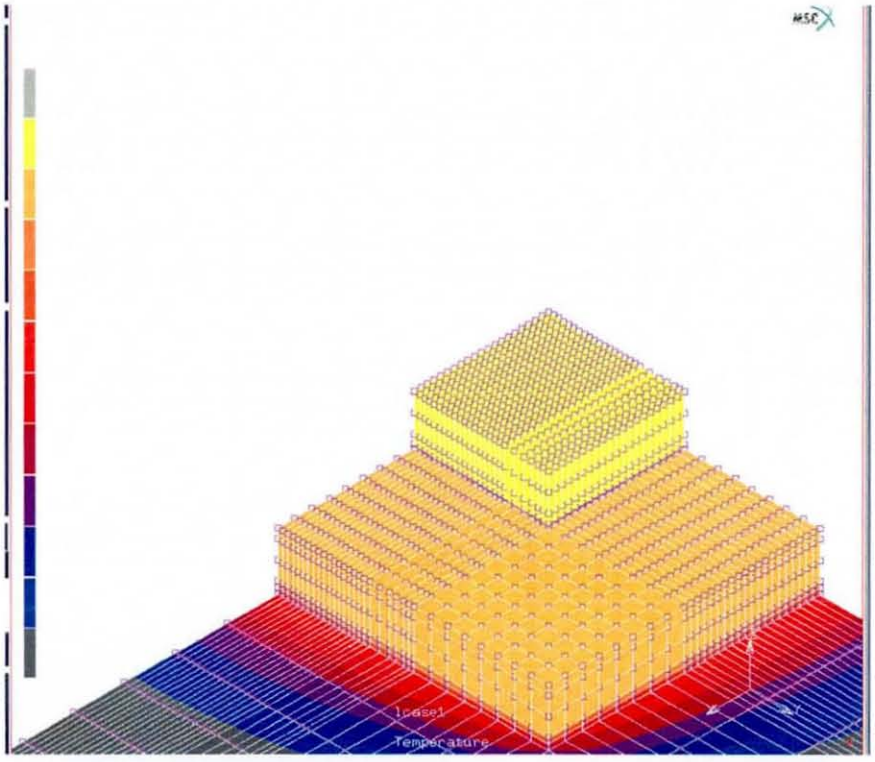


Figure 6.2.1a: *Temperature distribution of FE model of the MCM on FR4 with 0.2W powering the chip. The minimum temperature displayed is 26.5°C while the maximum is 33.5°C . The view is isometric*

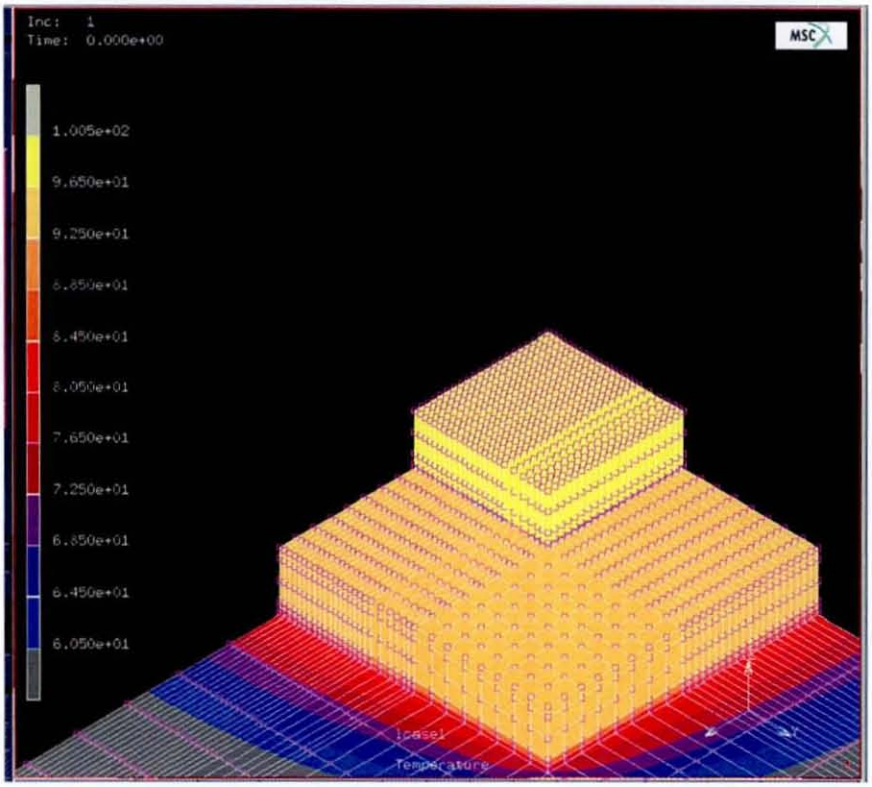


Figure 6.2.1b: *Temperature distribution of FE model of the MCM on FR4 with 1.2W powering the chip. The minimum temperature displayed is 60.5°C while the maximum is 100.5 °C . The view is isometric.*

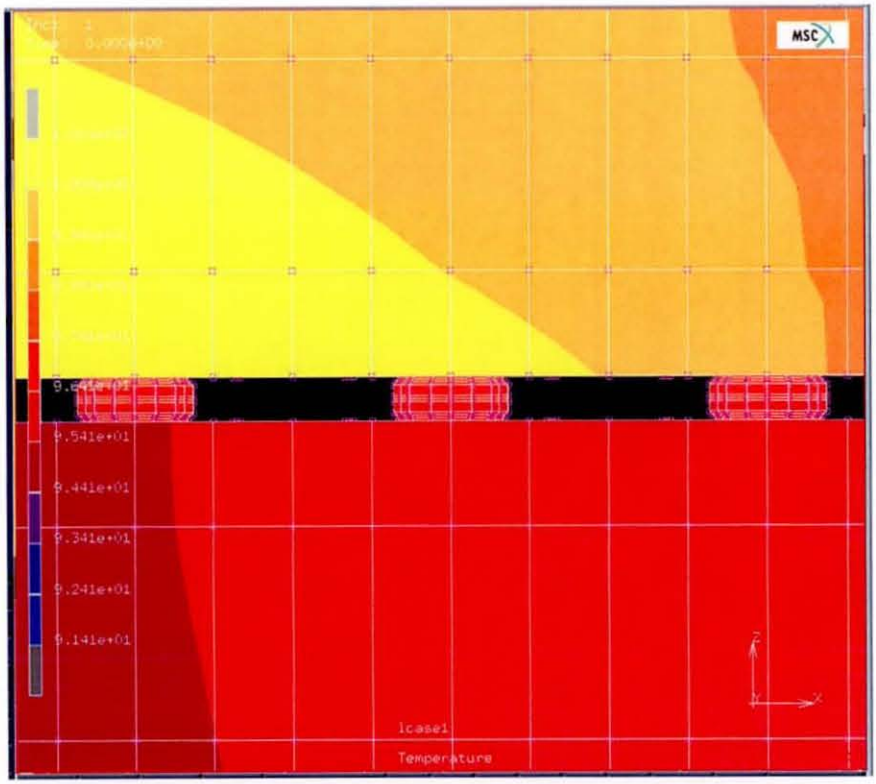


Figure 6.2.1c: *Close up of the interconnection layer when the MCM is powered with 1.2W. Note the significant temperature difference between the heater and carrier chips. The view is taken from the x axis (origin). The temperature scale is from 101 to 91°C*

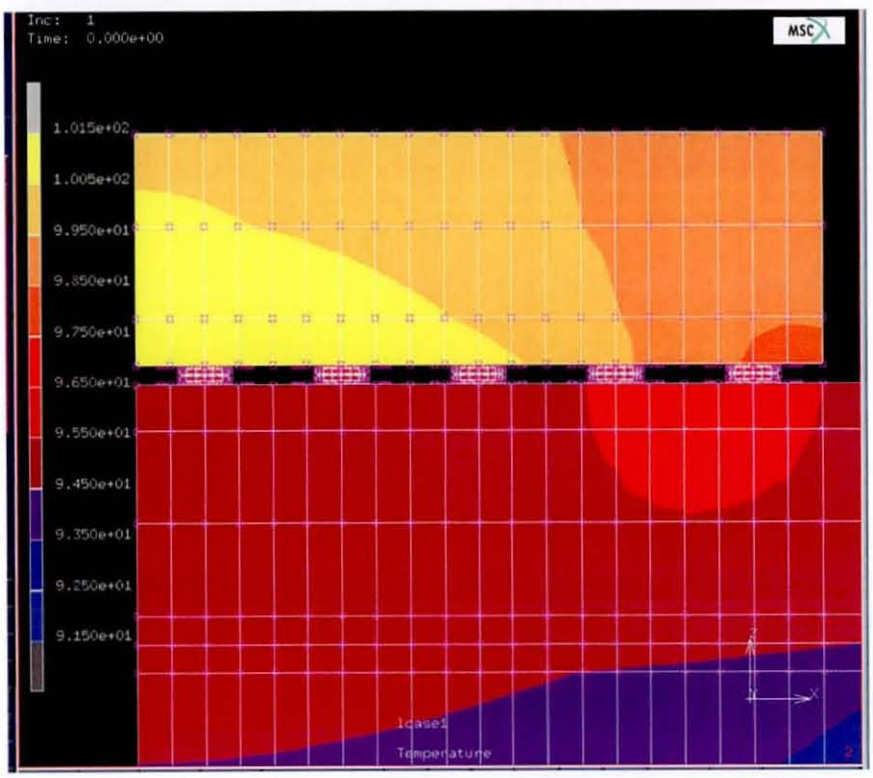


Figure 6.2.1d: *View of the interconnection layer with the heater and part of the carrier chip included. Note the warmer region of the carrier chip to the right hand side of the chip. As the bump on the RHS is close to the origin while the others are further away from the origin.*

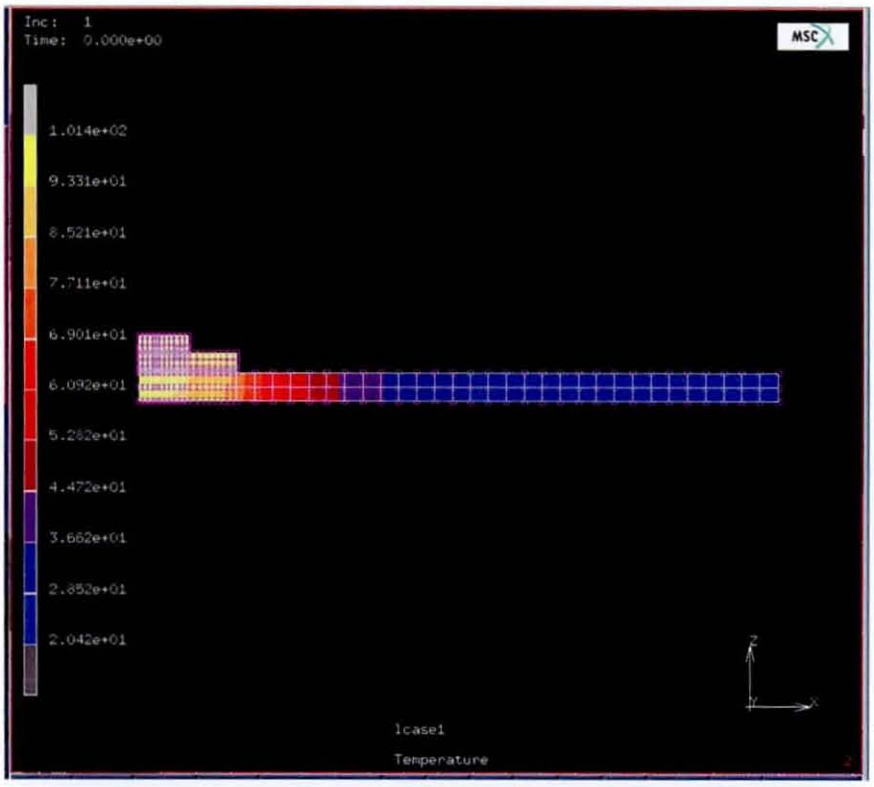


Figure 6.2.1e: *Full view of the side of the assembly. Note how on the substrate (FR4)the temperature decreases the further away from the MCM. The view is taken from the x-axis and the power level used was 1.2W and the temperature range is from 101 to 20°C*

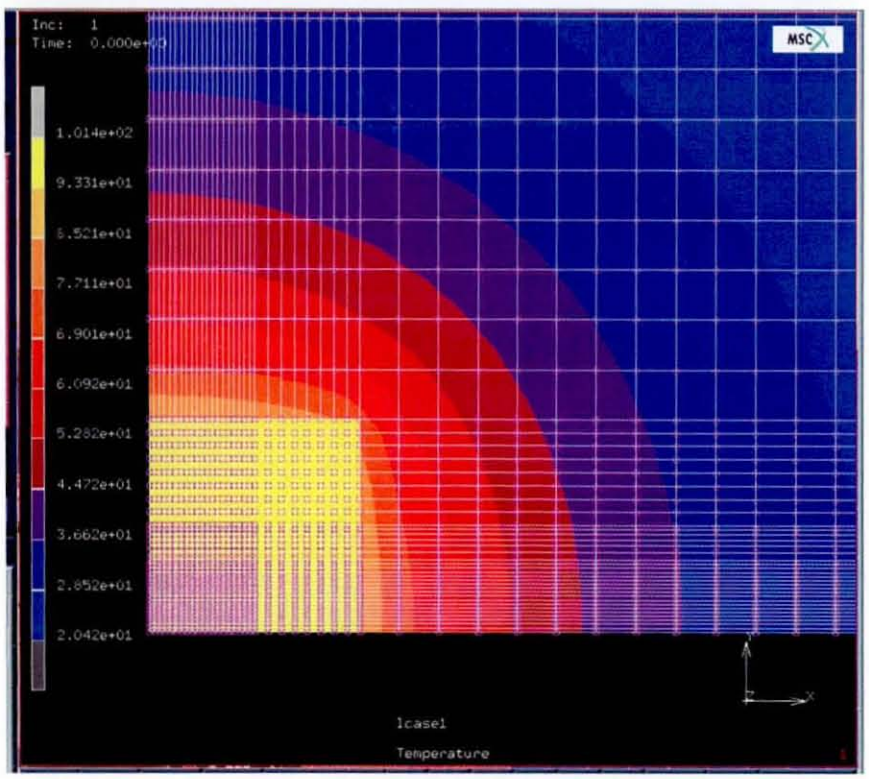


Figure 6.2.1f: *View of the MCM with part of the substrate (FR4). Note how the temperature appears uniform across the MCM but only part of the substrate close to the MCM is significantly warm. The view is a plan view (from above) with the power level of 1.2W used and the temperature range is 101 to 20°C*

Effect of Input Power Level on Temperature for MCMs on FR4 Substrate

Table 6.2.1 shows the temperatures obtained at the selected nodes for their respective power levels while graph 6.2.1g shows the relationship graphically. Despite the different temperatures reached, the models exhibit similar thermal profiles. Observing table and graphs 6.2.1 a linear relationship exists between the power input and the resultant temperature for the selected nodes. For the heater chip, a 6.7°C rise in temperature was observed for every 0.1W increase at the node on the heater chip (a). The temperature changes exhibited by the nodes located on the carrier chip were both seen to be over 90% of that of the node located on the heater chip; (the node on the side of the carrier chip (b) was observed to be 92% of (a) while the corner (c) was 91%) indicating that there is very little temperature change across the carrier chip and that there is substantial heat transfer from the heater to the carrier chip, whilst some heat is lost through convection. A more rapid temperature decline could be seen across the substrate, where a significant increase in temperature could be seen in the node close to the MCM (d), while there was an almost negligible temperature rise in the node located at the far corner (e). This was attributed to the poor conductivity of the FR4 substrate.

Equivalent power input (W)	***Heater chip node (a)*** (°C)	***Carrier chip-side node (b)*** (°C)	***Carrier chip corner node (c)*** (°C)	***Substrate (near MCM) node (d)*** (°C)	***Substrate (far corner) node (e)*** (°C)
0.1	6.7	6.2	6.1	3.4	0
0.2	13.4	12.4	12.2	6.8	0.1
0.3	20.1	18.6	18.4	10.2	0.1
0.4	26.8	24.8	24.6	13.6	0.1
0.5	33.5	31	30.7	17	0.2
0.6	40.2	37.2	36.9	20.4	0.2
0.7	46.9	43.4	43	23.8	0.3
0.8	53.6	49.6	49.1	27.2	0.3
0.9	60.3	55.8	55.3	30.6	0.3
1	67	62	61.5	34	0.4
1.1	73.7	68.3	67.6	37.4	0.4
1.2	80.4	74.5	73.7	40.8	0.4
1.3	87.7	80.7	79.9	44.2	0.4
1.4	93.8	86.9	86	47.5	0.5

Table 6.2.1a: *Table displaying the equivalent power input and the corresponding temperature rise above the ambient temperature (20°C)*

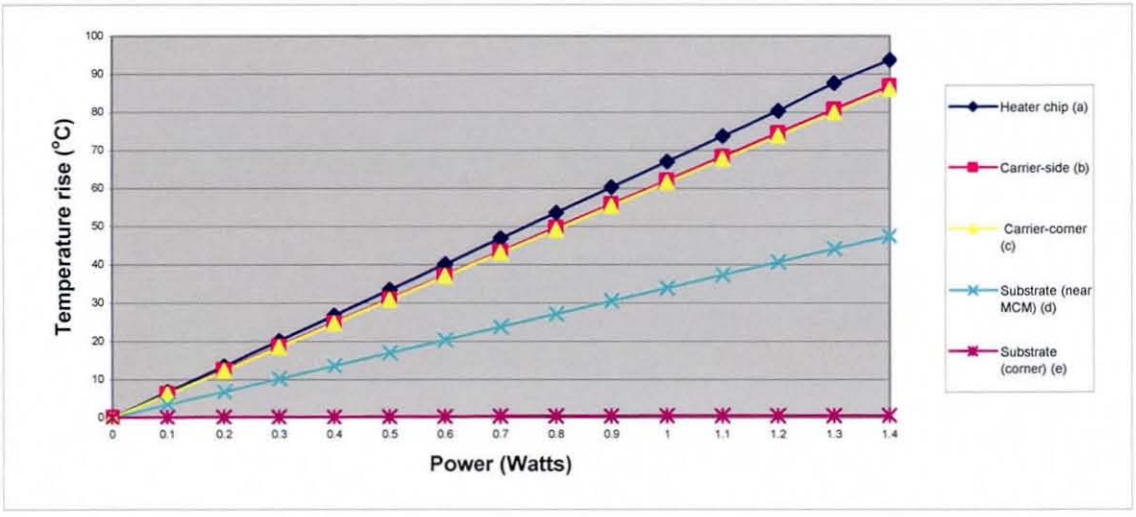


Figure 6.2.1g: *Graph of the power and temperature readings displayed in table 6.2.1*

6.2.2: Copper Substrate

The thermal profiles for the MCM on copper substrate are shown for 0.2 and 1.2W in figure 6.2.2a and b respectively. The temperature distributions are equivalent apart from the overall temperature difference. It can be seen that the temperature of the heater chip is significantly warmer than the carrier chip, as anticipated. Furthermore, while the carrier chip is fairly uniform, a mild temperature gradient can be seen on the heater chip. The interconnection layer is shown in figures 6.2.2c and d. It can be seen how the temperature is warmest in the heater chip and cools through the solder joints and then through the carrier chip; the solder joints show a clear temperature gradient. Figures 6.2.2e & f are concerned with the substrates thermal characteristics; a uniform temperature can be seen across the substrate as anticipated due to the high thermal conductivity of copper.

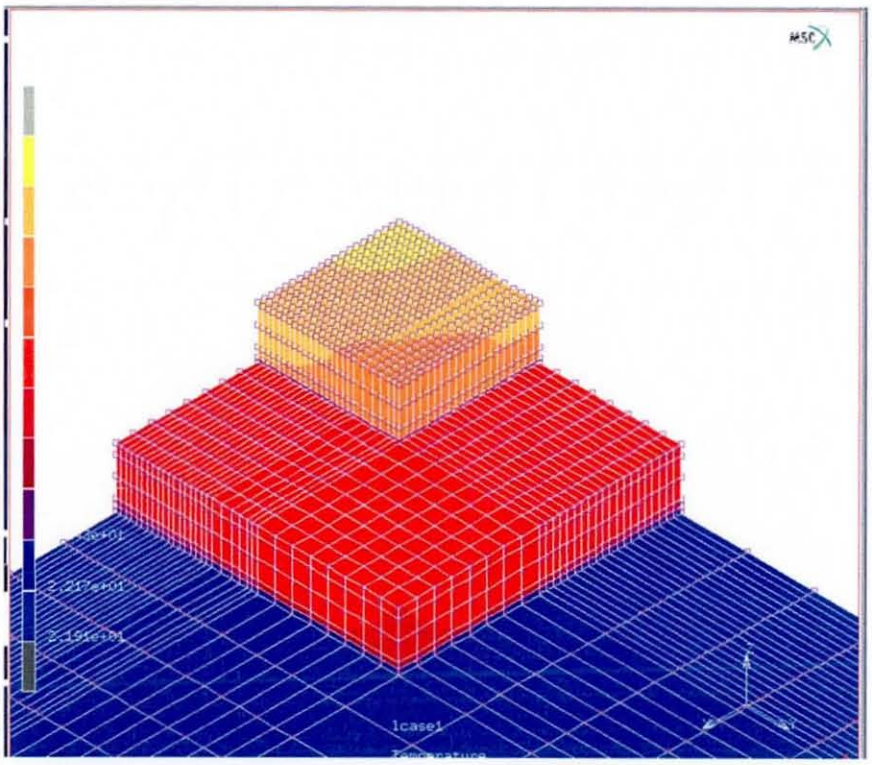


Figure 6.2.2a: *Temperature distribution of FE model of the MCM on copper with 0.2 W powering the chip. The minimum temperature displayed is 21.5°C while the maximum is 24.5°C . The view is isometric*

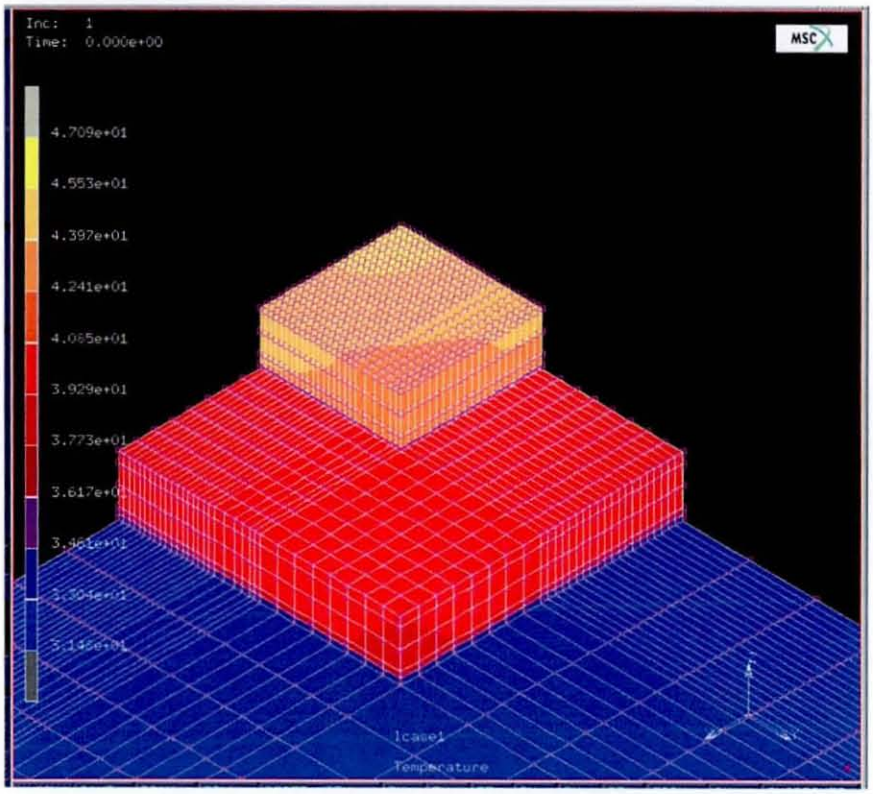


Figure 6.2.2b: *Temperature distribution of FE model of the MCM on copper with 1.2 Watts powering the chip. The maximum temperature displayed is 47.1 °C while the minimum is 31.5 °C. The view is isometric.*

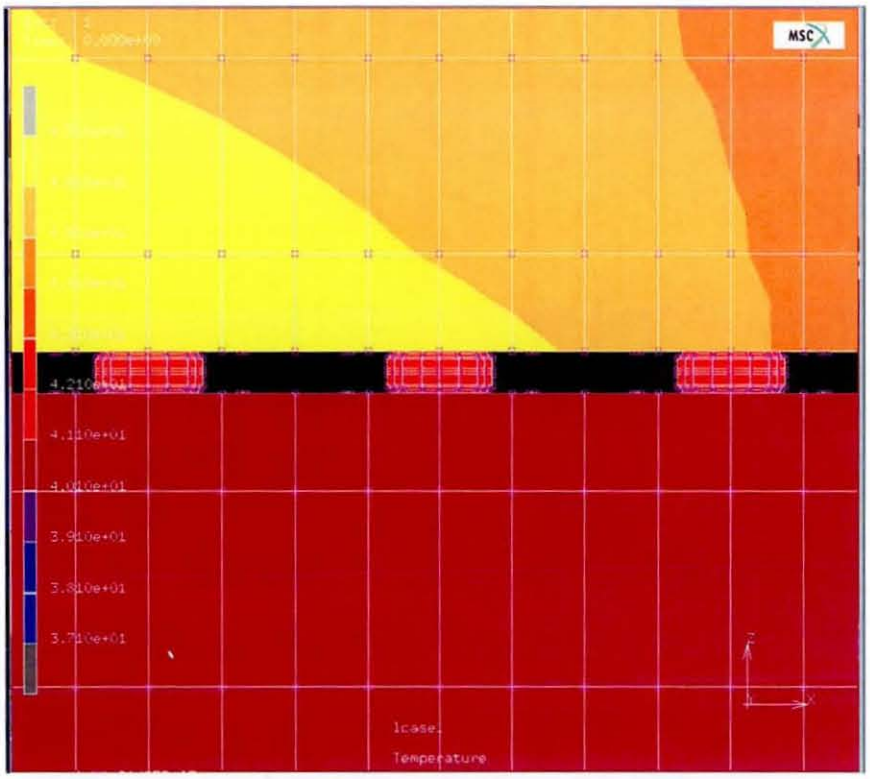


Figure 6.2.2c:
Close up of the interconnection layer. Note the significant temperature difference between the heater and carrier chips. The view is taken from the x axis (origin). The temperature scale is from 47 to 37°C (1.2W)

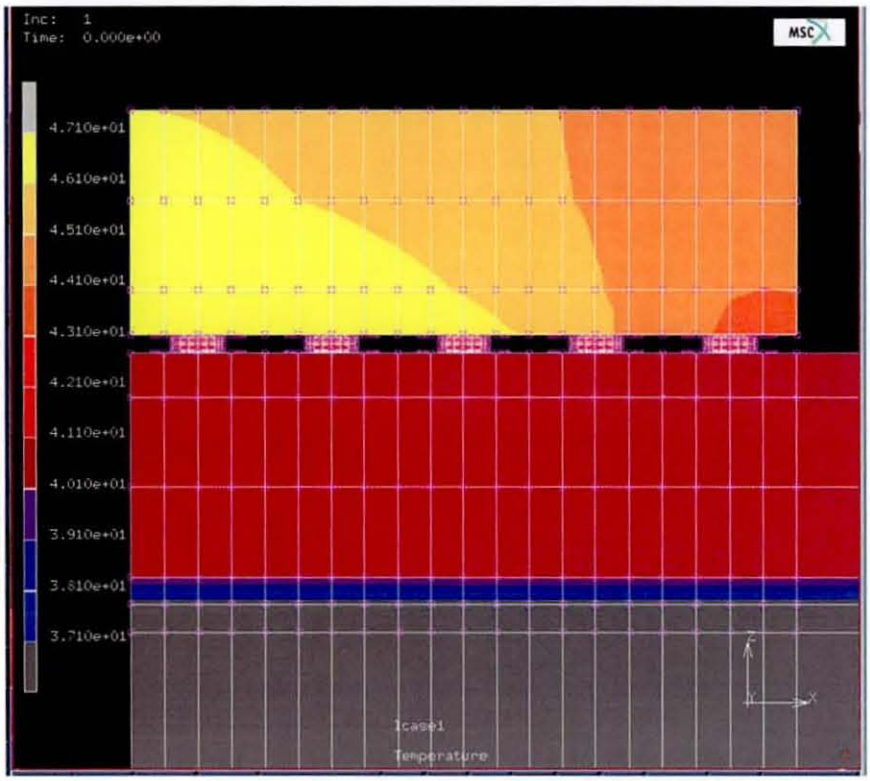


Figure 6.2.2d:
View of the interconnection layer with the heater and part of the carrier chip included. Note the warmer region of the carrier chip to the right hand side by the chip. The temperature scale is from 47 to 37°C (1.2W)

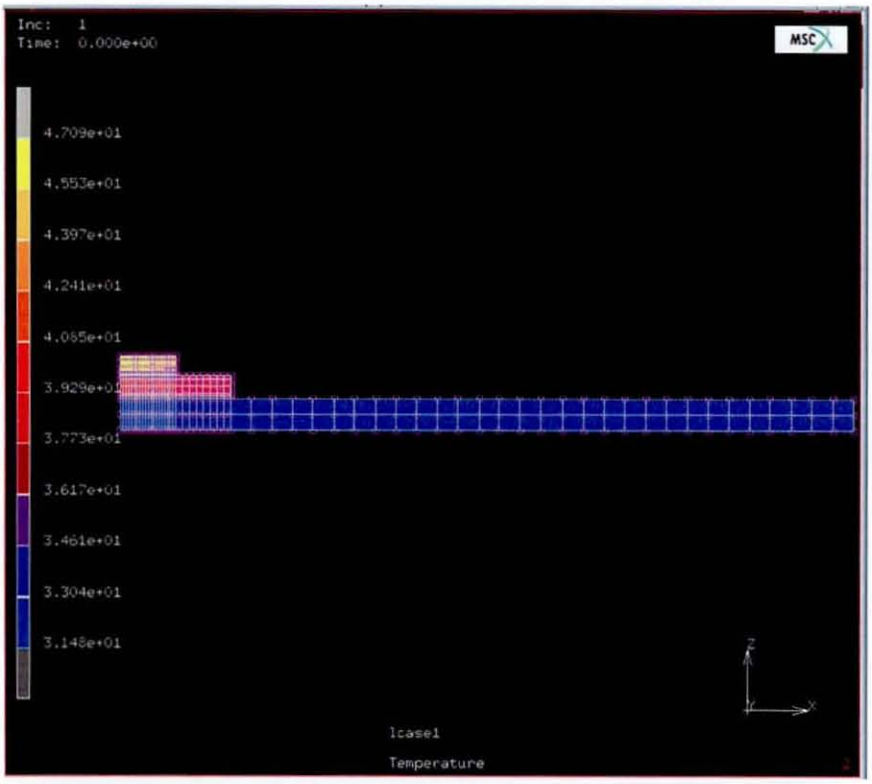


Figure 6.2.2e: *Full view of the side of the assembly. Note the uniform temperature of the substrate. The view is taken from the x-axis and the temperature range is from 47 to 31.5°C (1.2W)*

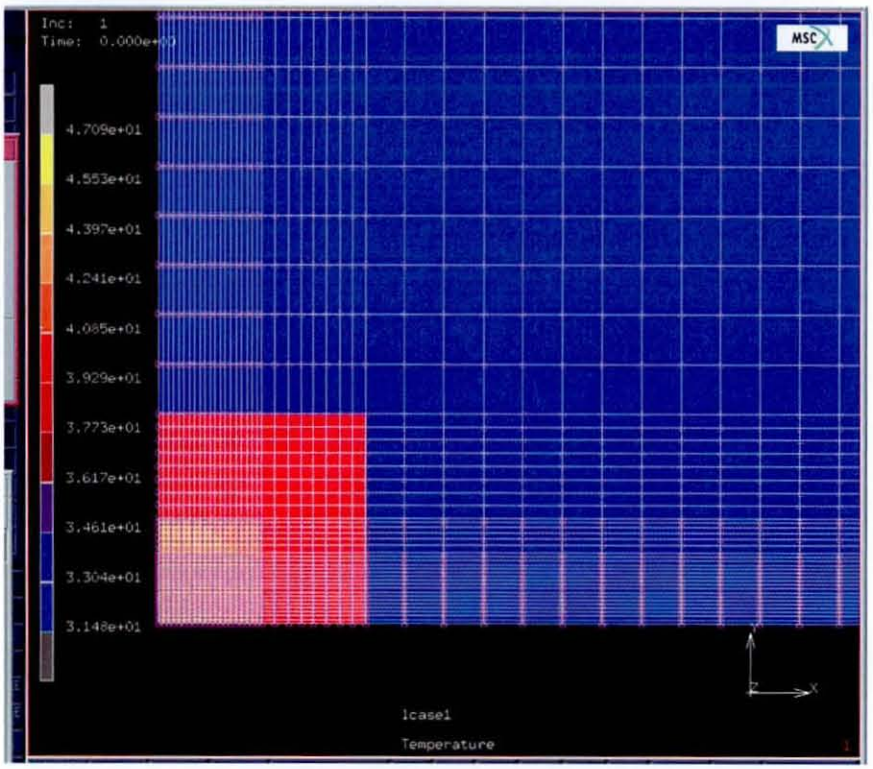


Figure 6.2.2f: *View of the MCM with part of the substrate. Again the substrate is of uniform temperature. The view is a plan view (from above) and the temperature range is 47 to 31.5°C (1.2W)*

Effect of Power Input on Steady State Temperature for the MCM on Copper Substrate

When the substrate material properties were changed to those of copper, the data shown in table 6.2.2a was obtained and the relationship can be seen in graph 6.2.2g. As with the FR4 substrate, the plots for the copper substrate also demonstrated a linear relationship between power and temperature. However, the temperature rise was much lower than for FR4 and an increase of 2.18°C was recorded for each 0.1W power increase in the heater chip at node (a). Furthermore, within the MCM there was a larger temperature gradient between the carrier chip and the heater chip. On the carrier chip at node (c) the temperature increase was 73% of that exhibited by the heater chip.

Equivalent power input (W)	***Heater chip (a)*** (°C)	***Carrier chip- side (b)*** (°C)	***Carrier chip corner (c)*** (°C)	***Substrate (near board) (d)*** (°C)	***Substrate (far corner)(e)*** (°C)
0.1	2.18	1.64	1.61	1.	0.96
0.2	4.35	3.28	3.21	2	1.91
0.3	6.53	4.92	4.82	2.99	2.87
0.4	8.7	6.57	6.43	3.99	3.83
0.5	10.88	8.21	8.04	4.99	4.78
0.6	13.05	9.85	9.64	5.99	5.74
0.7	15.23	11.5	11.25	6.99	6.7
0.8	17.4	13.14	12.86	7.98	7.66
0.9	19.58	14.78	14.46	8.98	8.61
1	21.75	16.42	16.07	9.98	9.57
1.1	23.93	18.07	17.68	10.98	10.53
1.2	26.1	19.71	19.29	11.98	11.48
1.3	28.28	21.35	20.89	12.97	12.44
1.4	30.45	23	22.5	13.97	13.4

Table 6.2.2a:
Table showing the equivalent power input and the resultant temperature rise relative to the ambient at the specified nodes.

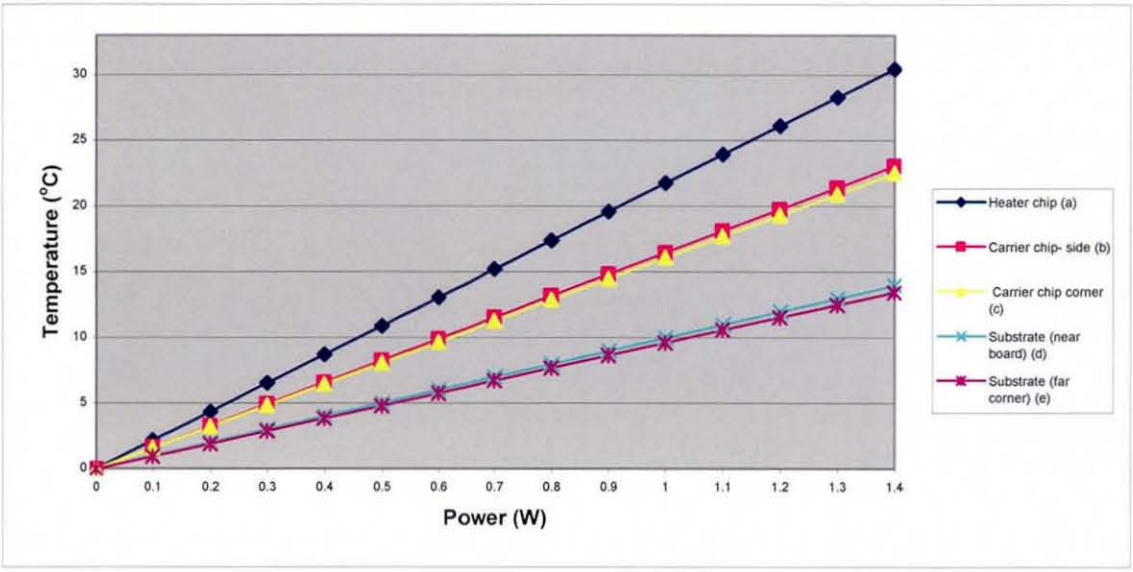


Figure 6.2.2g: *Plots of the temperature readings displayed in table 6.2.2a*

6.3: Transient Finite Element Results

Transient Thermal FE Models

The FE simulations were started from rest and as with the steady state procedure, an ambient temperature of 20°C was used. Three complete power cycles were simulated so that transient thermal profiles could be obtained. The FE model used 300 increments to simulate the three complete cycles totaling 384 seconds; therefore each increment represented 1.28 seconds. There were no adaptive increments and all calculations were made using an explicit solver.

The transient thermal results were displayed at a nodal level so that graphs could be directly compared to the results captured by the IRISYS thermal imager during the experimental power cycling. Nodes (a), (c) and (d) were selected to obtain the transient temperatures of the heater chip, carrier chip and the substrate area near the MCM. Nodes (b) located on the carrier chip and (e) near the far corner on the substrate were omitted from the analysis as: ***i)*** very little difference was seen between node (b) and (c) so the carrier chip was considered to be adequately represented by node (c) on its own and ***ii)*** node (e) showed a negligible temperature rise in the FR4 model and was seen to be very close to that of node (d) for the copper model. For

copper, one trial was performed with node (e) and it was seen to follow the profile of node (d)

For each section, there are transient graphs of power levels 1.4 and 1.2W. 1.2W was chosen as opposed to the lower power levels to enable a direct comparison with the experimental results; the lower power transient results for copper samples presented in section 4.3 were suspected to be contaminated with background "noise". In each section the thermal history from each node is shown as a graph. In addition to the graphs, two tables are presented to show the transient behaviour of the ON and OFF cycles. They show the temperature change noted for each interval from the previous. An interval was defined as 5.12 seconds within the cycle time. The interval started at the beginning of a particular part of a cycle (i.e. the start of either an ON or OFF part) and a total of 6 intervals were used for the purpose of this analysis, therefore the table characterized approximately ½ a cycle. Also a steady state analysis of each scenario (i.e. steady state profile of the MCM powered with 1.4W cooled with 10m/s air flow) was performed such that it could be seen how close each body came to its respective steady state temperature.

6.3.1: FR4 Substrate

Figures 6.3.1a & b capture the transient thermal profile of the MCM on FR4 substrate for both 1.4W and 1.2W power cycle scenarios. It can be seen that the profiles of the graphs are similar. Tables 6.3.1a and b characterize the rates of temperature change for the ON cycle and the OFF cycle respectively. Both graphs exhibit a large initial temperature change which gradually and steadily declines as the cycle time continues and a similar trend can be seen in the off period. Table 6.3.1c shows how close the assembly is to steady state at the end of the cycle while the thermal plots of the MCM at steady state are shown in figures 6.3.1c_i and c_{ii}. It can be seen that the assembly reaches over 95% steady state within the allocated transient cycle time (the MCM is seen to reach 99%). Furthermore, from examining table 6.3.1a it can be seen that the bodies initially heat at different rates at the start of the cycle (the heater chip node enduring the greatest temperature change with the substrate showing the least); though

during the 4th interval (i.e. after 20.48 seconds) the temperature change is the same for the nodes recorded.

The trends noted in the ON cycle were echoed during the OFF cycle: an initial large temperature decrease was noted for all the nodes in the first interval though the rate of temperature change declined during the later increments. Also from the first interval, it can be seen that the nodes started to cool at individual rates, with the heater chip the quickest, but by the 4th interval the entire assembly cools at the same rate.

Taking the first intervals for both the ON and OFF parts of the cycle into isolation, it was noted that the magnitude of temperature change was 8% greater for the OFF cycle than the ON cycle for the MCM. This was considered to be due to the small thermal mass and high thermal conductivity of the MCM compared to those of the substrate, allowing convection and the cooler substrate temperature to have a large effect on the MCM at that specific moment. The subsequent intervals imply a temperature change equivalent to those of the ON cycle.

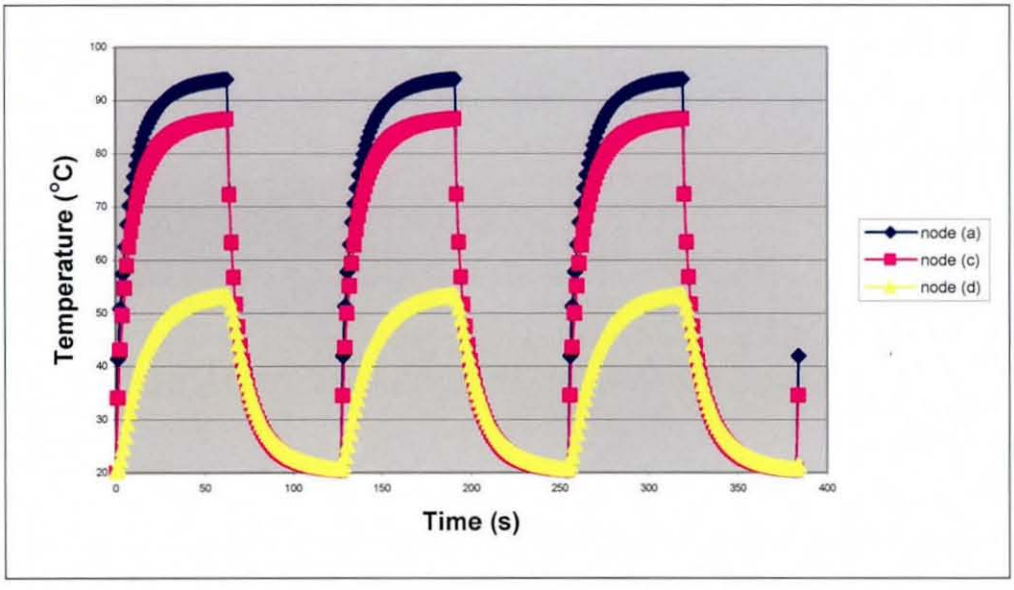


Figure 6.3.1a: *Transient profile of FR4 substrate model, power cycled with 1.4 Watts.*

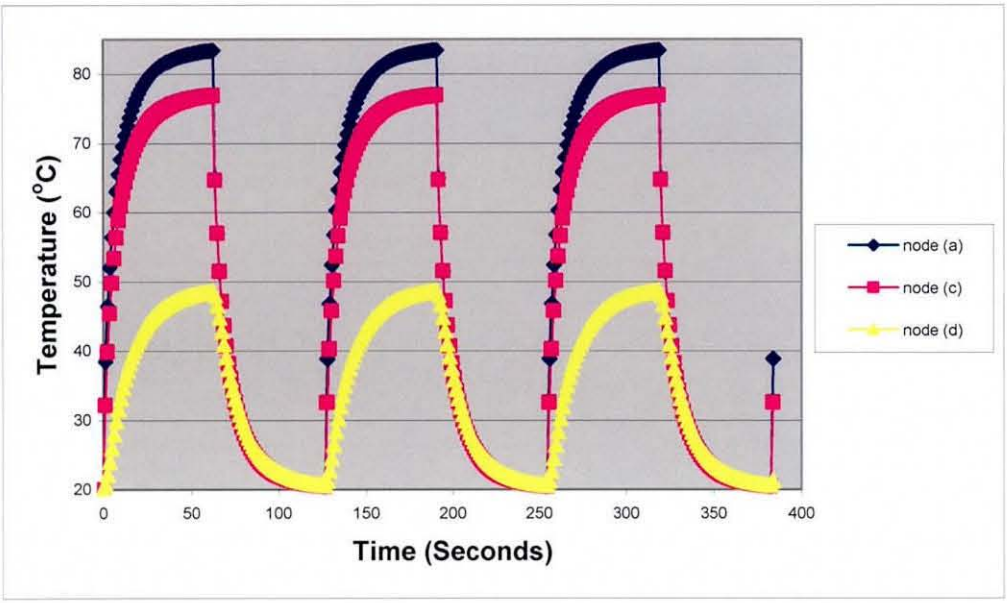


Figure 6.3.1b: *Transient profile of FR4 substrate model power cycled with 1.2Watts*

Node	Interval number (power on) (ΔT °C)					
	1	2	3	4	5	6
(a)	42.6	13.1	7	4.1	2.5	1.6
(c)	34.8	13.2	7	4.1	2.5	1.6
(d)	10	8.4	5.8	4.1	2.5	1.6

Table 6.3.1a: *Temperature change between intervals (interval is 5.12sec) when MCM is power cycled with 1.4 Watts (ON part)*

Node no	Interval number (Power off) (ΔT °C)					
	1	2	3	4	5	6
(a)	-46.6	-11	6	3.6	2.1	1.5
(c)	-38.9	-11	6	3.6	2.1	1.5
(d)	-9	-7.7	5.2	3.5	2.3	1.6

Table 6.3.1b: *Temperature change between intervals (interval is 5.12sec) when MCM is power cycled with 1.4Watts (OFF PART of cycle).*

Node no	Maximum temp in cycle time (ΔT °C)	Steady state (ΔT °C)	% steady state reached
(a)	74.0	74.52	99%
(c)	66.4	66.93	99%
(d)	33.4	34.8	95%

Table 6.3.1c: *Comparison of maximum temperature reached during power cycling and steady state temperature. The steady state models were of 1.4W cooled with airflow at 10m/s*

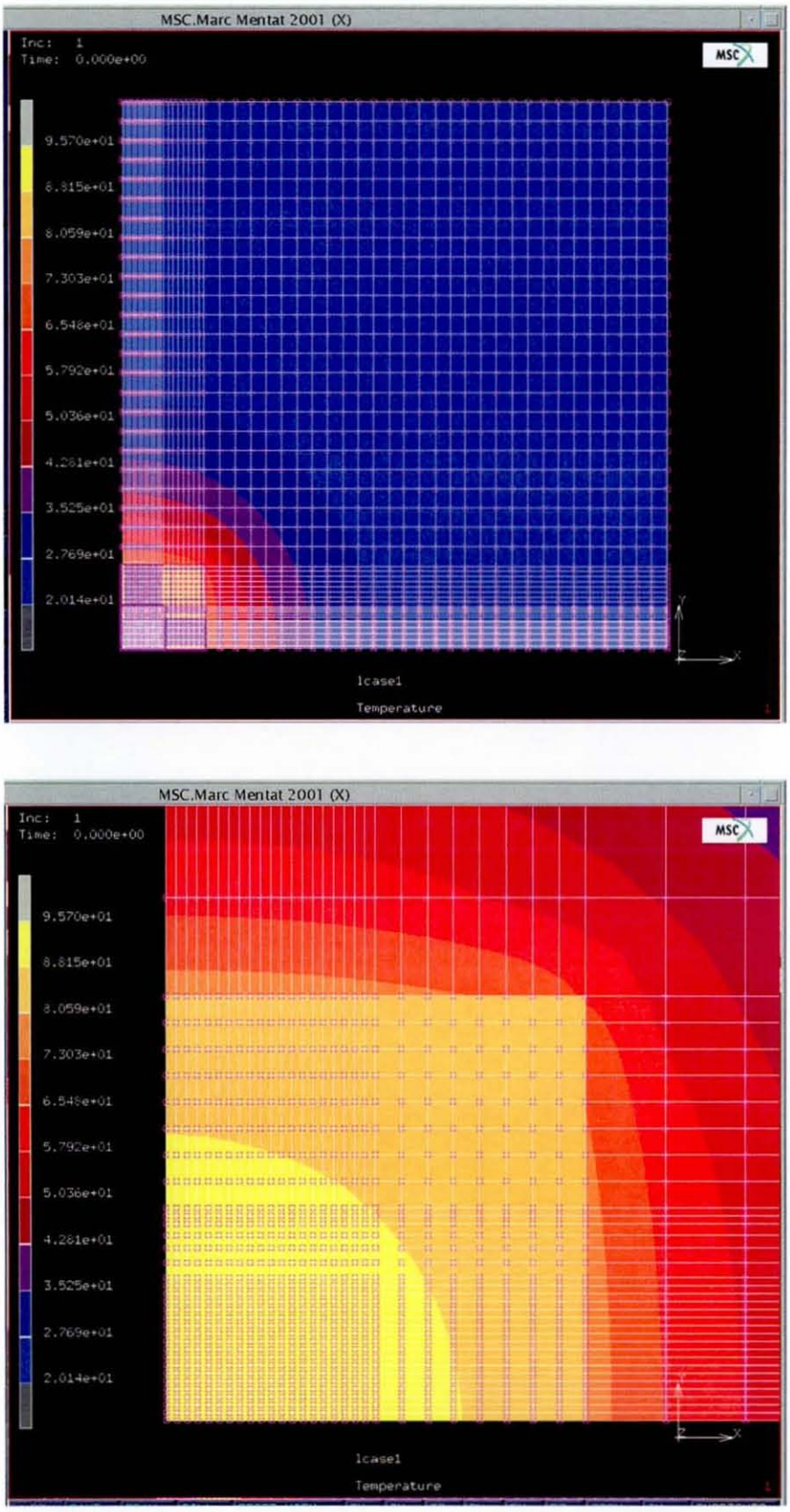


***Figure 6.3.1c_i and c_{ii}**: Thermal profiles of MCM steady state when cycled with 1.4W and airflow at 10m/s (plan view of assembly)*

Figure 6.3.1d shows the temperature difference between the heater chip and the carrier chip as a function of time. During the ON cycle, it can be seen that the temperature difference peaks after one second and remains almost constant afterwards; for the remainder of the cycle the temperature difference reduces by 0.1°C. For the OFF cycle, the temperature difference is negative, implying that the heater chip is actually slightly cooler than the carrier chip. The heater chip is cooler than the carrier chip by 0.1°C also after 1 second, but the difference decreases as the cycle time increases. This could be attributed to the smaller thermal mass and the surface area in direct contact with the free flow convection.

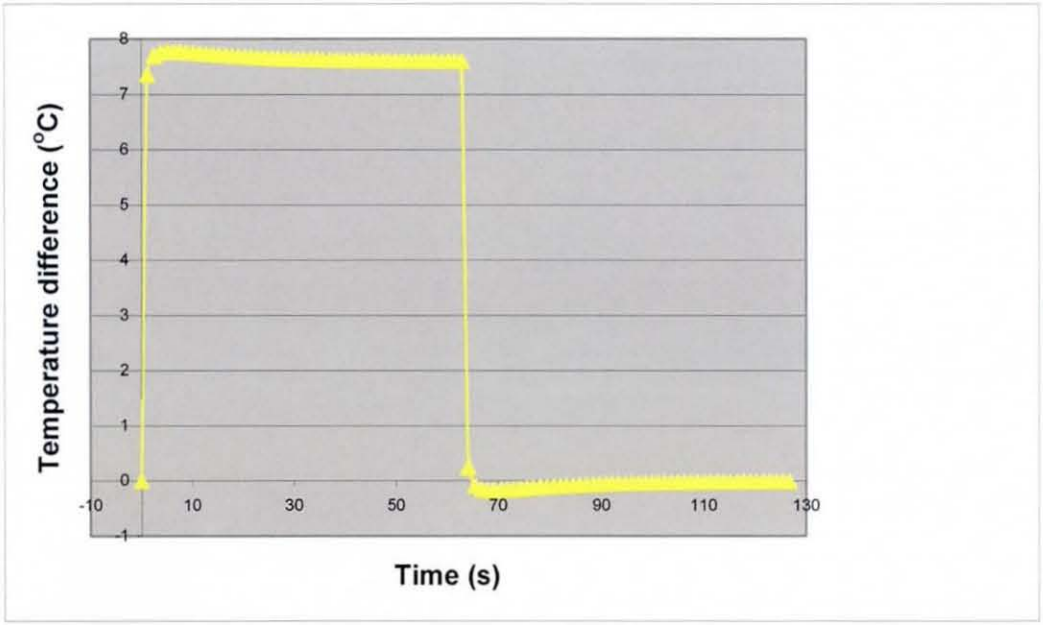


Figure 6.3.1d:*Temperature difference between heater and carrier chip for one cycle.*

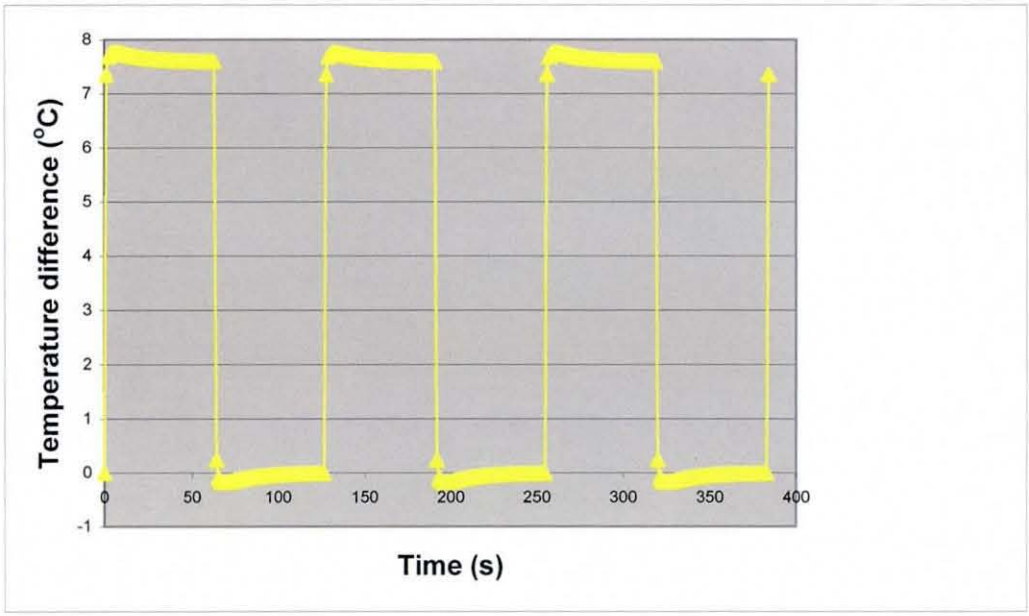


Figure 6.3.1e: *Temperature difference between heater and carrier chip for three cycles.*

6.3.2: Copper Substrate

Figures 6.3.2a & b show the transient thermal profiles for the assembly for when the power levels 1.4W and 1.2W were used respectively. Both graphs exhibit identical characteristics, apart from the final temperatures reached. From both graphs, it is clear that the MCM exhibits a large temperature change during the start of the ON cycle. However, there is an almost instantaneous and sharp change in the subsequent temperature profile. It can be seen that the rate of temperature change was slower throughout the assembly for the remainder of the cycle time. The temperatures of the assembly at different time intervals during the ON and OFF cycles are shown in tables 6.3.2a and b respectively. The bodies within the assembly are shown to initially heat up at different rates (the heater chip being the quickest, the substrate being the coolest) however, the rates of temperature change are observed to be uniform across the assembly after the second interval. The obtained thermal plots are shown in figures 6.3.2c_i and c_{ii} where the copper substrate can be seen at uniform temperature. The comparison with the steady state temperature is shown in table 6.3.2c, where it can be seen that the assembly is significantly far away from steady state at the end of the ON cycle when copper substrate is used.

When the heat flux is removed, the MCM endures a rapid temperature drop until it is the same temperature as the substrate. Once the MCM has reached this temperature, it then continues to cool at a rate governed by the substrate. From table 6.3.2b it can be seen that the assemblies cool in a manner consistent with the heating cycle. According to the graphs, it is also apparent that no part of the assembly returns to room temperature after the power cycling has started.

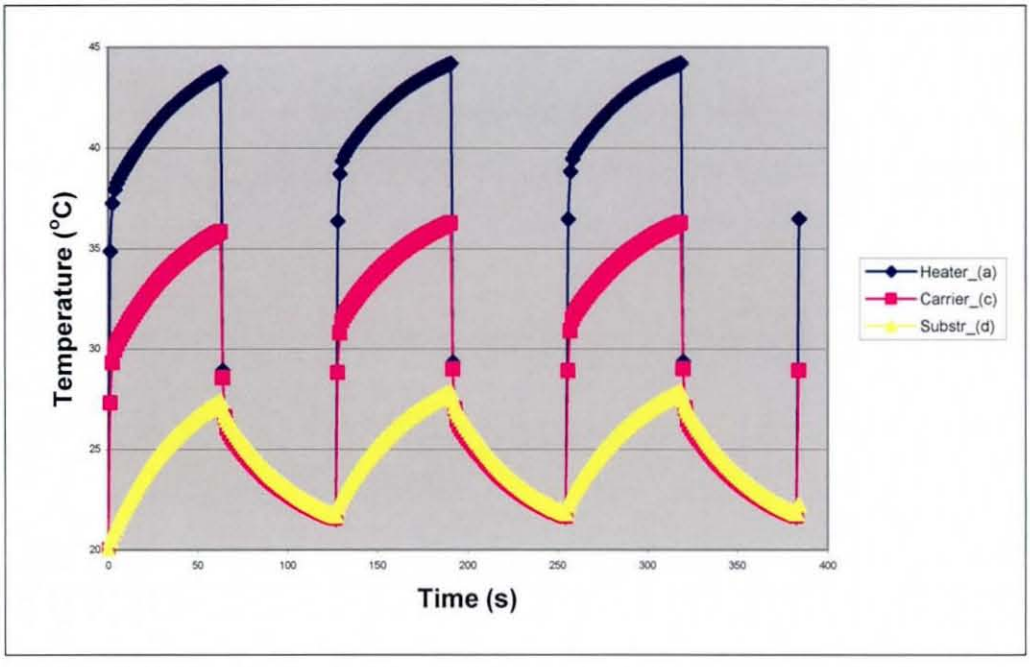


Figure 6.3.2a: *Transient profile of copper substrate model with 1.4W power cycling.*

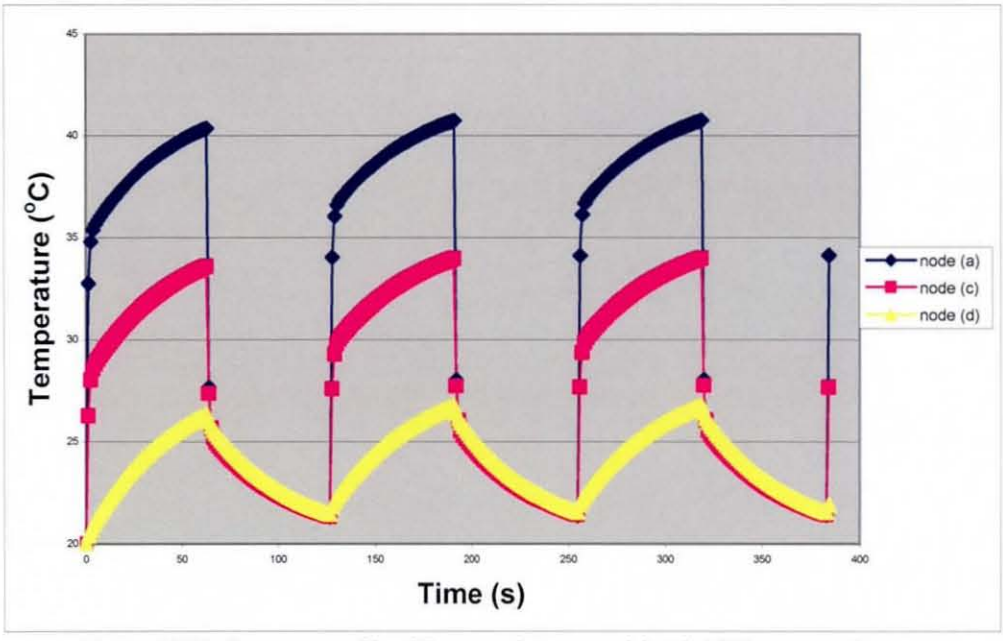


Figure 6.3.2b: *Transient profile of Copper substrate model with 1.2W power cycling.*

Node no	Interval (Power on) (ΔT °C)					
	1 (°C)	2 (°C)	3 (°C)	4 (°C)	5 (°C)	6 (°C)
(a)	18.3	0.8	0.7	0.7	0.6	0.5
(c)	10.3	0.8	0.7	0.7	0.6	0.5
(d)	1.4	0.8	0.7	0.7	0.6	0.5

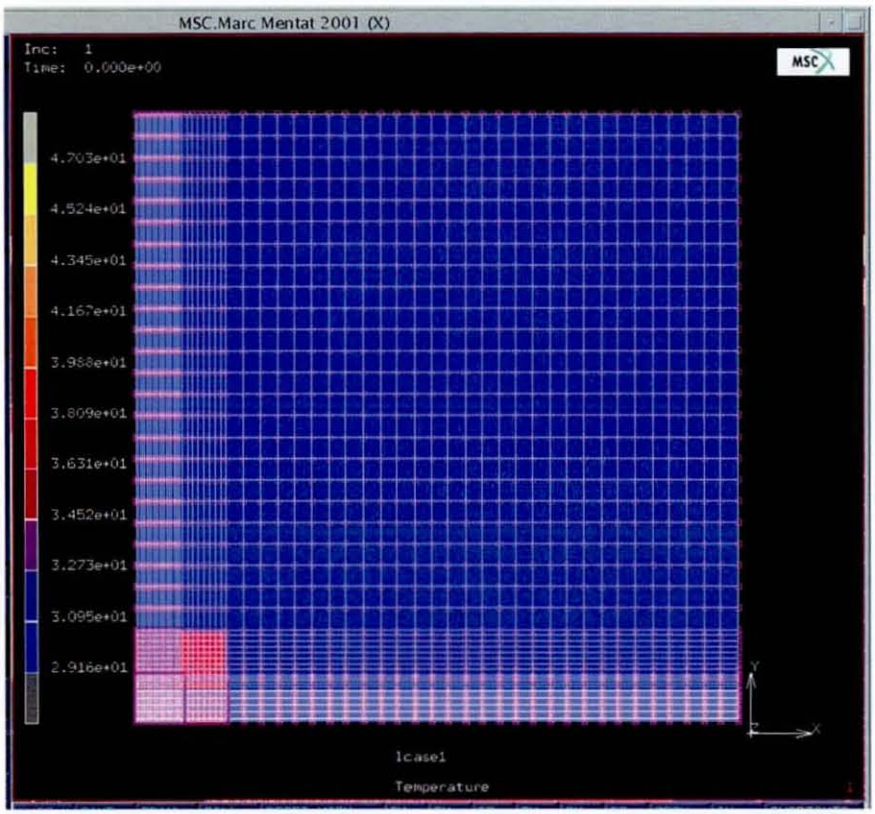
Table 6.3.2a: *Temperature change between intervals (interval= 5.12sec when MCM is power cycled with 1.4W (ON PART).*

Node no	Interval (Power off) (ΔT °C)					
	1 (°C)	2 (°C)	3 (°C)	4 (°C)	5 (°C)	6 (°C)
(a)	-18.6	-0.5	-0.5	-0.4	-0.4	-0.4
(c)	-10.5	-0.5	-0.5	-0.4	-0.4	-0.4
(d)	-1.2	-0.5	-0.5	-0.4	-0.4	-0.4

Table 6.3.2b: *Temperature change between intervals (interval is 5.12sec) when MCM is power cycled with 1.4W (OFF PART).*

Node no	Maximum temp in cycle time (ΔT °C)	Steady state temp (ΔT °C)	% steady state reached
(a)	23.8	25.87	91%
(c)	15.9	17.98	88%
(d)	7.4	9.72	76%

Table 6.3.2c: *Comparison of steady state temperature and final temperature reached in assembly. The steady state model was of the MCM on copper board power cycled with 1.4W and an airflow rate of 10m/s*



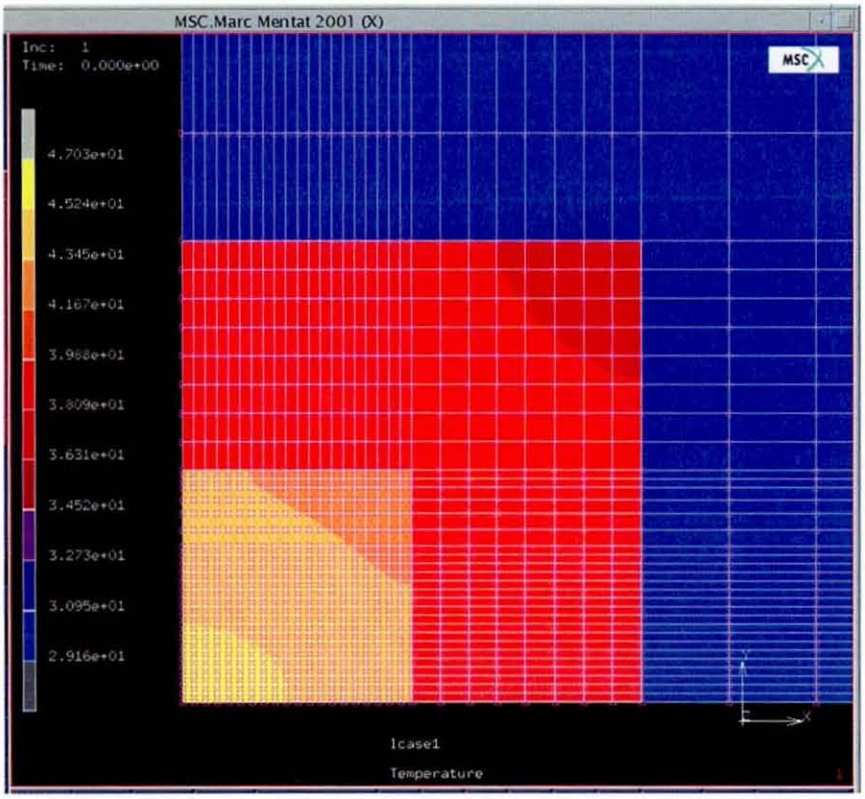


***Figure 6.3.2c$_i$ and c$_{ii}$** : Steady state thermal profile of assembly on copper board powered with 1.4W and airflow at 10m/s(plan view)*

The transient temperature difference between the heater and carrier chips are shown in figure 6.3.2d and e. While the temperature increase is marginally greater than that of FR4 (8°C) the temperature difference remains constant throughout the cycle. Also on the OFF cycle, there is no temperature difference between the heater and carrier chips after the initial rapid cool down.

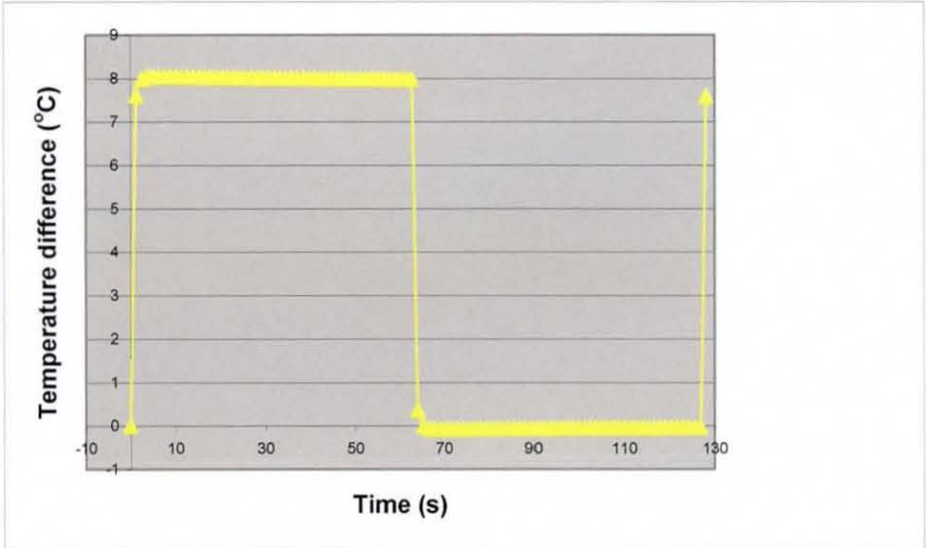


***Figure 6.3.2d:** Temperature difference between heater and carrier chip for the first power cycle*

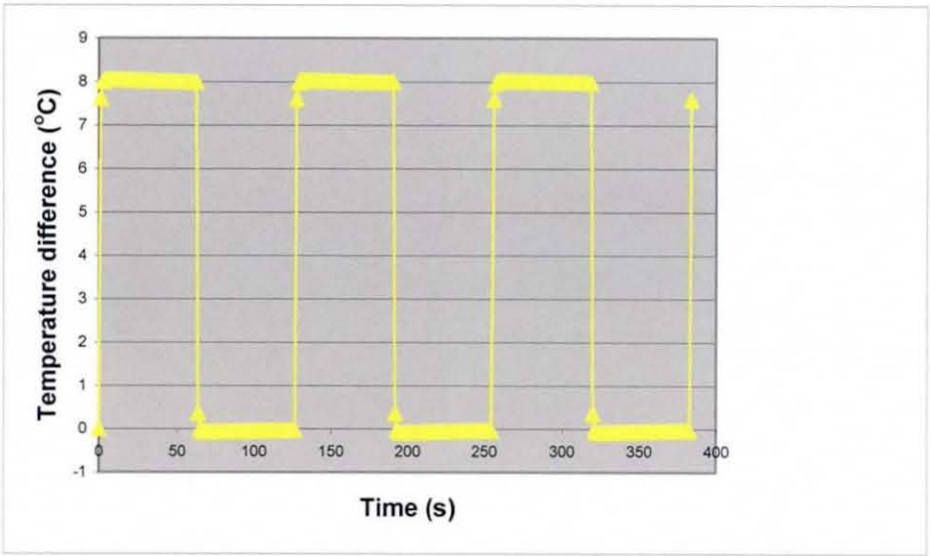


Figure 6.3.2e: *Temperature difference between heater and carrier chip for three cycles.*

6.4: Discussion

6.4.1: Effect of Copper Substrate Compared with FR4

6.4.1.1: Steady State Results

Table 6.4.1.1a compares the FE thermal results for the 1.4W power simulation. It could be seen that the use of the copper substrate had a greater effect in cooling the carrier chip than the heater chip. This was attributed to the greater thermal conductivity and mass of the copper substrate drawing heat away from the carrier chip. There was also a significant reduction in the temperature reached by the substrate (compared with the MCM), however this was very uniform across the copper substrate (node (e) was observed to reach 96% of the temperature rise of node (d)), which was attributed to the high thermal conductivity of copper.

Node	FR4 model (ΔT °C)	Copper model (ΔT °C)	Temperature rise from copper model as a percentage of FR4 model
(a) heater chip	93.8	30.45	32%
(c) carrier chip	86	22.5	27%

Table 6.4.1.1a: *Table comparing the temperature increase exhibited by the bodies on the MCM.*

Node	FR4 model (ΔT °C)	Copper model (ΔT °C)	Temperature rise from copper model as a percentage of FR4 model
(d) close to MCM	47.5	13.97	30%
(e) far from MCM	0.4	13.4	3350%

Table 6..4.1.1b: *Table comparing the temperature increase exhibited by the substrate*

Table 6.4.1.1b displays the final temperature differences recorded from the nodes near the MCM (d) and far away (e). The poor conductivity of the FR4 substrate impeded the thermal pathway of the assembly therefore the primary escape route for the accumulated heat was by convection through the surface of the MCM. As only the substrate area local to the board was heated, there was very little convection heat transfer from the substrate due to the small volume of the substrate actually heated. However when the material properties of copper were used, there was a viable alternative path for heat to escape; the thermal energy was spread over a much greater

area. Furthermore, the convection was applied to a much larger surface area as the entire substrate was heated.

6.4.1.2: Transient Results

The effects of changing the substrate properties from those of FR4 to copper are apparent. Namely the maximum temperature is reduced significantly and the overall profile of the assembly has changed dramatically. For comparison, the transient profile of the FR4 and copper are shown for the nodes located on the heater chip, carrier chip and the substrate in figures 6.4.1.2a, b and c respectively. From the graphs containing the MCM nodes, a smooth curve can be seen on the nodes from the case when the FR4 substrate is used as opposed to the sharp change in gradient for the assembly with the copper substrate. When the FR4 substrate is used, it exhibits a temperature change in character with that of the corresponding MCM, though steady state temperature is not reached. When the copper substrate is used, the substrate does not have the immediate temperature rise as exhibited in the case for FR4.

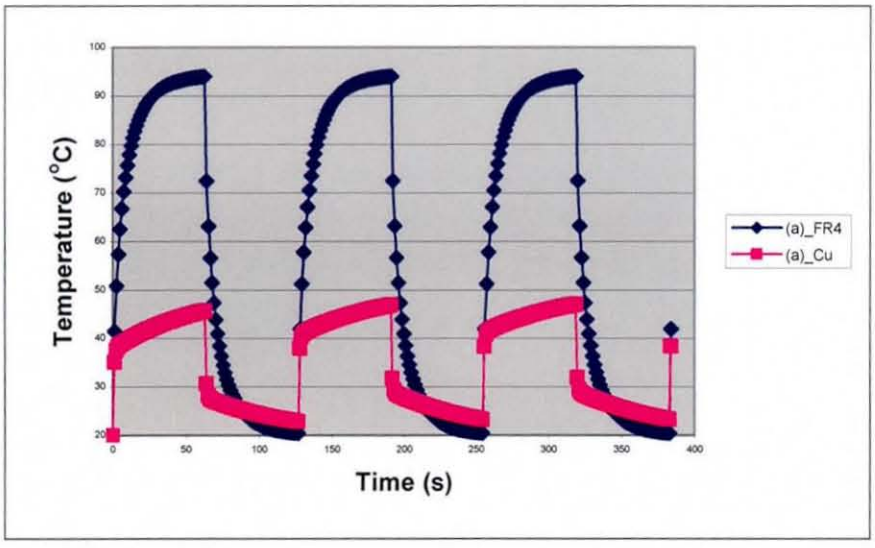


Figure 6.4.1.2a: *Temperature profiles of heater chip on FR4 and copper substrate. The MCM was power cycled with 1.4 Watts*

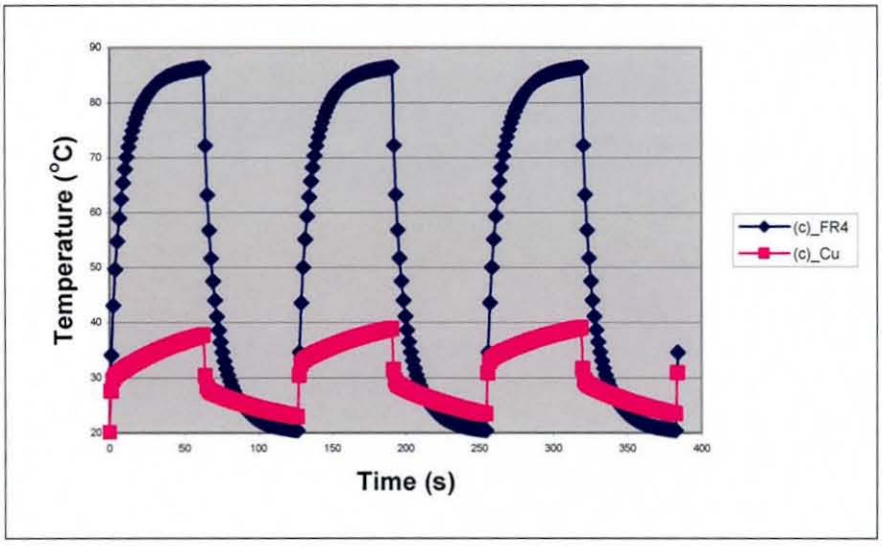


Figure 6.4.1.2b: *Temperature of carrier chip on FR4 and copper substrate (1.4Watts power cycled)*

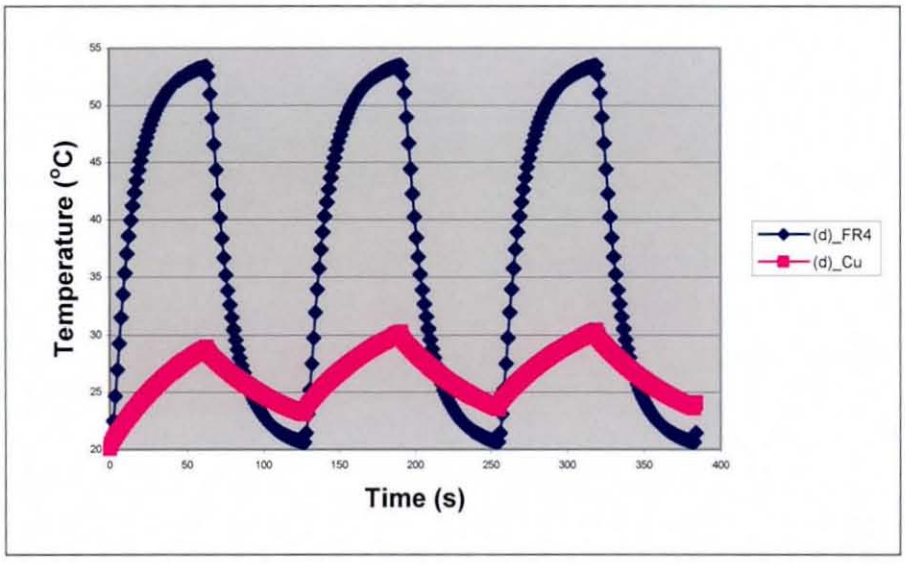


Figure 6.4.1.2c: *Temperature of FR4 and copper substrates (near MCM) when it is power cycled with 1.4 Watts*

As the power/heat source was located in the heater chip, the subsequent heat generated must first pass through the interconnection layer and then through the carrier chip before the relevant substrate material properties can significantly influence the MCM temperature. Therefore it was anticipated that for both cases, the MCMs would exhibit identical thermal profiles until some critical time was reached and then the influences of the substrate would become apparent. This characteristic is implied from the graphs above however to view this in more detail, a separate transient analysis was performed for the assemblies for a time of 1 second with 40 increments used. Figures 6.4.1.2d, e and f show the transient profiles for one second for the heater chip, carrier chip and the substrate. Considering figure 6.4.1.2d first, (heater chip) it can be seen that the temperatures of the nodes for both types of assembly are identical up to approximately 0.2 seconds after the heat flux has been

applied. A very steep thermal gradient is seen on both the assemblies for the first 0.2 seconds before the slope of the curves levels out to those of the nodes on the carrier chip.

On the carrier chip node, the divergence between samples on FR4 and copper was noted after 0.15 seconds. Comparing figures 6.4.1.2d and 6.4.1.2e it can be seen that a marginally sharper divergence exists on the carrier chip than in the case of the heater chip. This was considered to be due to the carrier chip being situated closer to the substrate, and the solder interconnection layer would impede the substrate effects on the heater chip. Therefore the substrate influences on the MCM would be seen sooner on the carrier chip than on the heater chip.

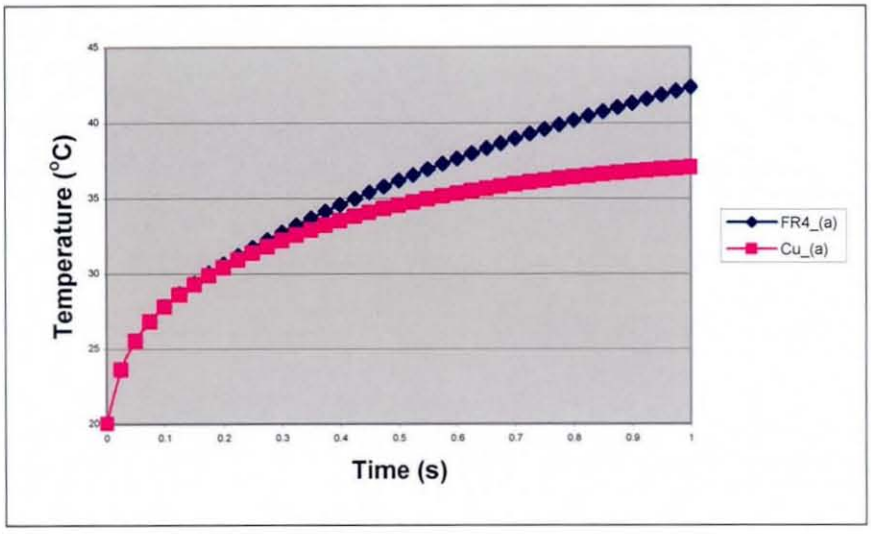


Figure 6.4.1.2d: *Temperature of the heater chip on FR4 or copper substrate power cycled with 1.4W*

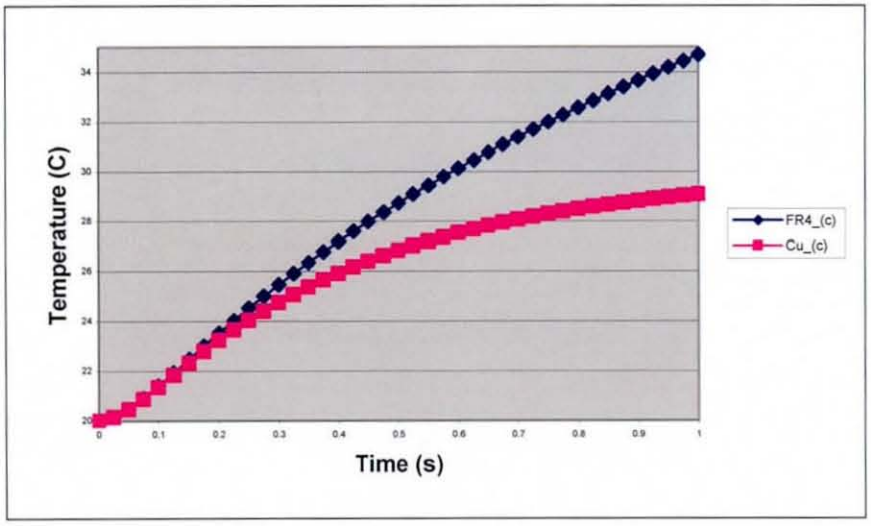


Figure 6.4.1.2e: *Temperature of the carrier chip on FR4 or copper substrate power cycled with 1.4Watts applied for 1 second*

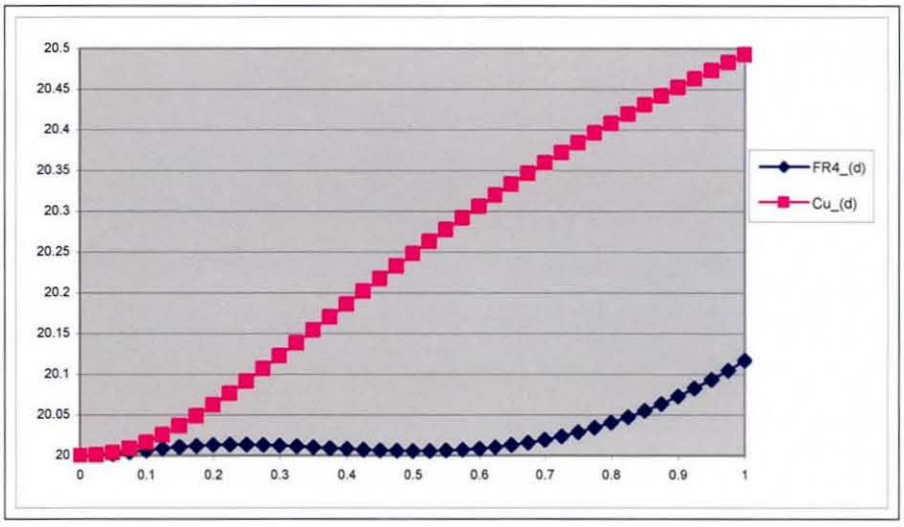
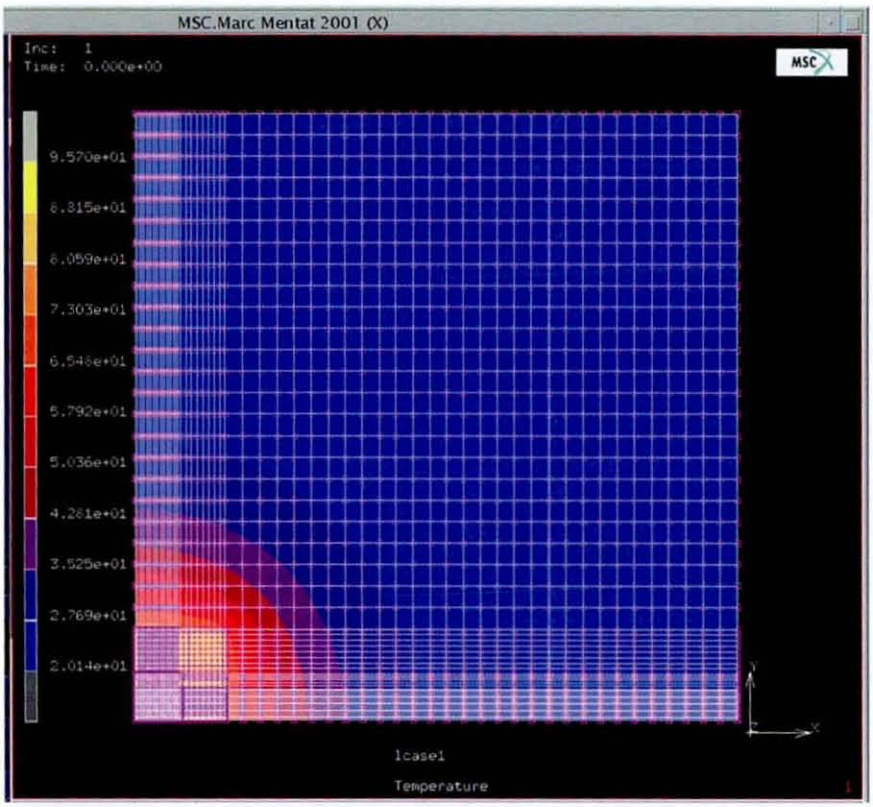


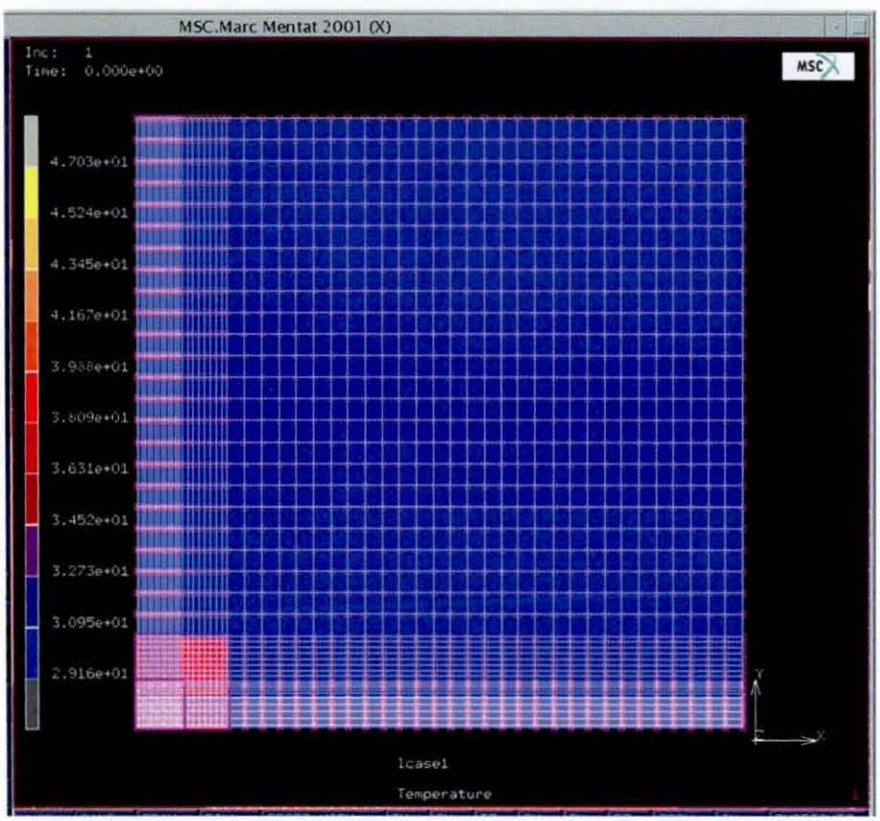
Figure 6.4.1.2f: *Temperature of the FR4 and copper substrates (near MCM) when it is power cycled with 1.4 Watts for 1 second*

ON CYCLE: The slower rates of temperature increase shown for the assemblies of substrates on copper board were anticipated to some extent. Figures 6.4.1.2a shows the final thermal profiles of the assemblies for FR4 and copper. For the FR4 case, it can be seen that only a small volume of the substrate is warmed by the heat flux from the MCM, as opposed to the copper substrate where the heat flux is diffused throughout the substrate. When FR4 is used, the heat flux is effectively applied over a reduced volume therefore a greater temperature rise and peak may be expected.

OFF CYCLE: Though both MCMs cool rapidly until they reach the same temperature of their respective substrates, the assembly characteristics after this point are markedly different. Whereas the FR4 assembly continues to cool off rapidly before levelling off, the copper substrate cools slowly. This was also considered to be due to only part of the FR4 substrate being heated; therefore the accumulated thermal energy is lost through the MCM. However, all of the copper material was heated and the only heat loss mechanism apparent in the substrate was convection, hence the slower cooling rate. It is also noted that the copper substrate assembly never returns to the initial, or ambient temperature; it remains significantly warmer than the ambient surroundings.



g_i FR4



g_{ii} copper

Figure 6.4.1.2g_i and g_{ii} : *Comparison of board temperature distributions for MCM powered with 1.4W and airflow of 10m/s. A temperature gradient can be seen on the FR4 board while the copper board is at uniform temperature.*

6.4.2: Comparison of FE Models with "Real" Thermal Profiling

6.4.2.1: Steady State

FR4 Samples

Maximum temperature: The thermal profiles of the assembly powered at 1.2W with 5m/s airflow from the experimental work and the FE models are shown in figures 6.4.2.1a and b respectively. It can be seen that both profiles identify the heater and carrier chips with distinct and uniform temperatures. The absolute values of the maximum temperatures from the thermal images were questionable for reasons explained in section 4.4.3 so the temperatures from the FE model were compared with the temperature obtained from the 4-point resistance measurements. The FE model predicted a maximum temperature change of 80°C throughout the model while the 4-point resistance measurement indicates that the temperature change was 90°C, therefore the FE model marginally under-estimated the final temperature reached. This may have been due to inconsistencies in the airflow in the experimental work caused by the large opening of the tunnel to allow the image to be taken and difficulties in positioning the airflow meter.

Overall thermal profile and relative temperatures: The FE model persistently predicts that the carrier chip will endure 92% of the temperature change shown on the heater chip, while the thermal profiles predict the temperature change to be between 85 and 90%. Both the models also show the rapid temperature decline in the substrate as the displacement from the MCM is increased, however this region of the FE model is very small compared with that shown on the thermal image. Furthermore it can be seen that the actual temperature of the substrate was significantly warmer than the FE model predicted (by an estimated 10°C). It should be noted that the geometry of the wires from the carrier chip and the substrate, as well as the copper tracks on the substrate were neglected from the FE model as it was anticipated these would unnecessarily complicate the FE model. From the thermal images obtained figure 6.4.2.1a the copper tracks are clearly considerably warmer than the remainder of the substrate; the wires were also at the top of the MCM in the image. Therefore the wires and copper tracks may provide an additional thermal path to the substrate, hence

accounting for the greater temperature difference and the higher substrate temperature shown on the experimental profile.

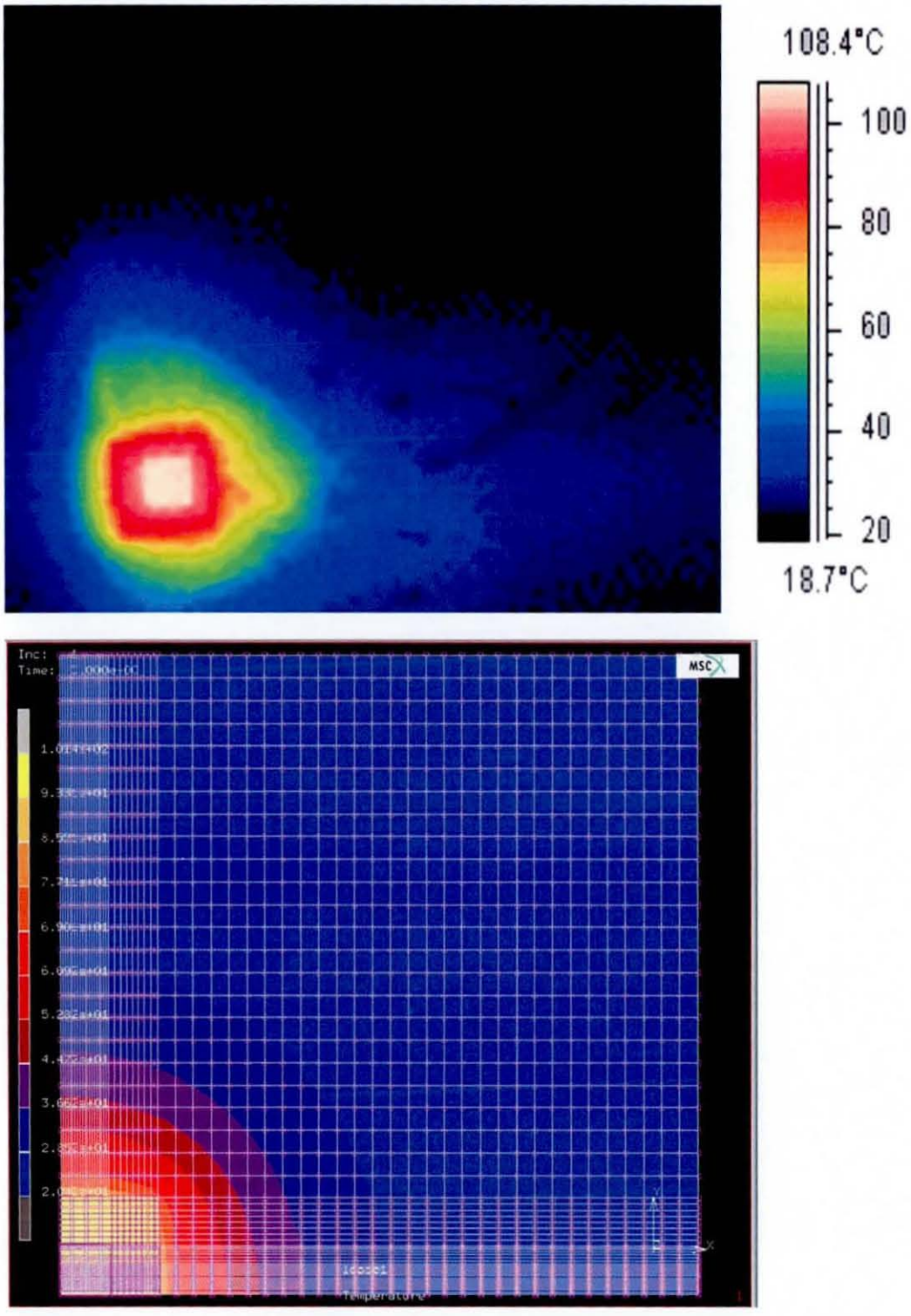


Figure 6.4.2.1a: *Comparison of temperature distributions of "real" assembly and equivalent FE simulation for 1.2 Watts. (FR4 substrate)*

Copper Samples

Maximum temperature: The thermal profiles of the assembly obtained from the experimental work (1.38W) and the FE models (1.4W) are shown in figure 6.4.2.1b. The FE model shows distinct and uniform temperatures on the heater and carrier chip bodies though this cannot be confirmed with the low-resolution camera. It can be seen that the model predicts a temperature change of 30°C while both the camera and the 4-point resistance measurements show the change to be approximately 40°C. The wind tunnel was set to produce an airflow rate of 5m/s with the panel in place. This was difficult to accurately set the airflow speed as the variable fan rate controller was very sensitive in this region, so there may be discretion due to inconsistencies in the obtained airflow rate.

Thermal distribution: Due to the low resolution of the camera, it was not possible to obtain any fine detail of the assembly, though the results from section 4.4.1 imply that the carrier chip endures a temperature change of 75% of that of the heater chip; the FE model predicts that the temperature change is 71%. Both the images clearly show the copper substrate to be of a uniform temperature.

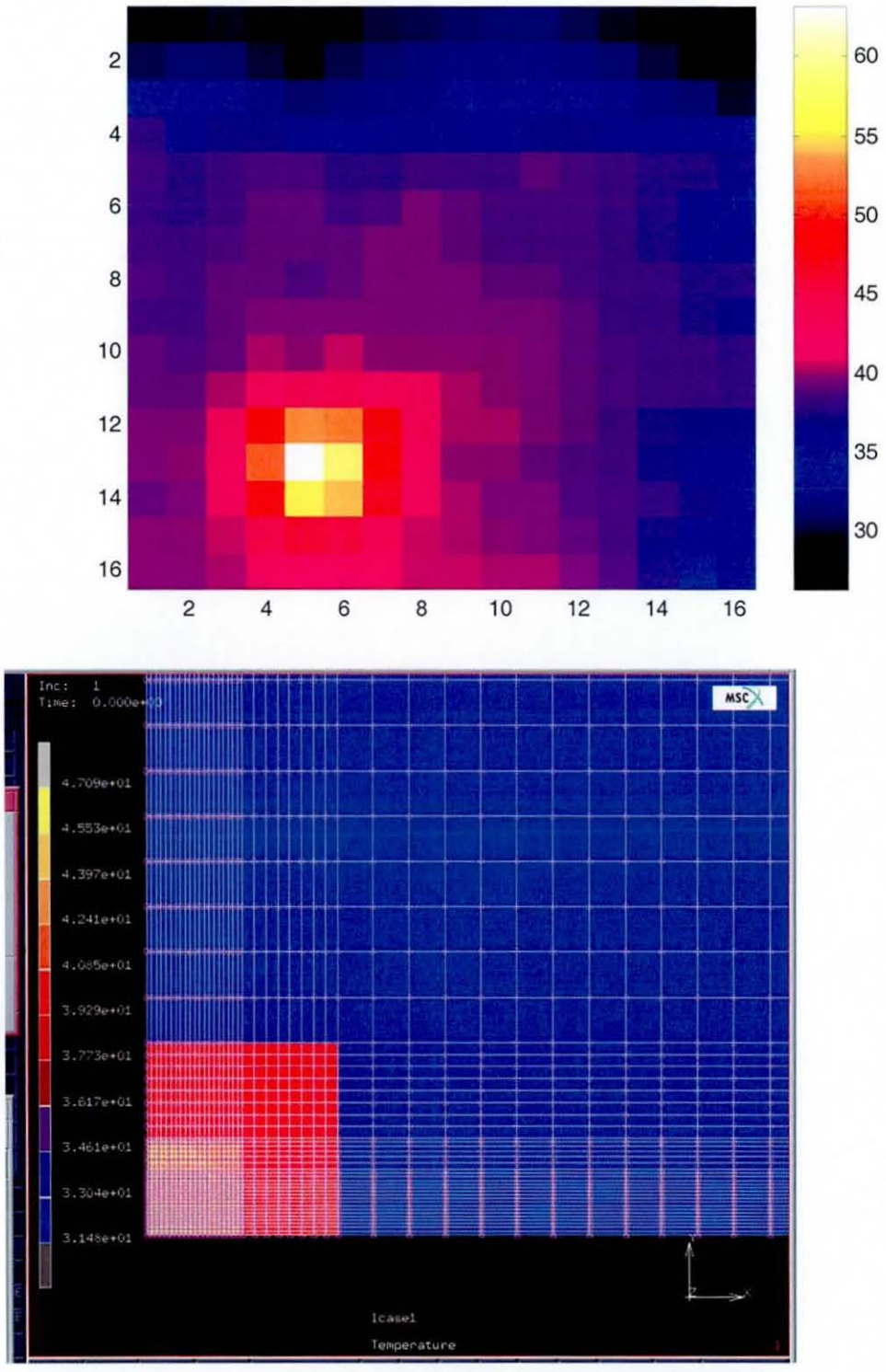


Figure 6.4.2.1b: *Comparison of temperature distributions from experimental work and FE models of MCM on copper substrate powered with 1.4W (real assembly powered with 1.38W)*

6.4.2.2: Transient Analysis

FR4 Samples

Figures 6.4.2.2a and b show the obtained thermal graphs from the experimental work and FE simulations respectively. It can be seen that there is generally good agreement with the temperature changes and the overall profiles of the graphs. However, upon closer examination it is apparent that the experimental graphs have yet to reach steady state whereas the FE results show that steady state is effectively reached at the end of the cycle. The omission of the copper tracks from the FE model as described in the previous section may account for the under-estimation of the substrate temperature. In addition, it may also serve as an explanation for the failure for the assembly to reach steady state, as the effects of the heat flux may be apparent on the copper tracks, as a direct thermal path exists.

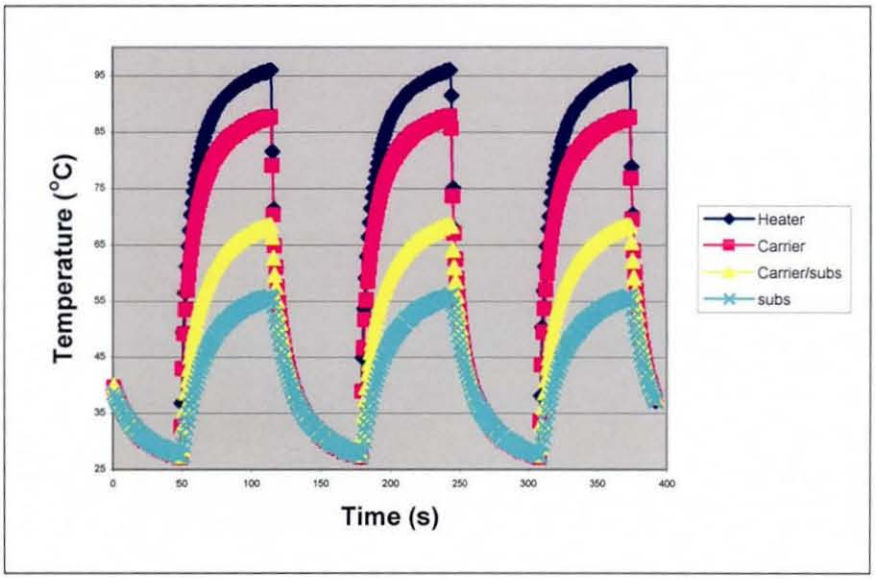


Figure 6.4.2.2a: *"Real" thermal profile of MCM on FR4 board. Power cycled with 1.4 Watts. The carrier/substrate legend represents a potential overlap where a pixel averages the temperature of the carrier chip and the substrate. The substrate legend is an area of FR4 board.*

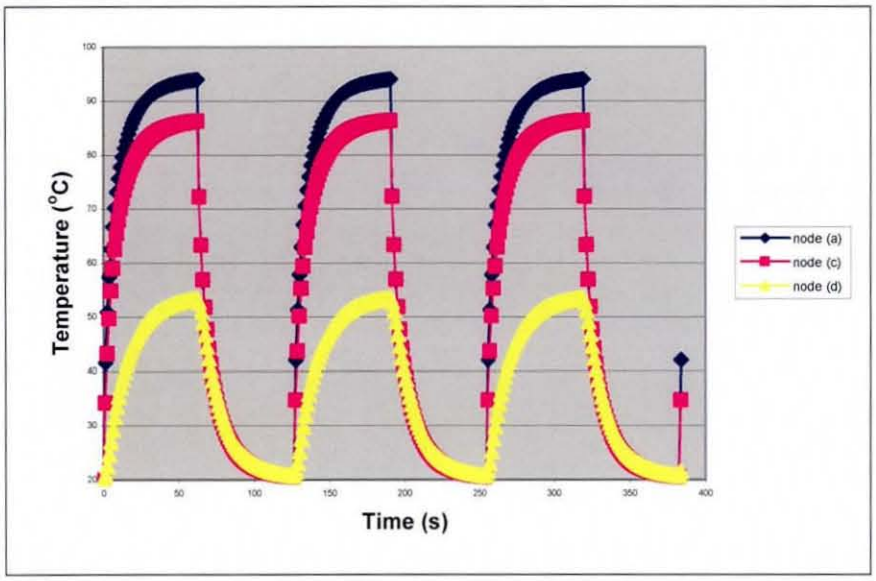


Figure 6.4.2.2b: *Obtained profile from FE model. 1.4 Watts simulated. Node (d) was anticipated to correspond with the legend "subs" from figure 6.4.2.2a*

Copper Substrate Samples

The experimental and the FE profiles of the assemblies with copper substrates are shown in figures 6.4.2.2c and d, it can be seen that the FE model over estimates the temperatures at the end of the cycle. The first cycle of the FE results exhibits a greater change in temperature than the subsequent cycles, this is due to the assembly stabilising. Comparing the subsequent cycles, it can be seen that the general characteristics are in good agreement. However, it can be seen that the FE model over estimates the temperature change of the copper substrate; a 5-degree temperature fluctuation is predicted by the model while the results from the actual cycle shows that this amplitude is 4 degrees. This over estimation was attributed to the attachment of the MCM to the substrate. The adhesive pad used required a uniform pressure to secure bonding. However the MCM was applied by using tweezers in the manner described in section 3.5. This may account for reduced heat transfer to the substrate.

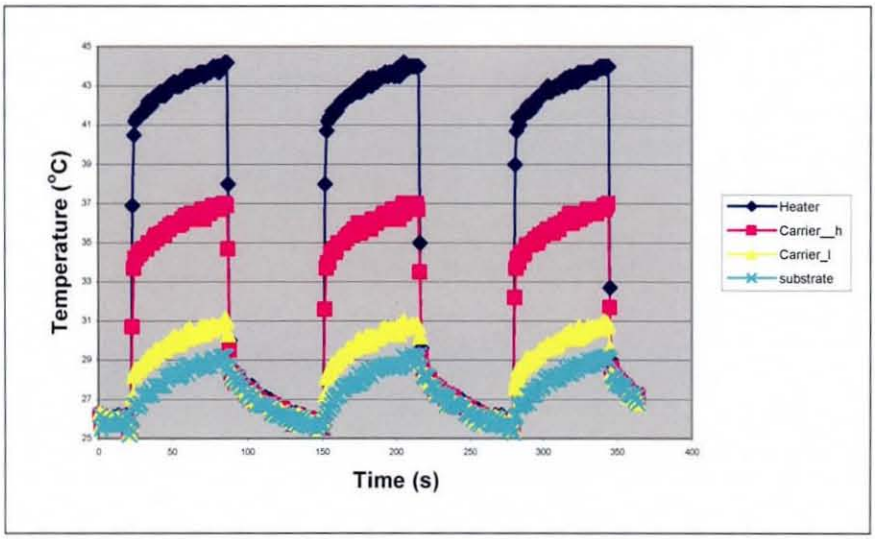


Figure 6.4.2.2c: *"Real" thermal profile of MCM mounted on copper substrate. Power cycled with 1.4 Watts. The carrier_l legend is anticipated to be some overlap between the carrier chip and the substrate while the substrate legend is thought to confidently represent the temperature of the copper.*

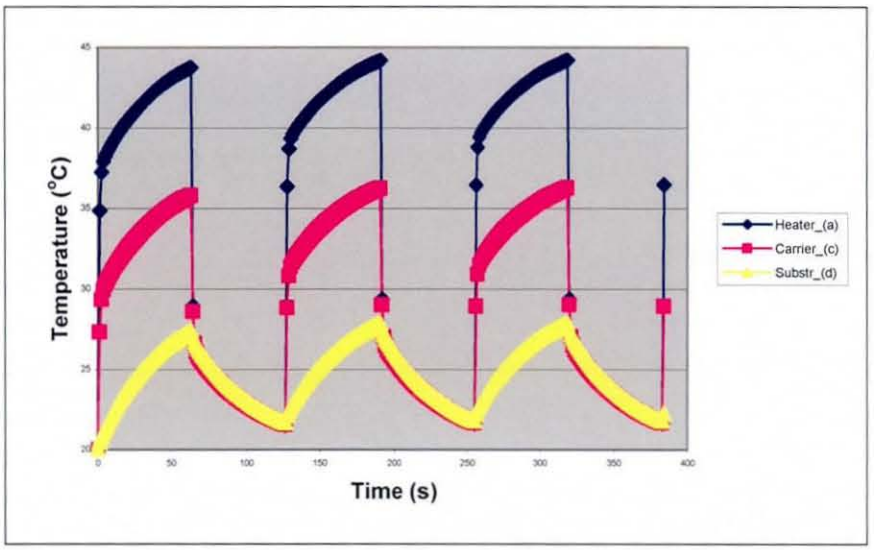


Figure 6.4.2.2d: *Obtained thermal profile from FE model power cycled with 1.4 Watts*

6.4.3: Comparison with Other Studies

The detailed data capture of the FE models show the rates at which the individual bodies heat up and cool down. It can be seen that the heater chip does indeed endure a more instantaneous, rapid and larger temperature change than the carrier chip in agreement with the work of Trigg and Corless (1).

Previous published attempts at capturing a FE power cycle of a flip chip assembly were dependent on test vehicles being manufactured and thermally profiled to obtain the heat transfer characteristics. Some models used the recorded temperatures of the chip and substrate at any given time as the necessary boundary conditions (2), others used a natural convection heat transfer coefficient based on the temperature difference observed between the die and cooling air (3,4). Either way, the models are strongly dependent on a test piece previously being manufactured and profiled. The FE models created in this experiment eliminate this prerequisite as the convection heat transfer coefficient was obtained independently of an observed temperature change. Furthermore, the FE models that used a convection heat transfer coefficient failed to correctly capture the transient profiles (3,4). This was attributed to the temperature dependent heat transfer coefficient used. The models concerned based their heat transfer only on the maximum temperature observed from the assembly while in reality, the heat transfer rate is likely to be changing throughout the transient period and does not become constant until steady state is reached. As the heat transfer coefficients used in this example were not temperature dependent this allowed the transient behaviour to be a better representation of the true profiles exhibited by the assemblies.

6.4.4: Direct Chip Attach

The work presented so far has shown that the FE models can predict the overall thermal distribution for the MCM given its specific characteristics. How does the relationship between the FE model prediction and trial results compare for the case of a die mounted directly on board, as in the case for direct chip attach (DCA)? Lenkkeri & Jaakola (5) power cycled dies attached to either alumina or low

temperature co-fired ceramic (LTCC) substrates with varied power levels and standoff heights as summarised in table 6.4.4a. It was found that there was a relationship between the power levels and cycles to failure for a given substrate and (where applicable) standoff height. It was decided to create FE models of the test vehicles used from this DCA power cycling work in order to provide further validation of the modelling method. The models were created with the following assumptions:

- Assemblies were cooled by natural convection
- The temperature of the die was fairly uniform (i.e. differed by no more than one degree)
- The ambient temperature was 20°C
- Convection rates were constant throughout the cycle time (the maximum value)
- Non-symmetrical entities were ignored such that the model could be simplified to ¼ of the actual geometry (i.e. there were 21 bumps on the "real" assembly. In the model, it was assumed to be only 20 therefore with this simplification, only 5 bumps need be created)

The die used was a PST2 test-chip with 21 pads (though this was simplified to 20 to reduce the size of the model). The dimensions of the die were 3.8x3.8x0.63mm with pad diameters of 160μm and the varying pitches are shown in figure 6.4.4a along with the symmetry lines. The alumina substrate was 16.5x19x0.89mm and had rectangular pads at either 200x100μm or 200x150μm depending on the final standoff height, while the LTCC substrate measured 25x25x0.78mm and had circular pads with a diameter of 200μm; both boards had their pads to match the pitch of the pattern on the die. The models were created in accordance with the meshing procedure described in section 5.5.1 though the model only consisted of the die, the solder joints and the substrate, with all other parts omitted. The exclusion of one of the joints from the chip (i.e. considering 20 joints instead of 21) allowed the model to be simplified to ¼ such that the number of elements required was reduced, in addition the mechanical boundary conditions described in section 5.5.2 were directly implemented. The power simulation was implemented by simulating a heat flux over the required area. The convection heat transfer coefficients were obtained by using the ΔT values obtained

for each scenario from the experimental work and implementing them in the relevant equations for natural convection from section 5.3.3. (i.e. first obtaining the Grashof/Rayleigh numbers, then the Nusselt numbers and finally the convection coefficient). In section 5.3.3, the orientation of the surface was shown to substantially influence the convection rate therefore it was necessary for the die's horizontal surfaces to have a different coefficient from vertical surfaces; therefore each model utilised three different h values as shown in figure 6.4.4b.

It is known that the natural convection heat transfer coefficient is dependent on the temperature difference between the surface of the body and the ambient temperature. Only the final temperatures of the die were reported (i.e. no transient data was available) therefore the heat transfer coefficients were obtained based purely on the final temperature. Consequently good agreement with the final temperatures was anticipated, however the prediction of accurate transient profiles was less certain.

Substrate	Power input (W)	Substrate pad dimensions (μm)	Bump height (μm)
Alumina	5.29	200x150	70
Alumina	5.35	200x100	80
Alumina	6.06	200x150	70
Alumina	6.14	200x100	80
LTCC	3.39	200	80
LTCC	4.03	200	80
LTCC	4.18	200	80

Table 6.4.4ai: *list of each model parameters*

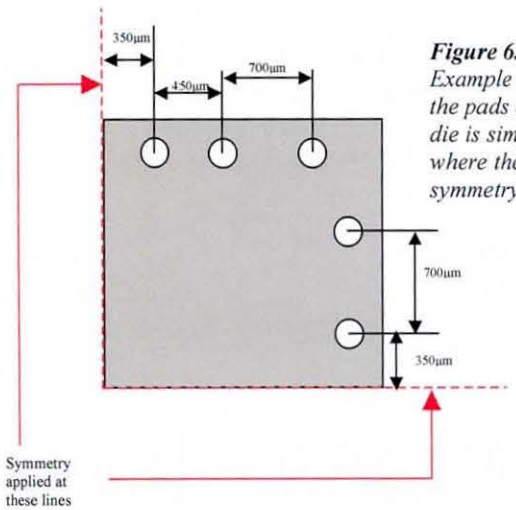


Figure 6.4.4a: *Example of the pitches of the pads on the die; the die is simplified to ¼ where the lines of symmetry are shown.*

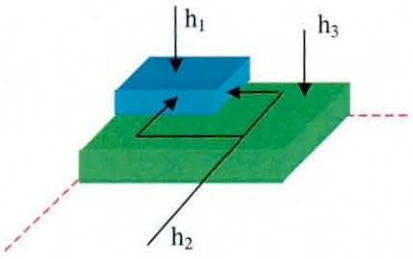


Figure 6.4.4b: *illustration of ¼ FE model and how convection coefficients were applied separately to the horizontal and vertical surfaces of the die. The substrates were anticipated to show much smaller temperature changes due to the large thermal mass, and (comparatively) low conductivity, therefore the substrate had only one convection value applied.*

Substrate Input power (W)	h_1 (W/m^2)	h_2 (W/m^2)	h_3 (W/m^2)
Alumina (5.29)	782.946	1043.63	173.205
Alumina (5.35)	784.955	1046.23	173.205
Alumina (6.06)	808.449	1076.7	173.205
Alumina (6.14)	810.68	1079.58	173.205
LTCC (3.39)	799.68	1057.5	172.673
LTCC (4.03)	820.646	1092.5	172.673
LTCC (4.18)	872.38	1101.25	172.673

***Table 6.4.4aii**: List of natural convection coefficients used.*

Thermal Results

Table 6.4.4b shows the temperature changes exhibited by the models compared with those from the recorded experiments. For the alumina substrates it can be seen that for some cases the FE model predicts very closely the temperature change for the experimental work (5.29W and 6.07W), while marginally overestimating the temperature change (5.35W and 6.14W). These discrepancies were very likely due to the approximations of the model. The FE model predicts more consistent results for the case with LTCC substrates, however on the model it over estimates the temperature reached by 11°C in all cases. The over estimates were possibly due to the omission of the cooling plate that the assemblies were placed on in the experiments. Furthermore, it was clearly evident that less power was needed for an equivalent temperature change when LTCC was used than with alumina.

The transient results from the models are characterised in figure 6.4.4b. It can be seen that the temperature profiles are different depending on whether the substrate was on Alumina or LTCC. The substrates on LTCC appeared to reach their maximum temperature very quickly. After only 10 seconds the thermal profile of the LTCC assemblies exhibit steady state characteristics. The profile of the chip when alumina was used implies that the assemblies are still in transit at the end of the ON cycle. This slower time to reach steady state was attributed to the higher conductivity of the alumina substrate compared to LTCC. This behaviour is in good agreement with the results of the thermal profiles of the MCM, where the more thermally conductive

substrate acts as a heat spreader (albeit to a lesser extent) therefore requiring more time to reach steady state.

Substrate	**Power (W)**	**ΔT (FE model)**	**ΔT (Experiment)** **(Lenkkeri and Jaakola)**
Alumina	5.29	94.5	92.1
Alumina	5.35	100.3	93.2
Alumina	6.06	107.3	107
Alumina	6.14	113.8	108.4
LTCC	3.39	108.6	98.1
LTCC	4.03	126.6	114.9
LTCC	4.18	130.6	119.5

Table 6.4.4b: *Comparison of the temperatures reached from a given power level*

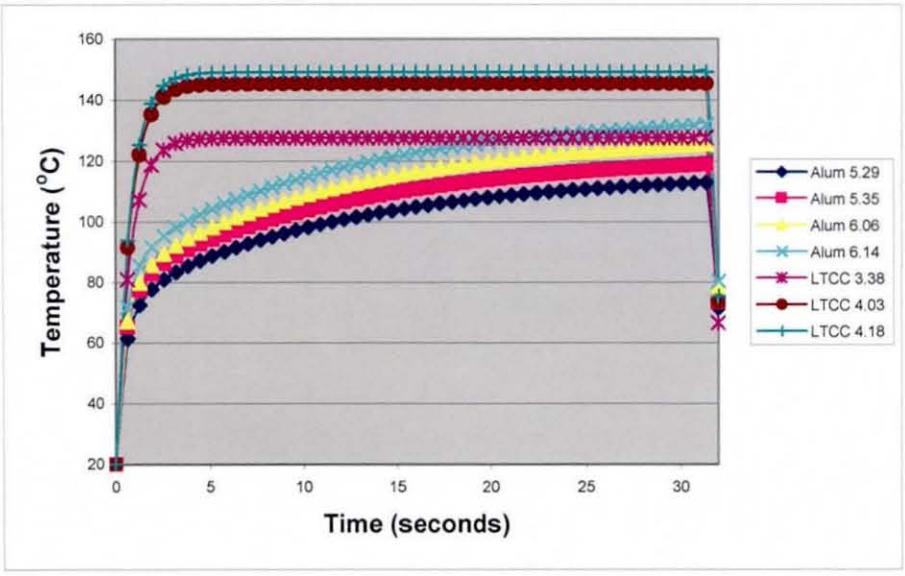


Figure 6.4.4b: *Graphs of transient thermal profile of each of the assemblies*

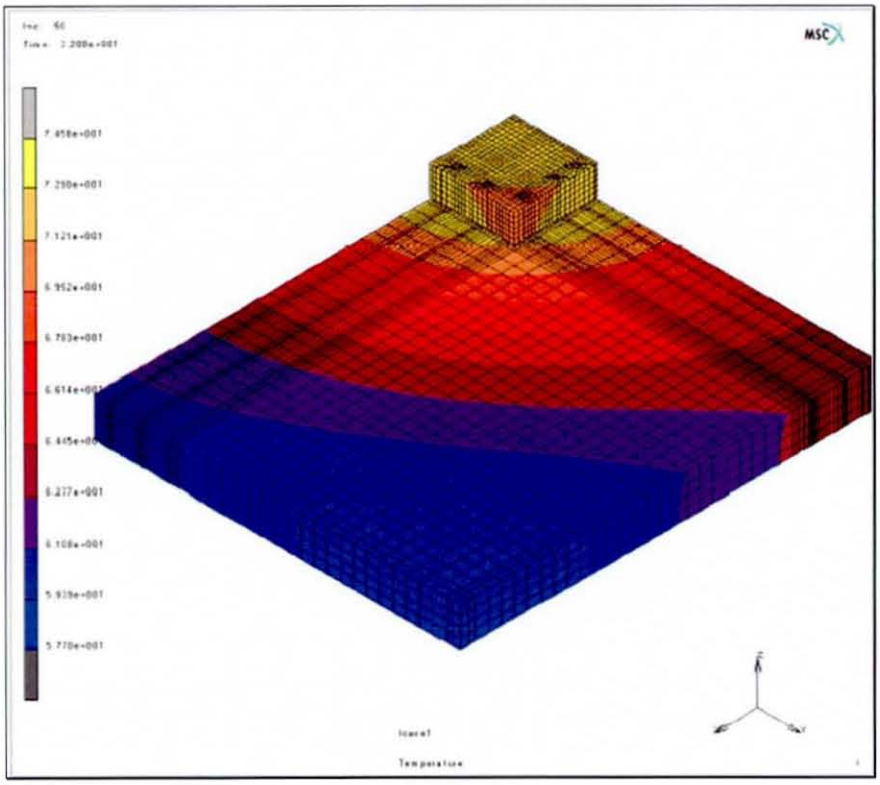


Figure 6.4.4c: *Profile of DCA assembly on Alumina substrate (5.29W powered). The temperature spans 17° C across the substrate (range is 74-57°C)*

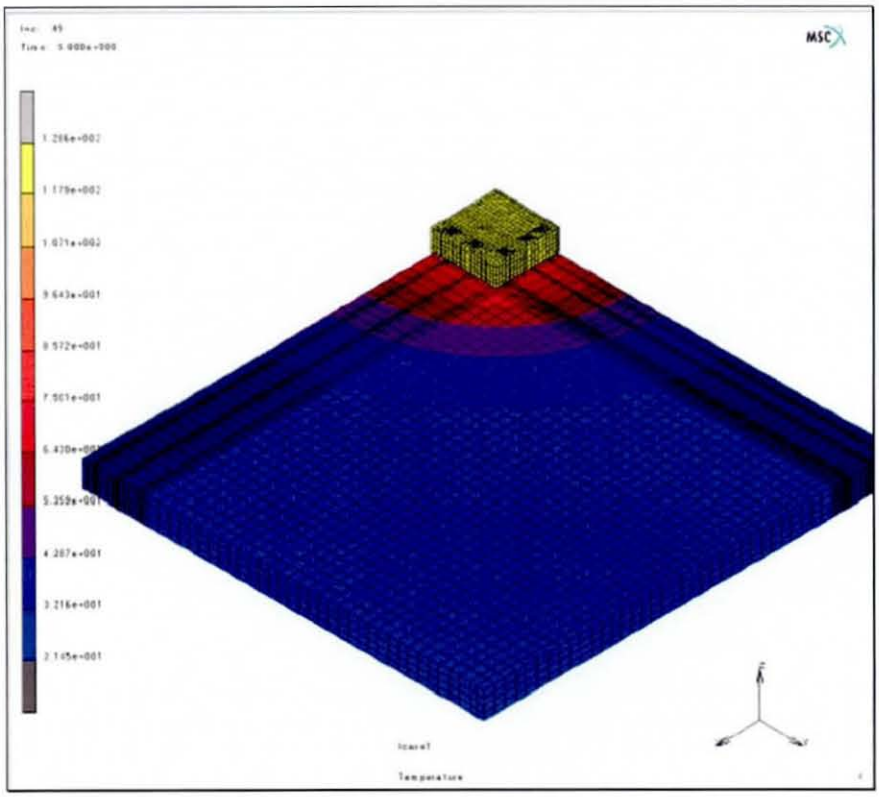


Figure 6.4.4d: *Profile of DCA assembly on LTCC substrate (4.18W powered). The temperature spans 80° C across the substrate (range is 100-21°C)*

6.5: Conclusion

With careful consideration of the modelling parameters, and when substantial care is taken when defining a convection heat transfer coefficient, FE models can be used to estimate the resultant temperature of the MCM assembly. The studies appeared to show somewhat better prediction for the assemblies on FR4 than on copper substrate though these may be attributed to chip manufacture or thermal capture defects. These studies are in good agreement with the findings of Trigg and Corless (1) where a substantial temperature difference between the two chips is still apparent. Furthermore, the transient profile is in agreement with the work by Sur and Turlik (2) where their finite element model predicted the shear in the opposite direction than anticipated originally. However, in this case there is no opposite shear stress as the MCM is CTE matched and the carrier is always cooler than the heater chip.

The steady state analysis confirms the linear effects of increasing power within a module as shown in section 6.4.4 to within over 90 % accuracy. The model has confirmed the influences of changing the substrate materials, the specifics or detail of the transient behaviour may differ marginally though the overall criteria are in good agreement. If required, the model is also capable of capturing fine detail such as when the amended material properties come into effect.

Based on a maximum operating temperature of 125°C for eutectic solder this allows a user to recommend a maximum power load for a given design if the relevant convection criteria are specified. This saves the time and expense of manufacturing "real" test devices and profiling real trials.

References

1) **A.D. Trigg & A.R. Corless** "Thermal Performance and Reliability Aspects of Silicon Hybrid Multi- Chip Modules" *40th Electronic Components and Technology Conference, 1990, pages 592-9*

2) **Biswajit Sur and Iwona Turlik** "Power Cycling and Stress Variation in a Multi-Chip Module" *IEEE Transactions on Components Packaging and Manufacturing Technology part B Volume 10 no 2, May 1995, pages 388-395*

3) **P. Towashiraporn, G. Subbarayan, B. McIlvanie, B.C. Hunter, D. Love, B Sullivan** "Predictive Reliability Models through Validated Correlation Between Power Cycling and Thermal Cycling Accelerated Life Tests" *Soldering and Surface Mount Technology, 14th March 2002, pages 51-60*

4) **S.J. Ham, M.S. Cho and S.B. Lee.** "Thermal Deformations of CSP Assembly during Temperature and Power Cycling" *International Symposium on Electronic Materials and Packaging, 2000, pages 350-57*

5) **Jaakko Lenkkeri and Tuomo Jaakola,** "Rapid Power Cycling of Flip Chip and CSP Components on Ceramic Substrate" *Microelectronics Reliability 41, May 2001, pages 661-8*

Chapter 7: Finite Element Stress Simulation

7.1: Introduction

The fatigue failure of solder joints in flip-chip devices is widely acknowledged to be as a consequence of crack propagation, which is known to be instigated by stresses resulting from thermal expansion mismatches that exist between the die and the substrate. The joint(s) where the stress is highest are typically where the first failure(s) can be found in following the physical testing of a manufactured device and these are almost always the corner joints, or those furthest away from the so-called neutral point.

The aim of this part of the study was to investigate the resultant stress and strain that would develop in the interconnection layer of each assembly by means of a stress FE simulation.

7.2: FE Model Refinement for Stress Analysis

7.2.1: Refinement Description

In order to obtain accurate stress/strain values in a reasonable analysis time, it was necessary to refine the existing FE models that had previously been used for thermal analysis but only in key areas. This process involved increasing the mesh density in the interconnection layer, while coarsening it in less important areas such as the substrate. To do this, the model was divided into different ***contact bodies***, which were defined such that they corresponded with the "real" parts on the test vehicle.

Contact bodies- Description and definition: - In addition to the standard FE pre and post-processing features, MARC-MENTAT has the additional facility of performing an automated solution of problems where contact of some description occurs between two or more bodies. Typically this type of contact is a problem where one body

moves or slides against another (as in common friction problems) though with the correct contact behavioural specification, it can be used to simplify problems where the contact bodies are joined together where the area of interest is very small compared with the overall size of the model, such as is often the case in electronics assemblies *(see **MARC MENTAT 2003** new features guide Chapter 17)*

Each contact body is a set of elements that have been grouped together by having some common properties as specified by the FE operator and the solver perceives each contact body as a FE problem in its own respect. Once all the contact bodies have been correctly defined, entering the contact conditions into a *contact table* specifies the appropriate nature of the contact. The MENTAT contact table allows three modes of contact: *no contact, touching* or *glued.* If there is no contact then this means that there is no physical contact between the two bodies; whereas touching implies that the bodies may come into contact. If *touching* bodies come into contact they may be stationary or they may move about if a suitable disturbing force is applied. Heat transfer or friction can occur between the two bodies. However, if the contact bodies are *glued* then the contact bodies are rigid and can only be displaced if a suitably large separation force is applied; MARC 2001 automatically specifies this separation force to be $1x10^{20}$ Pa.

The MARC contact algorithm does not require any special elements, programs or *tying equations* to be defined by the user. It automatically detects nodes entering or in contact and generates the appropriate constraints to insure no penetration occurs and maintains compatibility of displacement areas and touching surfaces. (MARC MENTAT 2001 user manual)

To determine the most suitable method for refining the model, a series of models were constructed and refined to different levels of detail as summarised in table 7.2.1a. The basic structure used to perform this verification work was the MCM mounted on FR4 board, power cycled with 1.4 Watts and cooled with an airflow of 10 m/s. The simulation time was 64 seconds (i.e. the first ON part of the cycle only). The refined models were evaluated globally in terms of the number of elements used, the stress and thermal results obtained, and the time taken to perform the analysis. Each stage of the refinement is explained below.

Model No	Attributes
0	Original model
R1	Substrate and adhesive coarsened
R2	Heater and carrier chip coarsened
R3	Interconnection layer divided and parts were coarsened
R4	Interconnection layer mesh density increased

Table 7.2.1a: Summary of the model refinements used

Refinement level 1 (R1): Substrate and adhesive coarsened: - In the original model used for the thermal analysis, the substrate and the adhesive parts of the assembly were modelled with 6496 and 1792 elements respectively, constituting over ½ the elements used in the original model. This was considered uneconomical, as the resultant stresses in these areas were unlikely to vary significantly (moreover the purpose of the substrates was to provide the necessary thermal characteristics for the MCM and the stresses there were of little interest). Therefore it was decided to reduce the number of elements in these areas as much as possible. As a result 98 elements were used to represent the substrate, while the adhesive utilised 128 elements. This resulted in three contact bodies used in the model as shown in figure 7.2.1a.

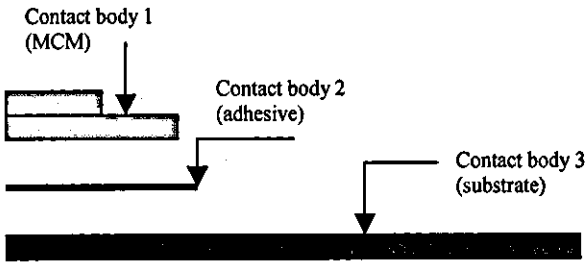


Figure 7.2.1a: - Schematic of the 3 contact bodies considered in the R1 FE model

Refinement level 2 (R2): Division of the MCM into different contact bodies and modification of the mesh densities: - Again, as the stresses within the silicon dies were unlikely to vary rapidly within the bodies, it was decided to investigate the effects of reducing the mesh densities in these areas as well. First, the MCM was divided into three separate contact bodies: the heater-top, the carrier-bottom and the interconnection layer as shown in figure 7.2.1b. The interconnection layer consisted of the solder, the UBMs, the bond pads and a layer of elements from the bottom of the heater chip and the top of the carrier chip. The heater-top was made up of the remaining elements from the heater chip, likewise for the carrier-bottom. After these

contact bodies were specified, the number of elements for the heater-top and the carrier-bottom were then reduced to 128 and 512 respectively. This formed the basis for refinement level 2.

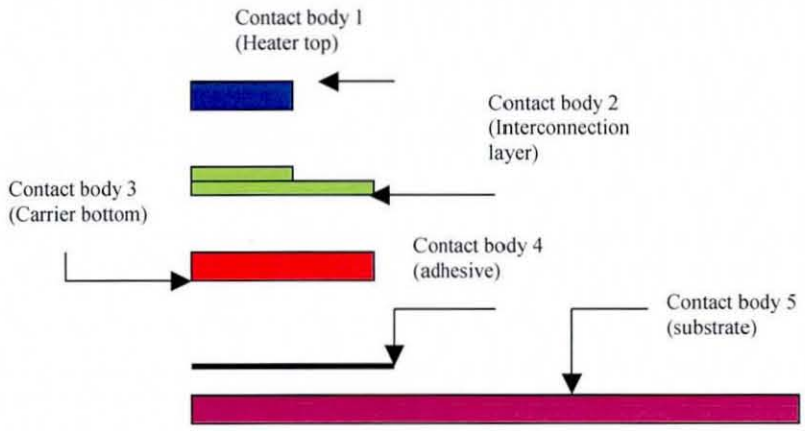


Figure 7.2.1b: - *Schematic of the 5 contact bodies considered in the FE model R2*

Refinement level 3: Division of the interconnection layer: - The interconnection layer was further divided into 4 contact bodies as shown in figure 7.2.1c. One of the bodies contained all the solder joints together with the remaining heater and carrier chip elements, while the other three consisted entirely of the remaining elements from the carrier chip. This enabled the contact bodies that were independent of the interconnection layer to be coarsened and their mesh densities were reduced as shown in figure 7.2.1c.

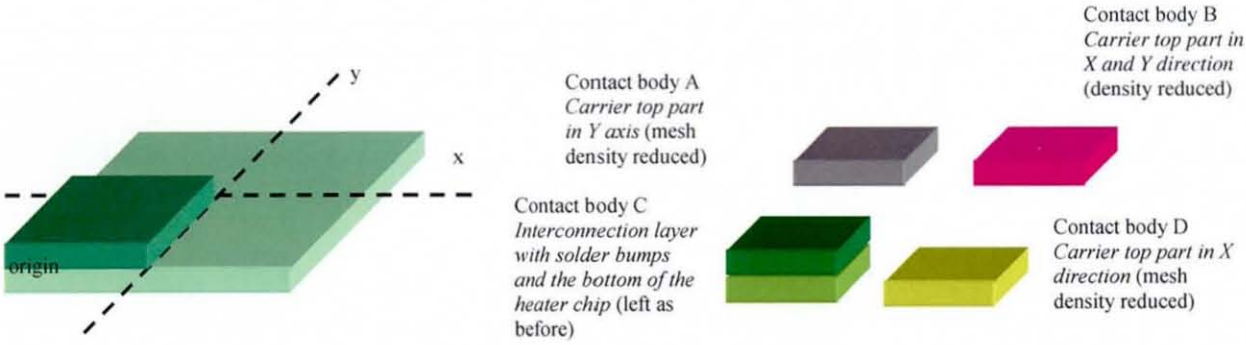
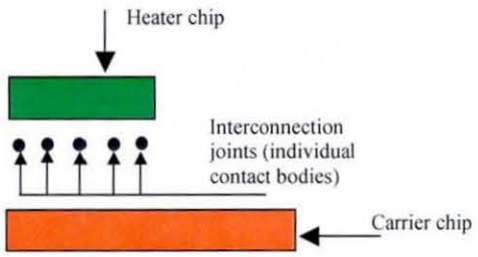


Figure 7.2.1c: *- Illustration of how the interconnection layer was modified for FE model R3*

Refinement level 4 (R4): Mesh density increased for the interconnection layer: - With the number of elements significantly reduced in the interconnection layer (contact body C in figure 7.2.1c) it was then subdivided to obtain a denser mesh at the interconnection layer for a more detailed picture of the solder joints.

Refinement Level 5 (R5): Each solder joint defined as a separate contact body: - The previous models preserved the integrity of the interconnection layer by keeping all the solder joints including the heating elements as one contact body. An alternative method involved defining each solder joint as an individual contact body (a joint consisting of two pads, two UBMs and the solder). By using this approach, the number of elements used to represent the heater and carrier chips could be reduced dramatically, while allowing the mesh density of each individual solder joint to be increased independently to provide a more detailed view of specific areas of interest.



	0	R1	R2	R3	R4	R5
Description of refinement	Original thermal model	Substrate and adhesive coarsened	Heater and carrier chip coarsened	Interconnection layer divided and some parts coarsened	Interconnection layer mesh density increased	Solder joints represented as separate contact bodies
Interconnection layer	4256	4256	4256	3840+(3×64)	15360+(3×64)	9x224
Heater chip	960	960	128	128	128	432
Carrier chip	1792	1792	512	512	512	768
Adhesive	1728	128	128	128	128	128
Substrate	6496	98	98	98	98	98
Total	15232	7234	5122	4898	16418	3442
Software used	MARC 2001	MARC 2001	MARC 2001	MARC 2001	MARC 2001	MARC 2003

Table 7.2.1b: *Summary of the total number of elements used in each refinement model (including contact bodies)*

7.2.2: Model Refinement Thermal Validation

Before the stress predictions from the refined models could be analysed, it was first necessary to verify that the models had adequately predicted the thermal characteristics of the heater and carrier chips. Therefore the temperatures of the heater and carrier chip were compared with the profiles from the original thermal model. This was done by recording the history of the nodes equivalent to those on the original heater and carrier chip as defined in Chapter 6 and comparing the final temperatures reached and the overall thermal profile.

The results of the maximum temperatures reached by the models are shown in table 7.2.2a while the transient profiles of the nodes located on the heater and carrier chips are shown in figures 7.2.2a and b respectively. It can be seen that in general, there is good agreement between the final temperatures reached by the contact body models and the original thermal model, although they marginally under estimate the final temperatures; more so for the case for model R4 with the increased mesh density. The graphs showing the transient profiles of the nodes (figure 7.2.2b), again exhibit an excellent agreement in the manner in which the heater and carrier chips warm up. The results indicated that all of the refinement methods were capable of accurately representing the thermal performance of the assembly and were therefore suitable for further assessment of their use for stress analysis.

Measured attribute	***Original thermal model***	***R1***	***R2***	***R3***	***R4***	***R5***
Time taken to solve (seconds)	25046	7578	7831	6874	51237	N/A
Node 1 (heater chip) °C	74	72.9	72.6	72.5	71.1	72.5
Node 2 (Carrier chip)°C	66.4	65.4	65.5	65.4	64.1	65.2

Table 7.2.2a: *Summary of time taken to run the models and the final temperatures reached by the heater & carrier chips. R5 analysis was performed on a more powerful workstation and therefore it was not possible to compare the time taken.*

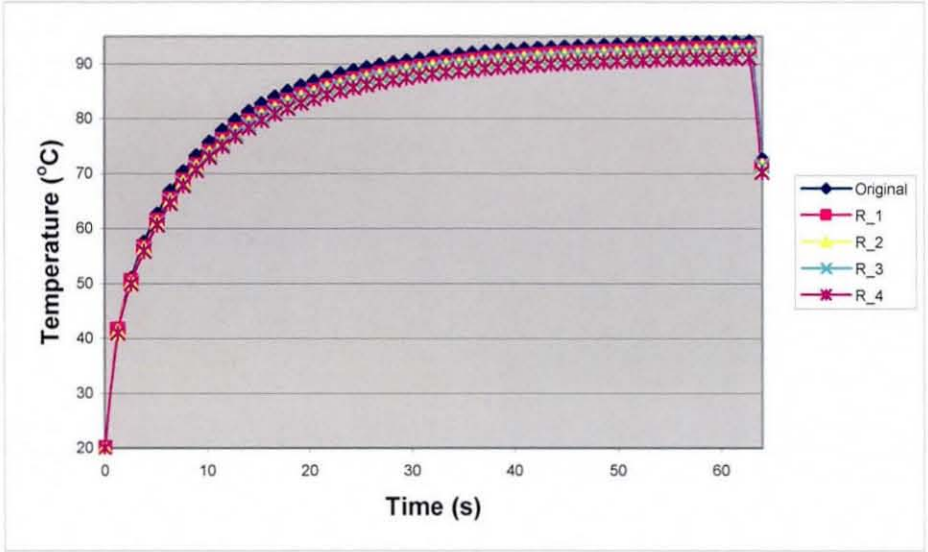


Figure 7.2.2b$_i$: *Graph of transient profiles of nodes on heater chip*

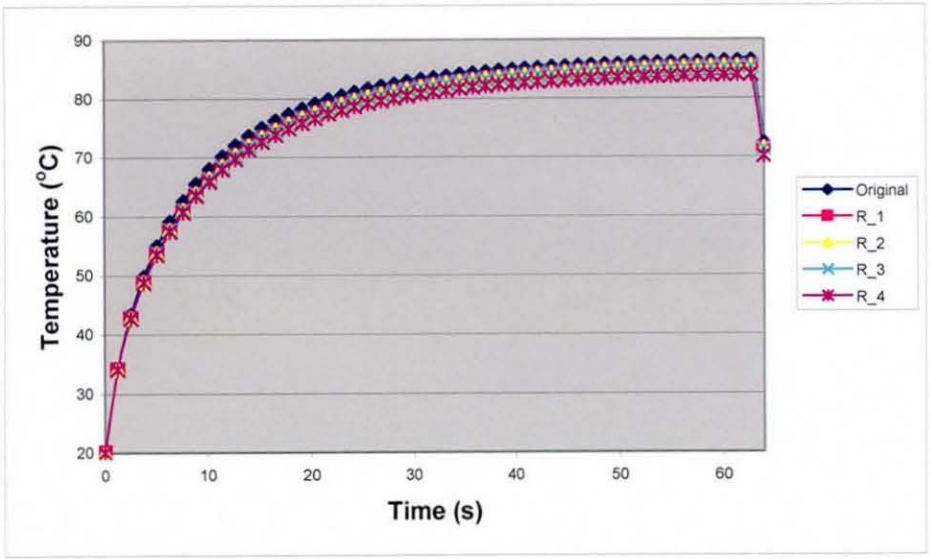


Figure 7.2.2b$_{ii}$: *Graph of transient profiles of nodes on carrier chip*

7.2.3: Model Refinement Stress Assessment

The predicted stress levels from each model were evaluated by examining the stress plots from 4 of the 9 solder joints from the penultimate increment of the model cycle (i.e. immediately before the power was "switched off"), so that stress distributions close to the maximum temperature were obtained. The layout of the solder joints from the FE model is shown in figure 7.2.3a and the solder joints were assigned the numbers shown. Solder joints 1, 4, 5 and 9 were initially chosen for analysis. For each model, the stress distribution within the solder joints was examined and compared with the equivalent joint from other models (figure 7.2.3). The accompanying scales for all of the figures were all standardised to allow a direct visual comparison with the other models.

It can be seen that from all the obtained plots, FE models R1 and R5 persistently predicted higher stress levels than R3 and R4. Furthermore, R5 persistently identified the same profiles for all the joints, irrespective of its displacement from the neutral point. R1 on the other hand, shows some bias of the location of the bump with respect to the neutral point as the location of the minimum stress has moved to the left of the profile in bump 4. Bump 5 shows the minimum stress in the right hand side and it is located in the centre of bump 9. R2 shows the most diverse stress distribution with a single peak value, similar to those in R1 & R5 whilst the plot also has the low stress regions shown in R3 and R4.

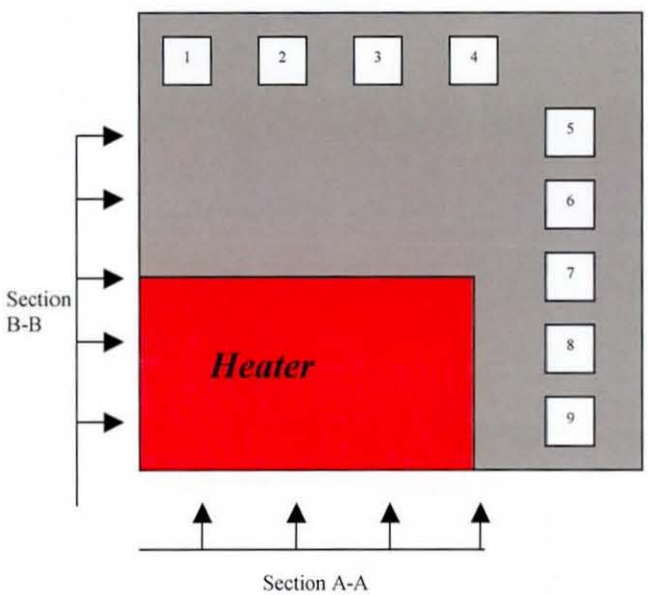
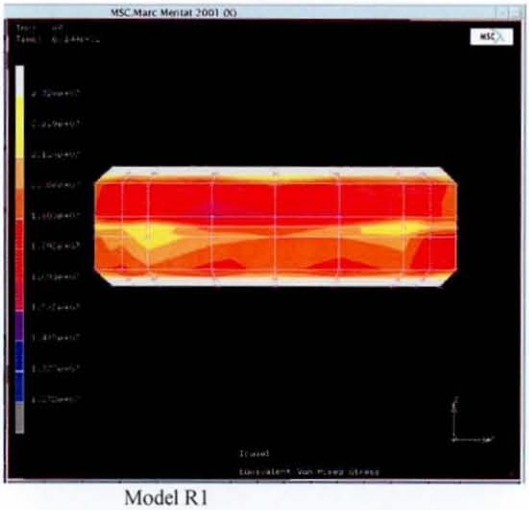
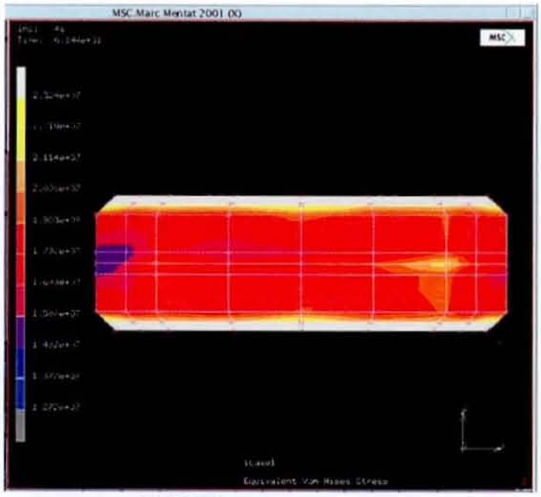


Figure 7.2.3a$_i$:
Illustration of the solder joint number allocation. The lines show the axes and where the views were taken from.

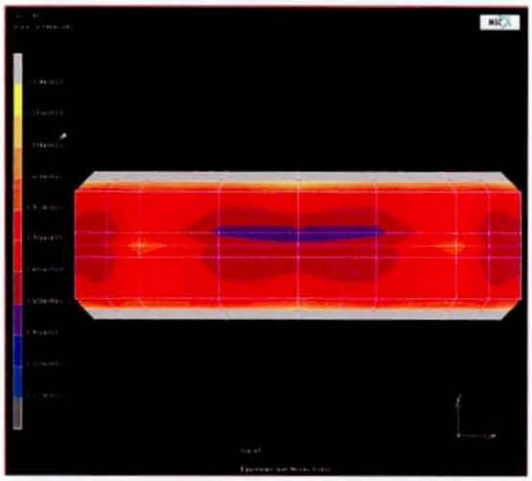
BUMP 1



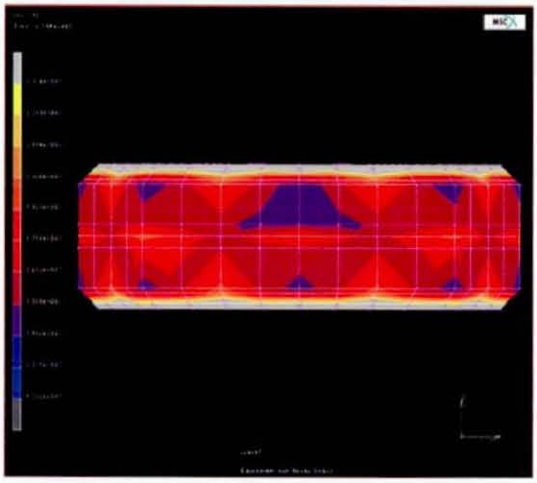
Model R1



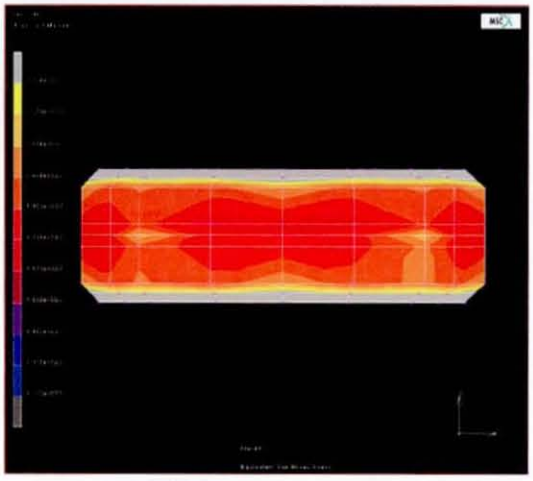
Model R2



Model R3



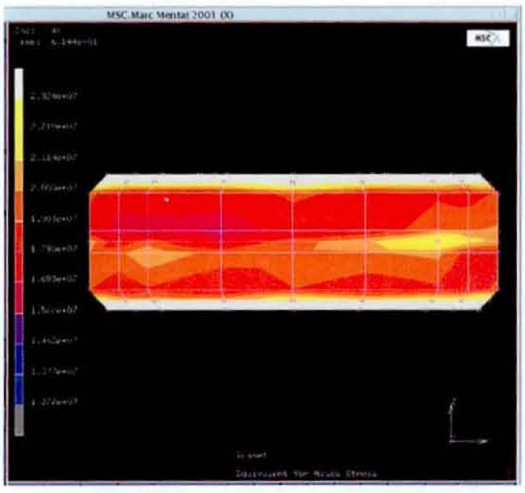
Model R4



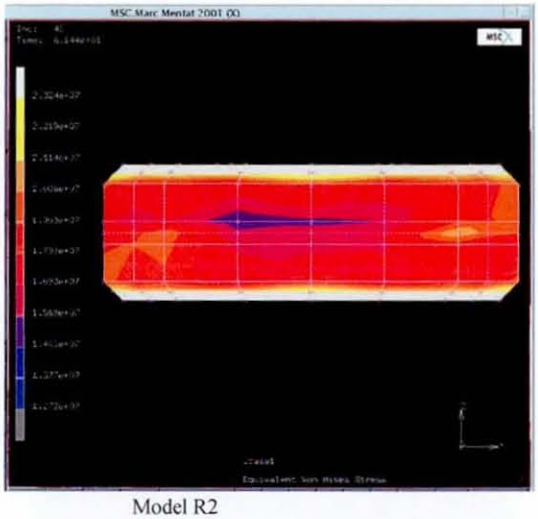
Model R5

Figure 7.2.3b$_i$:
The stress contour plots of bump 1 obtained from the different models as labelled beneath each picture. The contour range has been standardised to aid comparison. Note the similar profiles exhibited by models R1 and R5, both predicting peak stresses omitted by the others, also note the minimum stress located in the centre of the plot for both models. Models R2, 3,4 both show lower values, though R2 does show the most varied distribution, exhibiting one of the peak values shown in R1 & 5 as well as the low values of R3 and R4. The lower regions on R4 were attributed to the marginally lower temperature reached by this model in the thermal analysis. The view is the exterior of the bump taken from A-A.

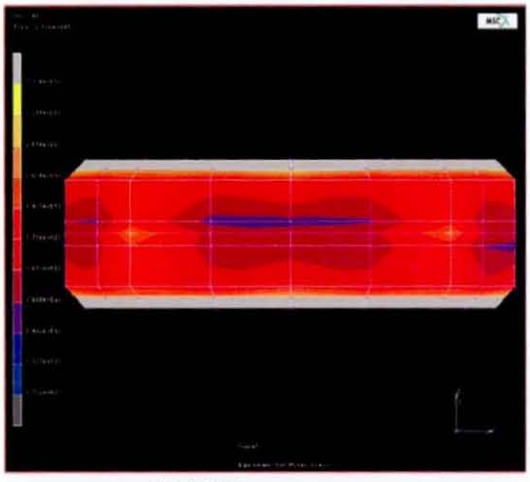
BUMP 4



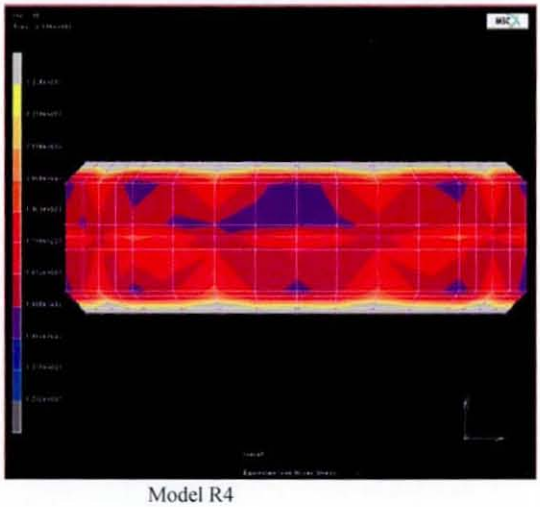
Model R1



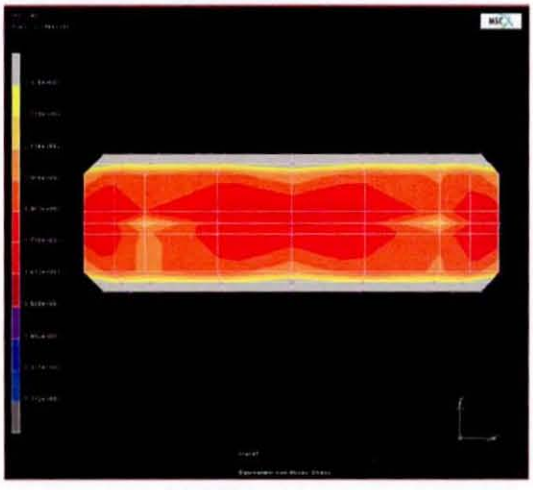
Model R2



Model R3



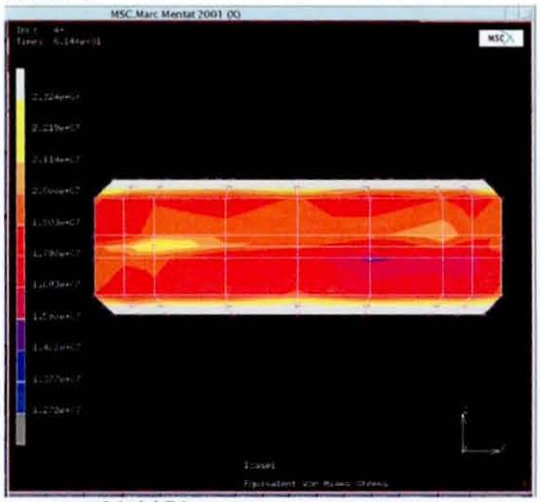
Model R4



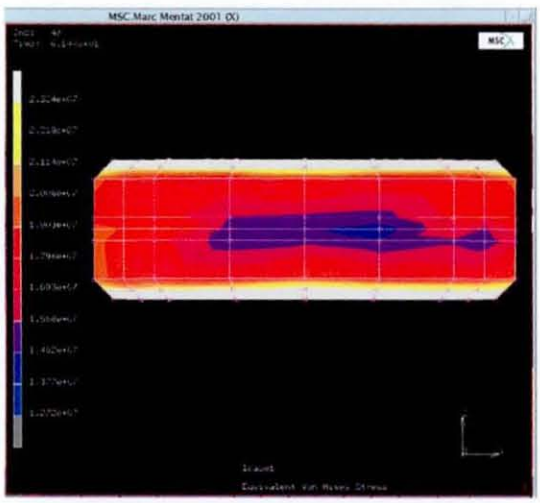
Model R5

Figure 7.2.3b$_{ii}$:
Stress contour plots of bump 4 obtained from the different models as labelled beneath each picture. The contour range has been standardised to aid comparison. It can be seen that R1 and R5 have predicted similar peak stress levels in the solder joints, similar to those exhibited in bump 1; however note that the location of the minimum stress has moved to the left of the plot for R1 while it remains in the centre for R5. The peak stress in R2 remains in the same place though the low region has moved from the side to the central region. The view is of the exterior of the bump taken from A-A.

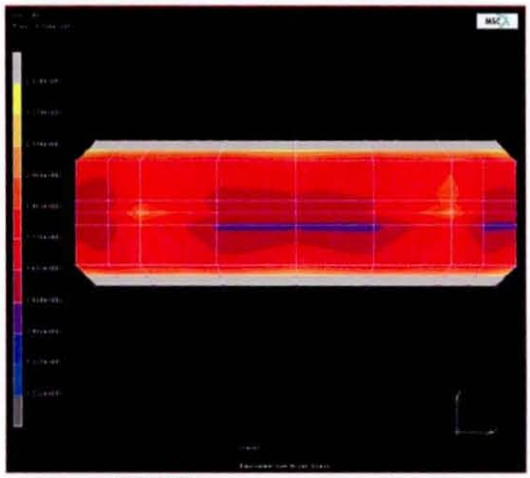
BUMP 5



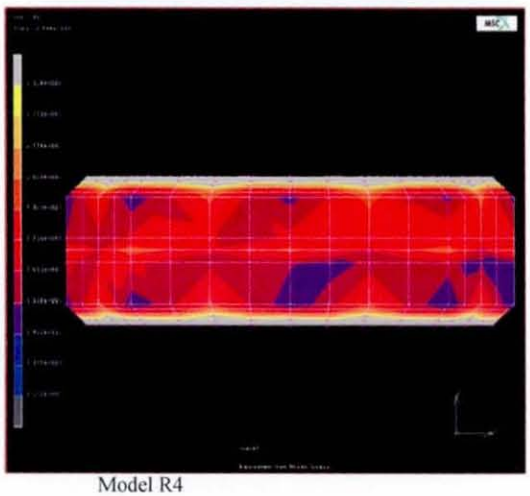
Model R1



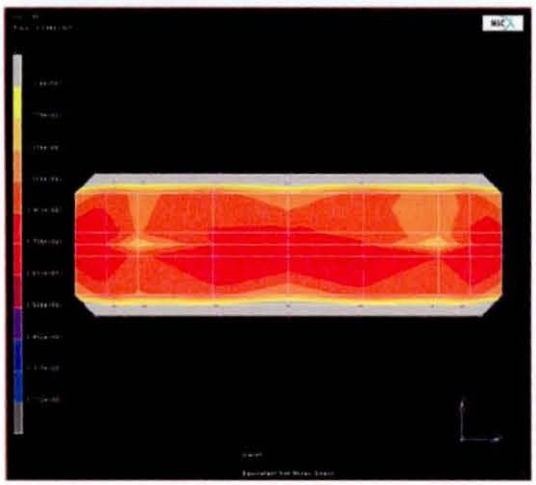
Model R2



Model R3



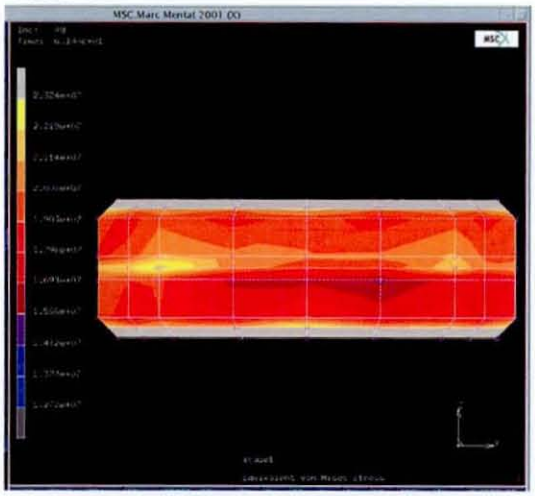
Model R4



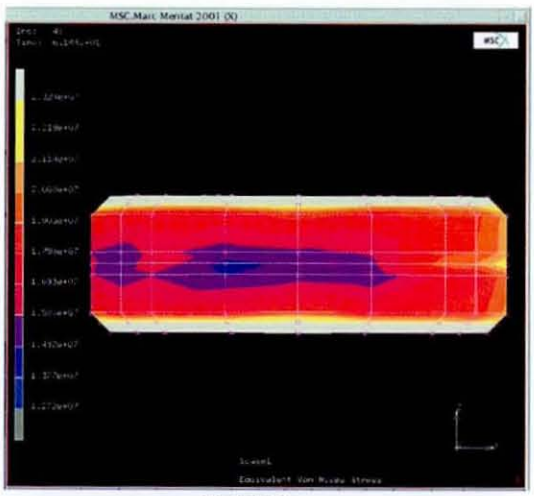
Model R5

Figure 7.2.3b$_{iii}$:
Stress contour plots of bump 5 obtained from the different models as labelled beneath each picture. The contour range has been standardised to aid comparison. It can be seen that R1 & R5 has predicted the same stress peaks as in the previous models, though the minimum stress in R1 has moved to the right of the plot. The minimal stressed areas in all the models are more prevalent in the lower regions of the bump. The peak stress in R2 bumps 1 & 4 is not present in this view. The view is of the exterior of the bump taken from B-B

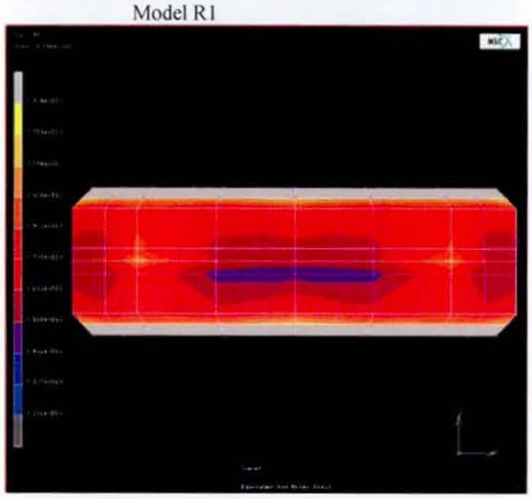
BUMP 9



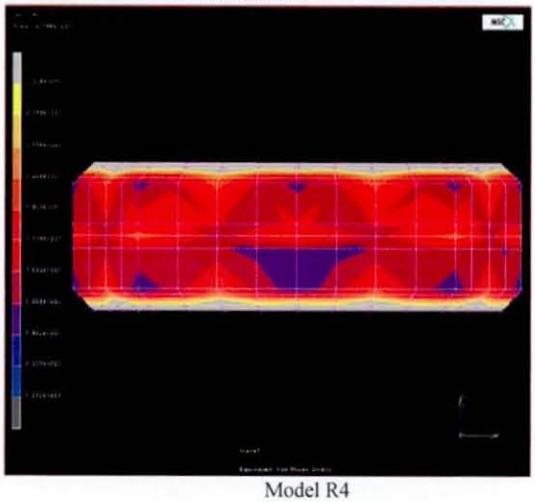
Model R1



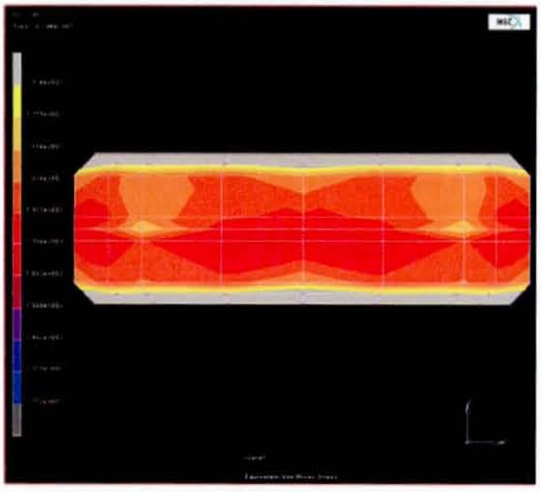
Model R2



Model R3



Model R4



Model R5

Figure 7.2.3b$_{iv}$:
Stress contour plots of bump 9 obtained from the different models as labelled beneath each picture. The contour range has been standardised to aid comparison. The profiles from R3 and R4 are very similar to those of Bump 5. From model R2 the maximum stress values have moved to the side of the joint. The locations of the maximum values from R1 and R4 are in good agreement. The view is of the exterior of the bump taken from B-B.

To provide a quantitative comparison between the models, the node depicted in figure 7.2.3c_{ii} was selected for all the joints and the stress value at this location was recorded.

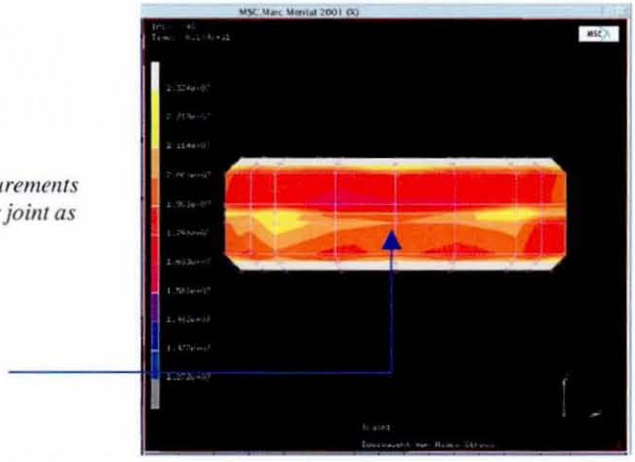


Figure 7.2.3c_{ii}:
Location of where the measurements were taken from each solder joint as described in the tables.

Joint number	***Original model***	***R1 (MPa)***	***R2 (MPa)***	***R3 (MPa)***	***R4 (MPa)***	***R5 (MPa)***
1	16.89	19.36	17.94	16.03	18.94	18.03
2	16.89	18.17	16.50	16.12	19.04	18.19
3	16.88	18.37	16.14	15.97	18.82	18.13
4	16.91	19.68	16.51	16.2	18.97	17.89
5	17.13	19.46	16.60	16.54	19.49	18.44
6	16.89	18.17	16.19	16.23	19.19	18.54
7	16.9	18.09	15.95	16.19	19.15	18.3
8	16.89	18.48	16.06	16.23	19.18	18.04
9	16.91	19.54	16.02	16.27	19.22	18.03

Table 7.2.3a: *Equivalent Von Mises stress (MPa) obtained for the node shown in figure 7.2.3c_{ii}*

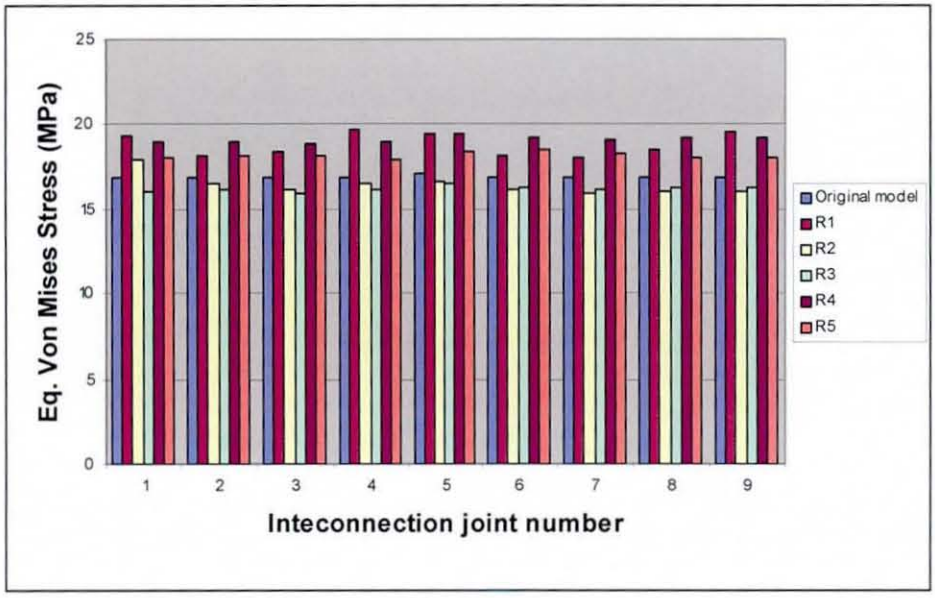


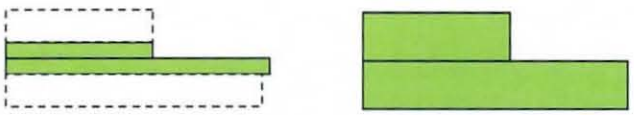
Figure 7.2.3d: *Bar chart of results from table 7.2.3a*

Stresses at the Joints

Comparing the relevant stress values from table 7.2.3a with the overall profiles, it can be seen that the values at the nodes are not a true representation of the overall profiles. However they do serve to illustrate how each change in the model affects the results obtained at each node. From the table and figures 7.2.3a it can be seen that there is general agreement with the final stresses values reached in the joint. Models R1, and R4 and R5 predict marginally higher stress values at the selected nodes than the original model, R2 and R3.

General MCM Deformation

To assess each model globally, the deformation plots of each model are also shown in figures 7.2.3 e_i to e_v. These plots offer an indication of the overall model integrity. The deformation is magnified 2000 times so it is emphasised. Looking closely at the plots for the different models it can be seen that there is some separation or fracture at some of the contact body interfaces on refinement levels R2, R3 and R4. This implies that the solver perceives the different contact bodies as separate entities even though the structure is known to be modelled as continuous. Therefore the FE model solved resembles that shown in figure 7.2.3d_i as opposed to the correct model shown in figure 7.2.3d_{ii}



7.23d$_i$ Model as perceived in the stress calculation. Broken line indicates different contact bodies perceived in refinement levels 2,3 &4 (stress calculation)

7.2.3d$_{ii}$ MCM as it should have been perceived by the solver, as was implemented in refinement level1 1 & the original

***Figure 7.2.3d**: Schematic of perceived problems against correct model*

The FE simulations of models R1 to R4 were performed on the original workstation where MARC MENTAT 2001 was installed and are shown in figures e_i to e_{iv}. On this edition, a separation force had to be specified in the governing contact table for each instance of two *glued* contact bodies. The default value of $1x10^{20}$ Pa was used initially and this produced the distortion plots obtained. Upon increasing the value to $1x10^{25}$ Pa, the distortion was still present but less prevalent than shown in the plots. When the separation force was increased to $1x10^{30}$ Pa, the distortion was further reduced. However, it should be noted that $1x10^{20}$ Pa is an exceptionally large value and upon subsequent examining of the stress profiles there was no evidence that such a stress value was present within the model. Moreover, the profiles of models R1 to R4 show evidence of nodal displacement at the bottom of the carrier chip where there symmetry lines were applied. This issue was raised with a MARC representative and this was attributed to a contact algorithm "bug" that was present within the 2001 edition; This along with several other bugs was corrected in the subsequent 2003 version. This version had the additional feature of allowing no separation between glued contact bodies so the bodies remained rigid regardless of the separation force

applied. With this in mind, the R1 refinement model was run again using MARC 2003 and compared against the 2001 version and is shown in figures $e_{v\text{-}a}$ and $e_{v\text{-}b}$. It can be seen that both the separation and the nodal displacement issues have been rectified and a more plausible deformation profile was obtained. Model R5 was also ran using MARC 2003 and the deformation plots were obtained and compared with those of model R1 (2003 version) and can be seen in figure $e_{vi\text{-}a}$ and $e_{vi\text{-}b}$. Comparing the profiles of Refinement level 1 and 5 from the models ran using MARC 2003 it can be seen that there is very little difference in the overall profile of the MCM assembly thereby implying that refinement level 5 is an adequate representation of model 1 using the updated software.

As a consequence, the stress simulations were based on the refinement method R1 where the whole MCM was kept as a single contact body and R5 where each solder joint was defined as an individual contact body. In terms of the actual assembly geometry, R1 and R5 offered the best representation as the heater and carrier chips did not exhibit any discontinuities as in the real case. Model R1 produced the best results where the orientation of each stress distribution in the joints can be seen, and in addition R1 gave the closest results to the original FE model. However, this model was moved to the superior workstation such that no separation allowed could be specified.

Model R5 generated results that were comparable with those from level 1; while they omitted the orientation of the stress profile, they did offer the potential benefit of using few elements to represent the heater and carrier chips. Furthermore this allowed the mesh densities of individual solder joints to be increased or reduced accordingly thus allowing greater confidence in the stress predictions; the details of how the two different refinement methods were utilised is specified at the beginning of section 7.3

R1

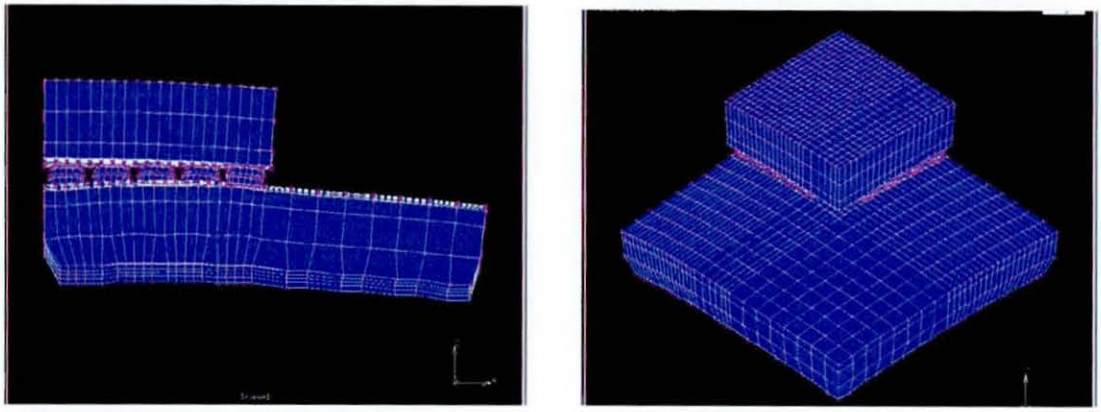
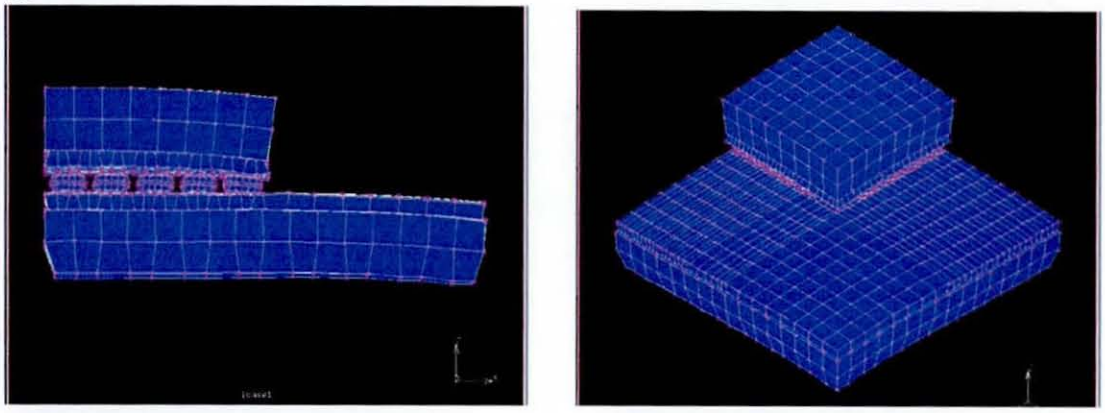


Figure 7.2.3ei: -
Plot of the deformation of the model when refinement level 1 is used. Note the continuation of the heater and carrier chip bodies near the interconnection layer. The contraction at the bottom is due to the reduced temperature at the bottom of the carrier chip The deformation is magnified 2000 times

R2



*Fig**ure 7.2.3eii:***
Plot of the deformation of the MCM when refinement level 2 is implemented. Note the discontinuities at the top of the carrier chip and the bottom of the heater chip on both plots. The distortion at the bottom of the carrier chip is due to the reduced temperature at the bottom. The deformation is magnified 2000 times

R3

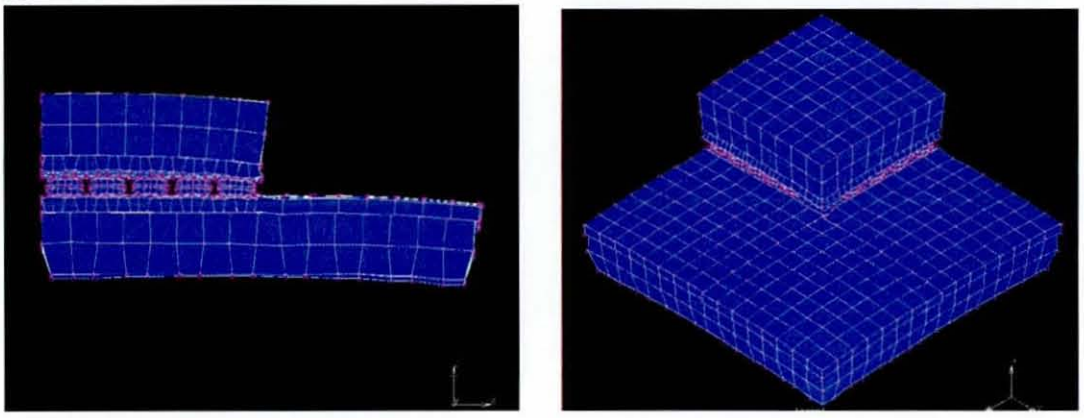


Figure 7.2.3eiii:
Plots of the deformation when refinement level 3 is implemented. Note the discontinuities from the first model are still evident, in addition, contact body C can be seen to be breaking away. The deformation is magnified 2000 times.

R4

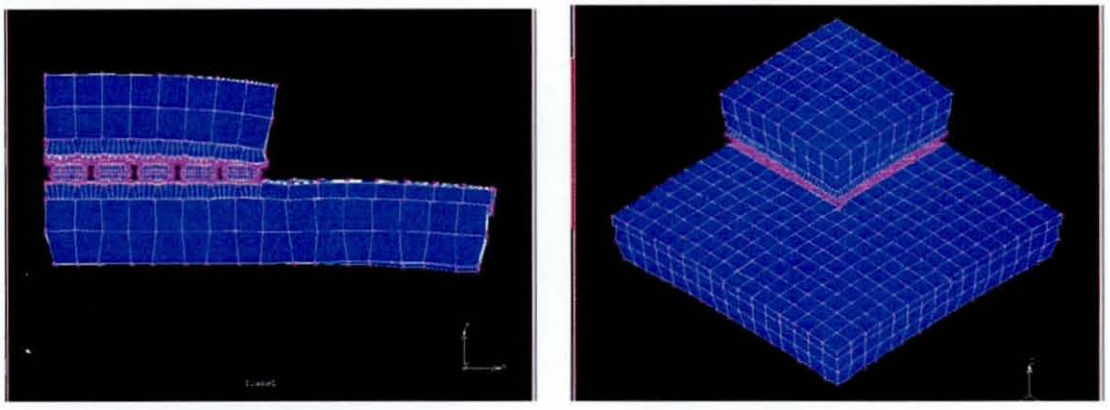


Figure 7.2.3eiv:
Plots of the deformation when refinement level 3 is implemented. Note the discontinuities from the first model are still evident, in addition, contact body C can be seen to be breaking away. The deformation is magnified 2000 times

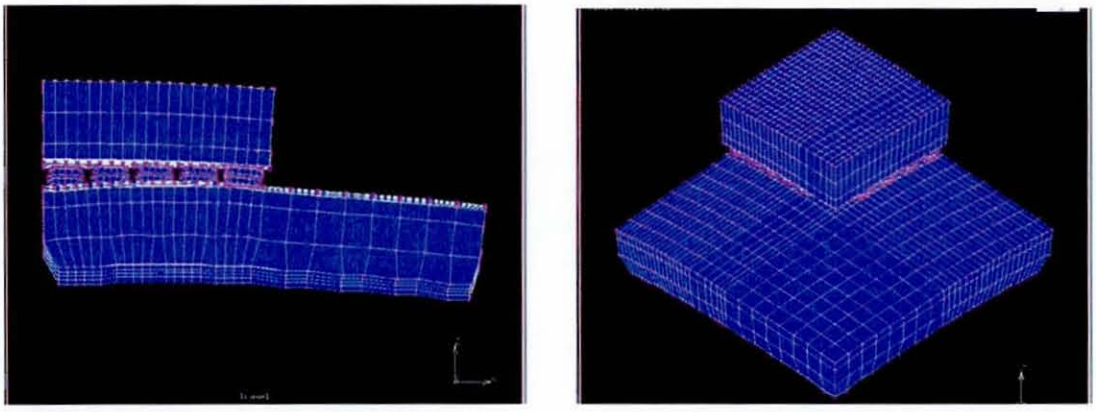


Figure 7.2.3ev-α:
Plots of the deformation when refinement level 1 performed on MARC 2001. The deformation is magnified 2000 times

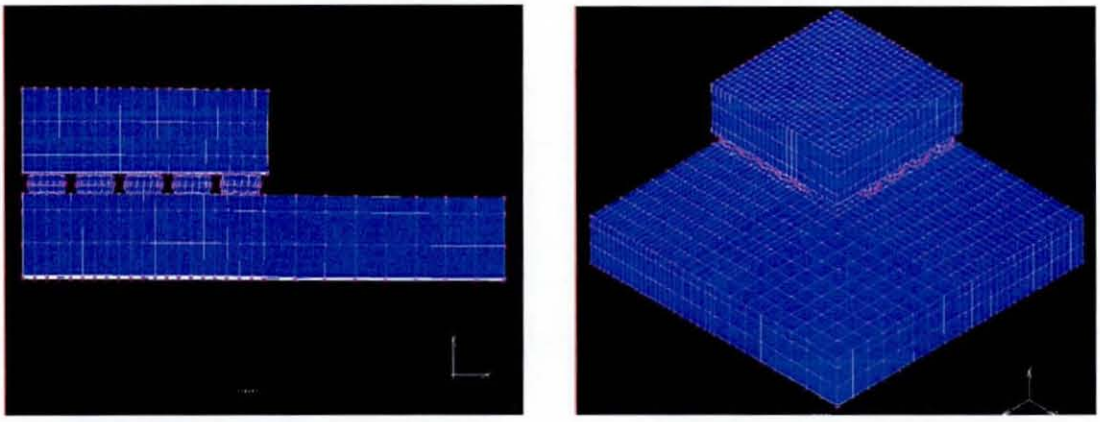


Figure 7.2.3eiv-β:
Plots of the deformation when refinement level 1 performed using MARC 2003. Note that the carrier chip no longer bends nor is there displacement of the nodes where the symmetry axes were applied. The deformation is magnified 2000 times.

R1

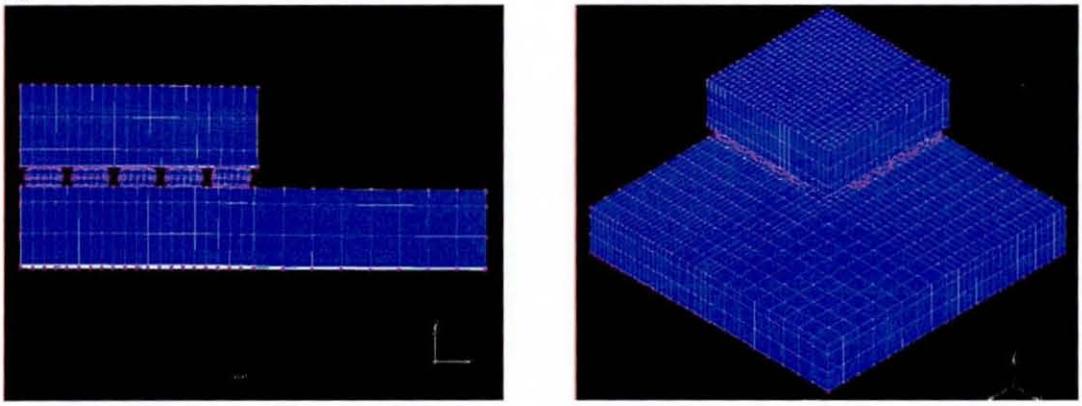


Figure 7.2.3evi-α:
Plots of the deformation when refinement level 1 performed on MARC 2003. The deformation is magnified 2000 times

R5

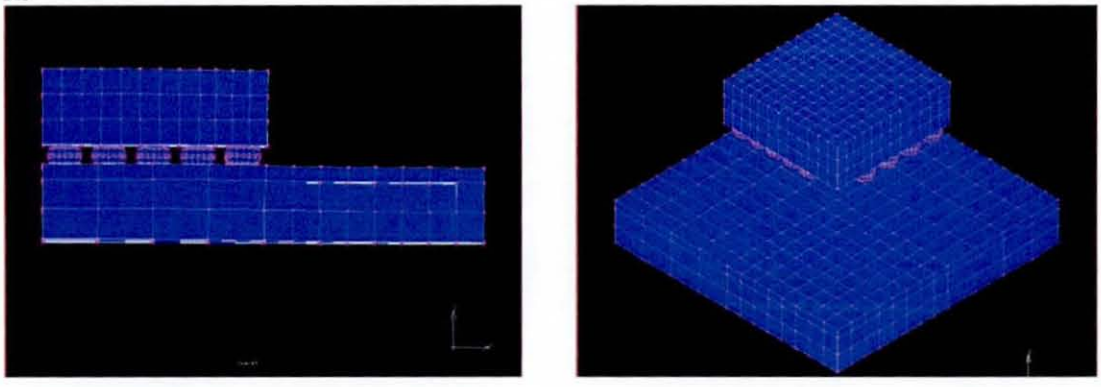


Figure 7.2.3eiv-β:
Plots of the deformation when refinement level 5 is implemented. Note that the deformation profile bears a good resemblance with that of R1, providing better confidence in the models . The deformation is magnified 2000 times.

7.3: Stress Modelling of MCM Assemblies

Model R5 was run with two of the solder joints (joints 4 & 5 from figure 7.2.3ai) meshed densely and was used to obtain a more detailed stress profile while the remaining solder joints were left as they were in the original models. Inevitably this significantly increased the time and memory requirements to run the model such that using the desired time steps (1.28s= 1 increment), it was not possible to run more than the first cycle with this model so model R1 was run for four cycles; as the nominal stress values were in good agreement, R1 was considered an acceptable model to use. The FE models were run for assemblies with both FR4 and copper substrates power cycled with 1.2W and 1.4W. The remainder of the test conditions were consistent with those from the previous transient thermal analysis work (i.e. 64 seconds ON/OFF and airflow at 10m/s).

In order to simplify the results analysis, the solder joints were nominally split into three different regions as defined in figure 7.3a. The **upper region**: that contained the solder elements and the UBM from the heater chip, the **central region** that consisted only of solder elements and the **lower region** that contained the elements of the lower region of solder and the carrier chip UBM.

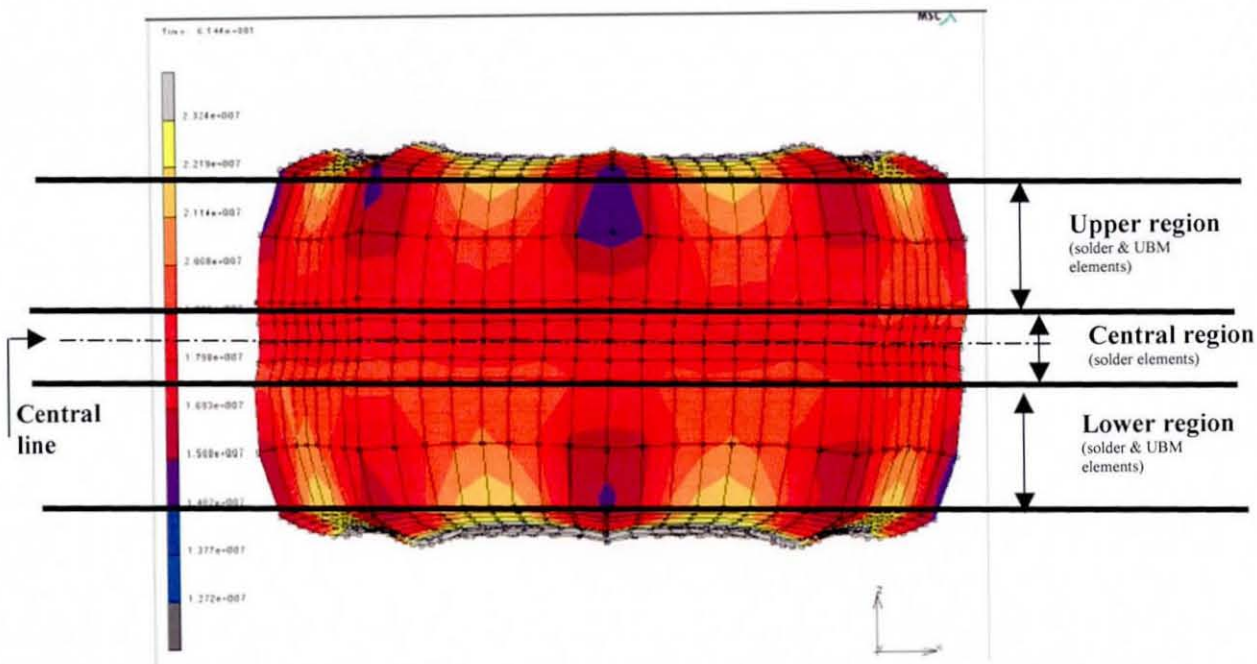


Figure 7.3a: *Definition of the different regions of the solder joint. Also shown is the central line.*

The cylindrical shape of the real solder joints was approximated by the octagonal shape as described in section 5.4. When the FE model or contact bodies were subdivided it was considered impractical to manually reposition all the nodes to form a better approximation of the desired shape. As a consequence, the shape of all the solder joints including those that were dense meshed remained in the original octagonal shape. Unfortunately this resulted in a visible "ripple" effect in the stress distribution and distortion.

Initial solder stress state: Previous work has shown that an equilibrium stress state exists in solder just after reflow and stresses arise when the solder cools down to room temperature before creep and stress relaxation reduces and eventually eliminates these stresses. The FE model was intended to replicate the conditions of the original experiment. In the experiment, the manufactured test vehicles were stored for a period of at least one month before they were power cycled. This was assumed to be a sufficient time period for the stress relaxation to occur therefore they were assumed to be at a neutral stress at the beginning of the experiment.

Intermetallics: Any intermetallics present within the solder joint would have been formed from the solder and the nickel UBM as there was no copper in direct contact with the solder to allow for tin-copper intermetallics. The prevalent compound at the UBM solder interface has been identified from previous work as Ni_3Sn_4 (1). It is also known that the solder/copper intermetallics (namely Cu_6Sn_5 and Cu_3Sn) grow much faster than the nickel based ones; a typical interconnection layer is of a nickel based UBM, solder and a copper pad, therefore most damage modelling work published has been performed with respect to the copper-tin intermetallics. He et al (1) conducted thermal aging tests to investigate the way of the Ni_3Sn_4 intermetallic layer propagates. It was found to grow linearly in proportion with the square root of time and is faster at higher temperatures, and also much faster depending on the lead quantity (tin silver eutectic solders had much higher growth rate than the eutectic tin lead solders). When they aged the 63/37 SnPb solder / UBM specimens for 625 hours at 130°C and 150°C, and examined the intermetallic growth layer. It was found that 130°C case had grown to 0.25μm and 0.75μm for the 150°C. The solder in these experiments would have reached a temperature no greater than 90°C for a total period of no greater than 400 hours.

The implications from He et al imply that the intermetallic layers will be very thin in this study and therefore there were assumed to be of little consequence by omitting them from the FE stress analysis.

7.3.1: Simulation Results for MCM Assemblies with FR4 Substrates

7.3.1.1: Linear Stress Results (1.4Watts)

The distributions of the resultant Von Mises stresses in the solder joints for the 1.4-Watt power cycle simulation are shown in figure 7.3.1.1ai. It can be seen that all the joints exhibit similar profiles with an overall stress value of 17-18MPa. It was interesting to note that the stress profiles of the bumps did not vary significantly with their position from the neutral point in the assembly (they all exhibited similar stress values), thereby implying that the joint position to the neutral point was less significant in this experiment than traditional thermal cycling results have indicated.

The model was first ran without plastic strain implemented. A stress profile of the surface of solder joint 4 surface from model R5 viewed from the direction AA of the MCM is shown in figure 7.3.1.1b_i. From both profiles, it can be seen that the stresses are marginally lower within the central region than in the upper or lower regions of the solder. It should be noted that in the central region, there are only solder elements at this area whereas in the upper and lower regions there are the UBM elements also. The higher stresses in these areas were attributed to a local CTE mismatch between the solder and the UBM; an effect which is not apparent in the central region due to the uniform material.

Figure 7.3.1.1b_{ii} shows the sectioned view of solder joint 4 taken from the direction AA. It can be seen that higher stress values were obtained for the solder in the central region; particularly at the very top and bottom of this region at the solder/nickel interface. A higher stress value was plausible due to the materials having different CTEs; however it should be noted that the solder mesh density, while better than the initial model was still low within this region and it is possible that the stress values within this region would be lower if a more suitable mesh density was used.

Furthermore, it is emphasised that the stress levels recorded are greater than the yield stress of 63/37 SnPb solder within the temperature range considered.

The Von Mises stress plots of the maximum node for time against temperature are shown in figures 7.3.1.1c and 7.3.1.1d respectively. It can be seen that the behaviour of stress with respect to time follows a similar profile of the temperature Vs time; with the initial steep change that levels off as the time continues. A strong relationship with temperature and stress was expected as the solder was assumed to be in a stress free state at room temperature and as the stresses would develop as the solder material expands and the shear effects of the temperature difference between the heater and carrier chip increase as they get hotter. Observing the stress characteristics of the OFF cycle however, it can be seen that the stress curves do not level out as they do at the end of the ON cycle. From the stress-temperature plot, it can be seen that there is not a linear relationship between the temperature and stress. The non-linear characteristics of the curve were attributed to the reduction of Young's modulus in the 63/37 SnPb solder as the temperature increases, therefore at higher temperatures (say 80°C) a larger temperature change will be needed to produce an equivalent stress increase than at 20°C.

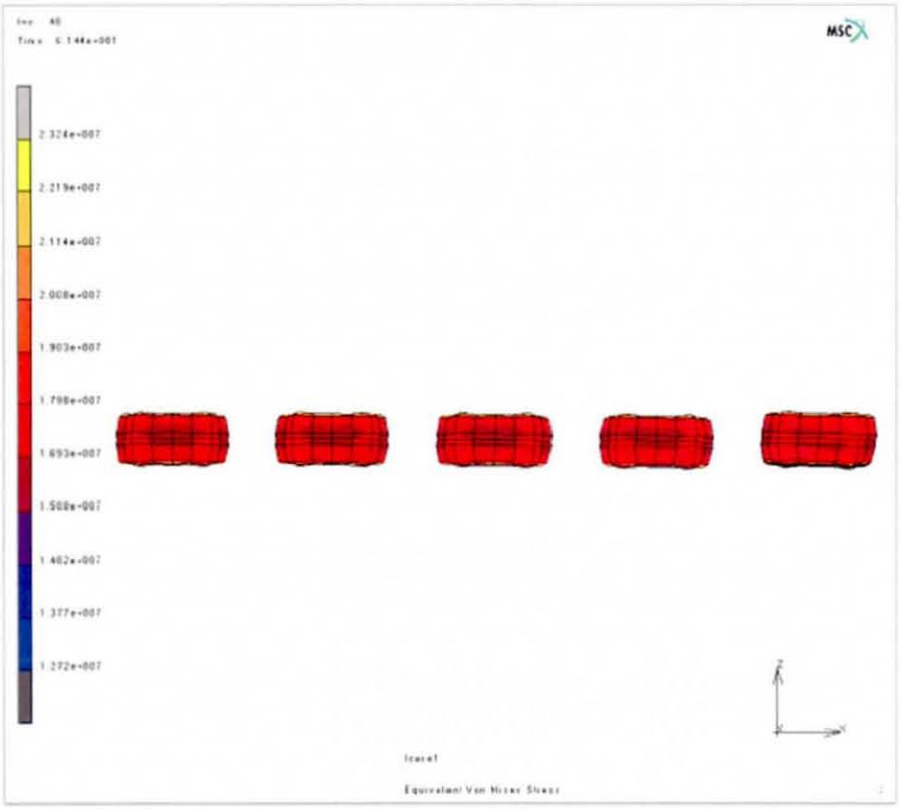


Figure 7.3.1.1a$_i$: *Von Mises stress of joints, taken from view AA. Note that the bumps all exhibit similar stress profiles, implying there is little dependence on the joint position with respect to the neutral point.*

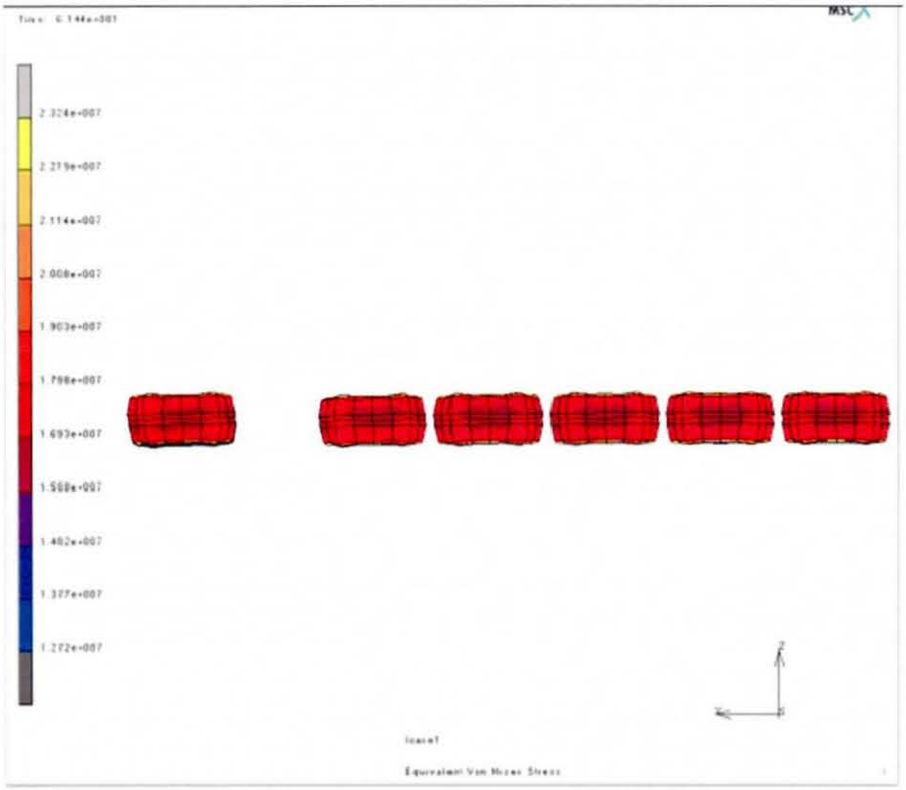


Figure 7.3.1.1.a$_{ii}$: *Von Mises stress of joints, the view was taken from BB*

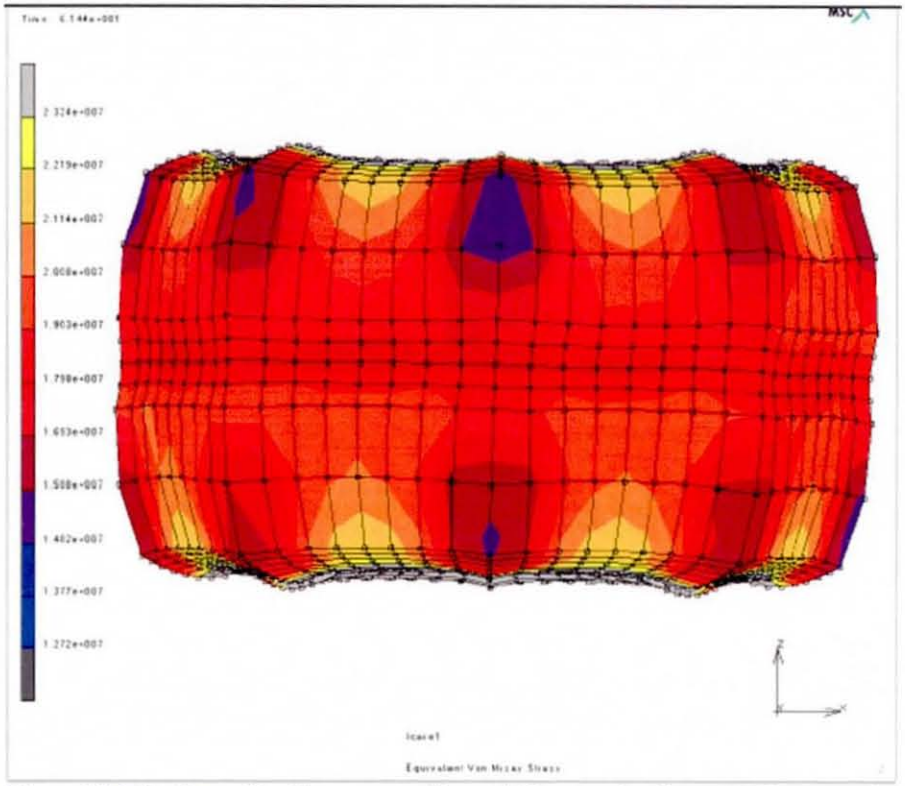


Figure 7.3.1.1b$_i$: *Linear Von Mises stress of bump 4. View is taken from AA. The unusual stress contours are of a consequence of approximating the solder joint as an octagonal shape.*

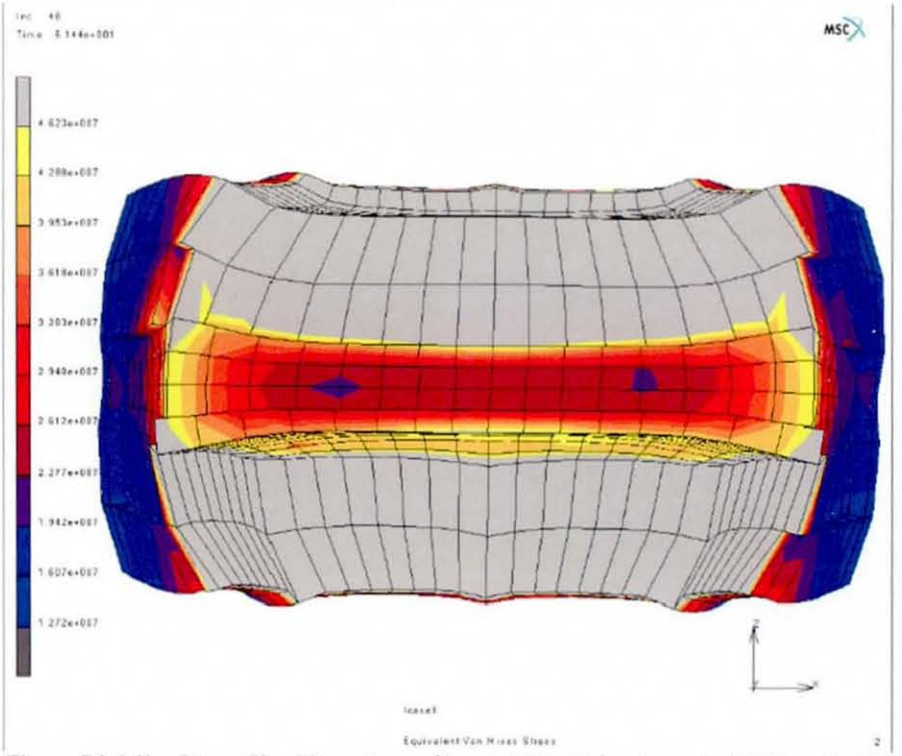


Figure 7.3.1.1b$_{ii}$: *Linear Von Mises stress of bump 4 (view taken towards AA) Note the higher stress levels of the solder sandwiched between the upper and lower UBMs. This was attributed to the CTE mismatches of the UBM and solder and the very low standoff height.*

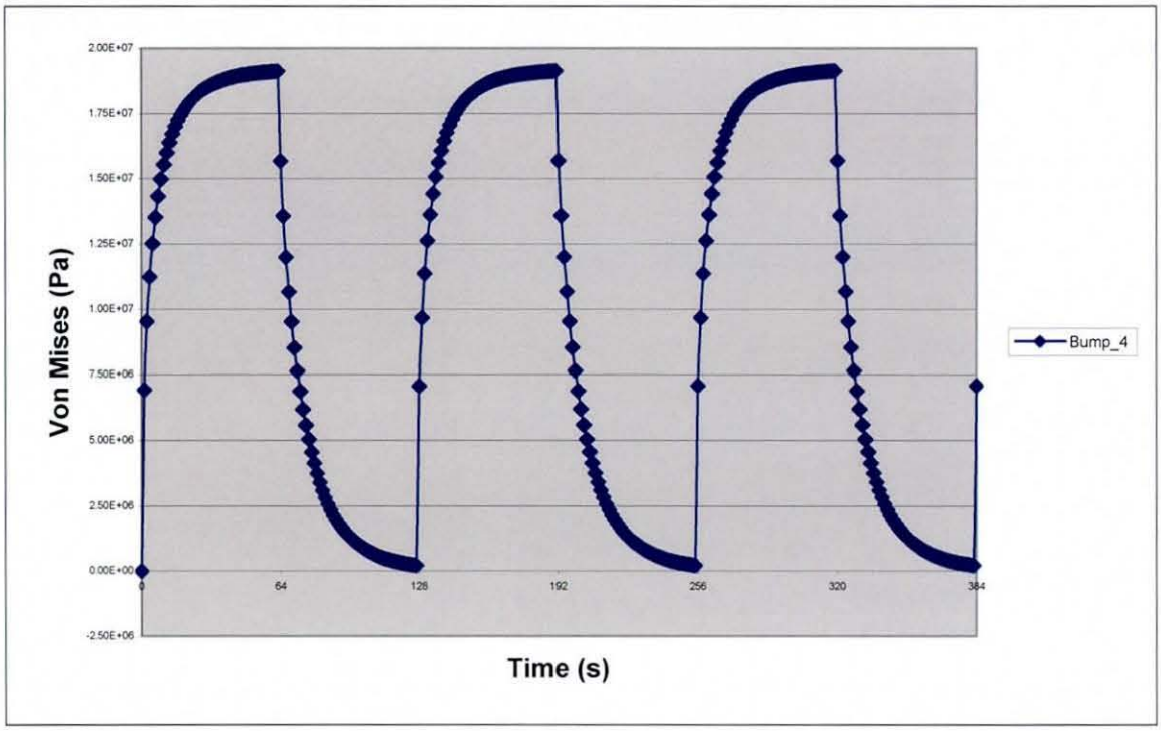


Figure 7.3.1.1c: *Linear equivalent Von Mises stress plot as a function of time. Note the near horizontalness of the curve at the end of the ON cycle in comparison with the steeper slope at the end of the OFF cycle; this was attributed to the temperature dependent material properties.*

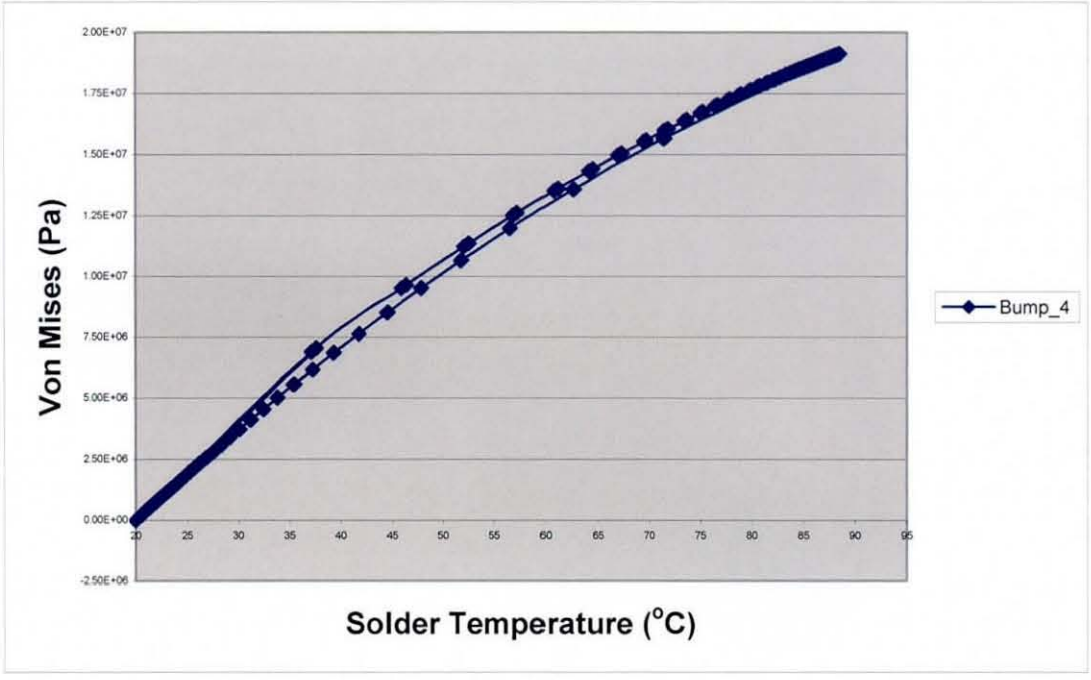


Figure 7.3.1.1d. Linear equivalent Von Mises stress plot as a function of temperature. Note the cyclic nature of the graph shown. The top curve is the ON cycle that exhibits slightly greater stresses for the same temperature compared with the OFF cycle.

7.3.1.2: The Influence of Non-Linear Material Properties

The solder yield stress for any particular FE model set-up (e.g. specific to convection values & power levels, the substrate used etc.) was implemented by initially running the model as a linear stress analysis (i.e. no plastic strain) and the node exhibiting the maximum stress had its history plot out. Figure 7.3.1.2 shows the maximum stress in a solder joint at a given temperature and the temperature dependent yield stress properties are also shown (2,3). It can be seen that for values below 85°C approx there will be no plastic deformation in the joint though for values greater than 85°C then the plastic deformation becomes apparent.

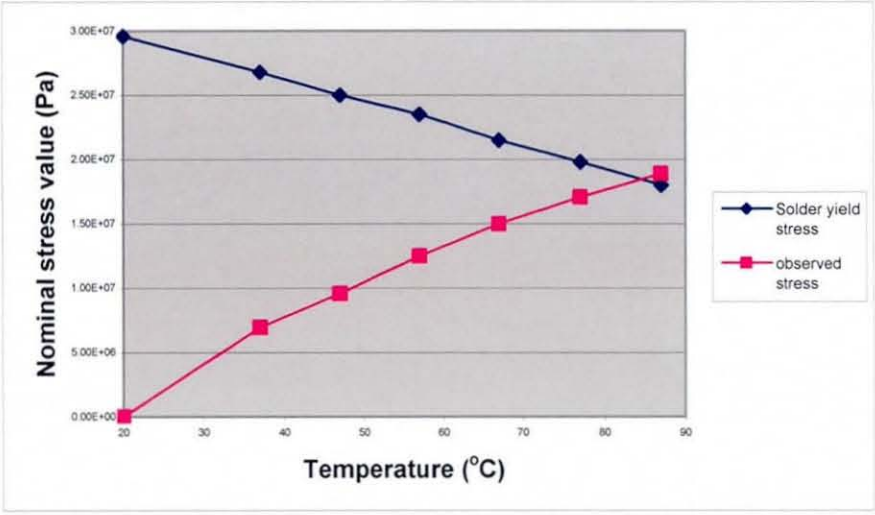


Figure 7.3.1.2: *Plot of observed stress in solder joint and solder yield stress against temperature.*

From the previous thermal modelling and profiling of the MCMs on FR4, the temperature of all the solder joints was known to reach approximately 90°C at the end of the ON cycle for 1.4W power input. According to material values in section 5.4, the yield stress of 63/37SnPb eutectic solder at this temperature is 17MPa and super plastic deformation was implemented in the model. Figures 7.3.1.2a, b, and c show joint 4 viewed from AA (from figure 7.2.3a_i), the complementary view and the sectioned view respectively. It can be seen that the stress levels are now uniform in the joint with a maximum value of approximately 17MPa detected in the joint at any part of the assembly. The stress values obtained at the central region are in good agreement with the linear analysis however the peak values that were detected in the upper and lower regions have substantially reduced. The plastic strain plots of bump 4

are shown in figures 7.3.1.2d and e for the inside and outside views. They show little resemblance to the non-linear stress plots, though they do show an agreement with the initial linear plots, identifying the greatest plastic strain at the parts that were shown to show the highest stress. A plastic strain value of $1.9x10^{-4}$ was recorded at the end of the first ON part of the initial power cycle.

Figure 7.3.1.2f shows the non-linear stress against time for three nodes selected from bump 4 of the original model. The node that showed the greatest plastic strain was chosen (max), the subsequent node showed less strain (int) and the third showed very little plastic strain (min). **Max node: -** From the profile obtained, it can be seen that the same stress value is reached on all the cycles. Observing the first ON part of the cycle, it can be seen that the profile of the curve changes after one second at approximately 17MPa, presumably due to the yield stress of the 63/37 SnPb solder being breeched. Note the almost linear increase of the stress. During the OFF cycle, it can be seen that the stress value decreases as would be expected initially, though after some critical point, the stress values increase during the OFF cycle. This phenomenon was attributed to the accumulated deformation altering the stress free profile of the joint. As the cycle continues, the minimum stress encountered during the OFF cycle increases incrementally. **Int node**: - has similar characteristics to those of the "max node" during the ON cycle as shown shown. However when the minimum stress value is reached during the OFF cycle, it remains constant as opposed to the increase shown from the "max node". Also the minimum value reached can be seen to rise during subsequent cycles. Min node: - During the OFF cycle, the minimum value can be seen to be the same for all the increments.

Figure 7.3.1.2g shows the variation of plastic strain with time for the nodes max, int and min. There was no plastic strain detected at the very start of the ON cycle, as the stress values reached were not sufficient to cause plastic strain. It can be seen that the greatest amount of plastic deformation occurs in the first increment and the plastic deformation from subsequent increments is reduced sequentially before the plastic deformation effectively ceases after an appropriate number of cycles. It can be seen that all the plastic deformation occurs towards the end of the cycle when the solder is warm and the stress endured is greater than the yield stress. The Max and Int nodes can clearly be seen to be accumulating plastic strain whilst the min node appears to

have only endured plastic deformation on the first increment and has then ceased to accumulate any further plastic strain.

The graphs appear to complement each other as there is no plastic strain at the very start of the first cycle, where the solder exhibits elastic properties, however as plastic strain is detected, the characteristics of the stress graph change accordingly. This process is repeated for all the plastic strain increments.

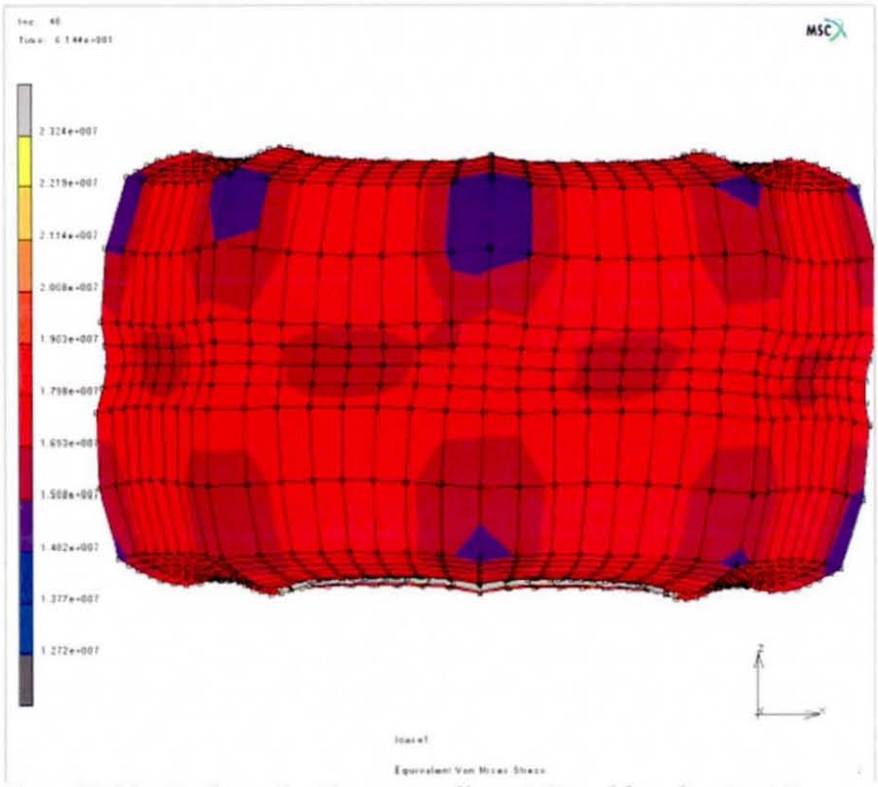


Figure 7.3.1.2a: *Non linear Von Mises stress of bump 4 (viewed from direction AA)*

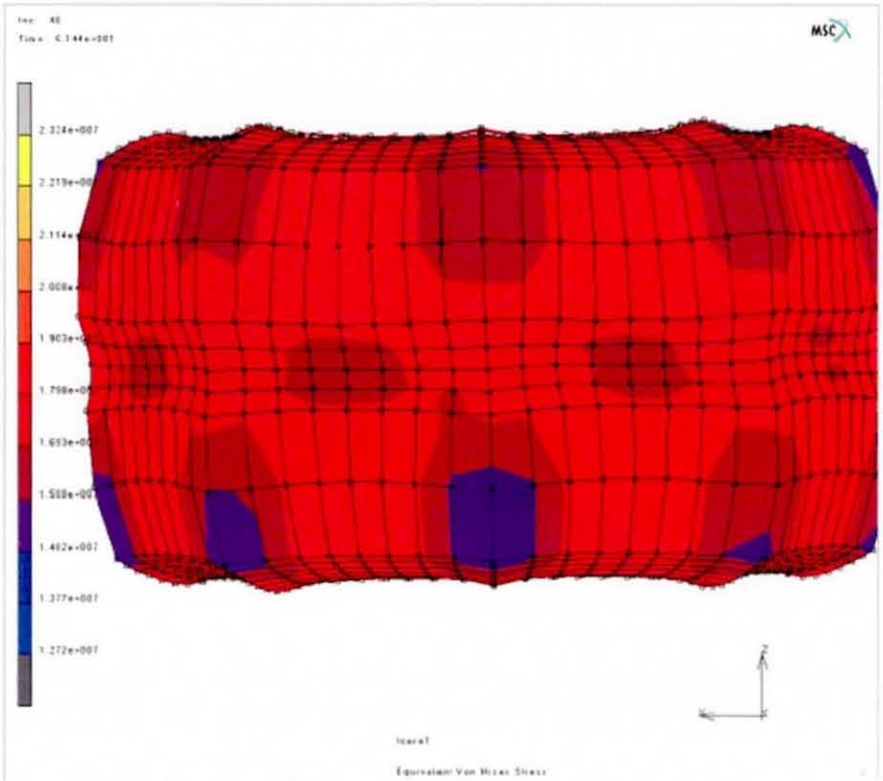


Figure 7.3.1.2b: *Non linear Von Mises stress of bump 4 (viewed towards AA)*

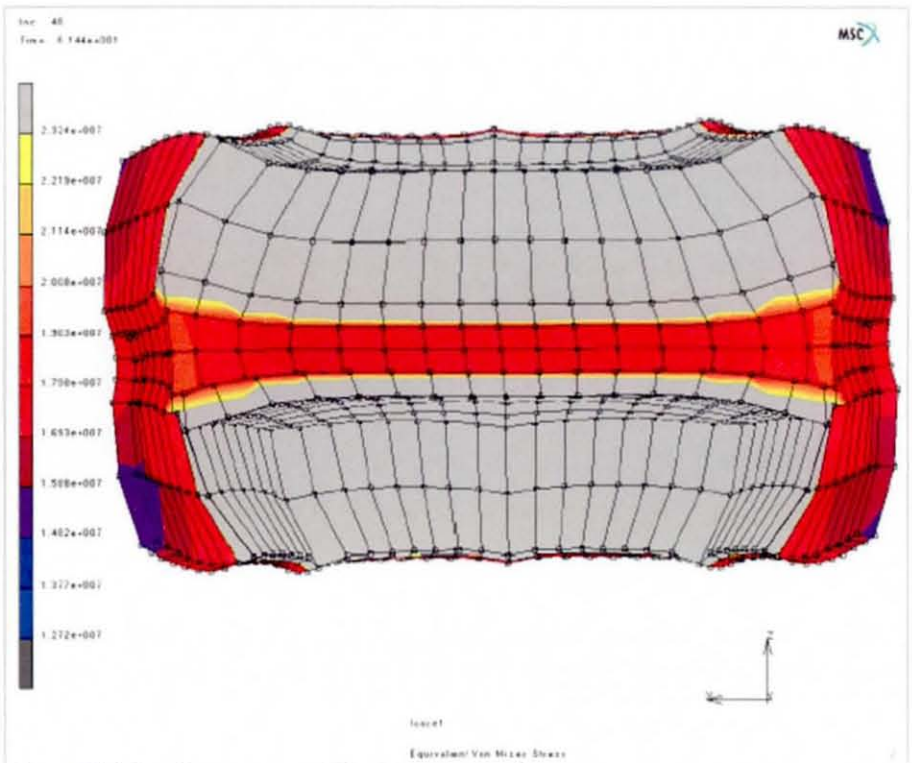


Figure 7.3.1.2c: *Cross-section of Von Mises stress of bump 4*

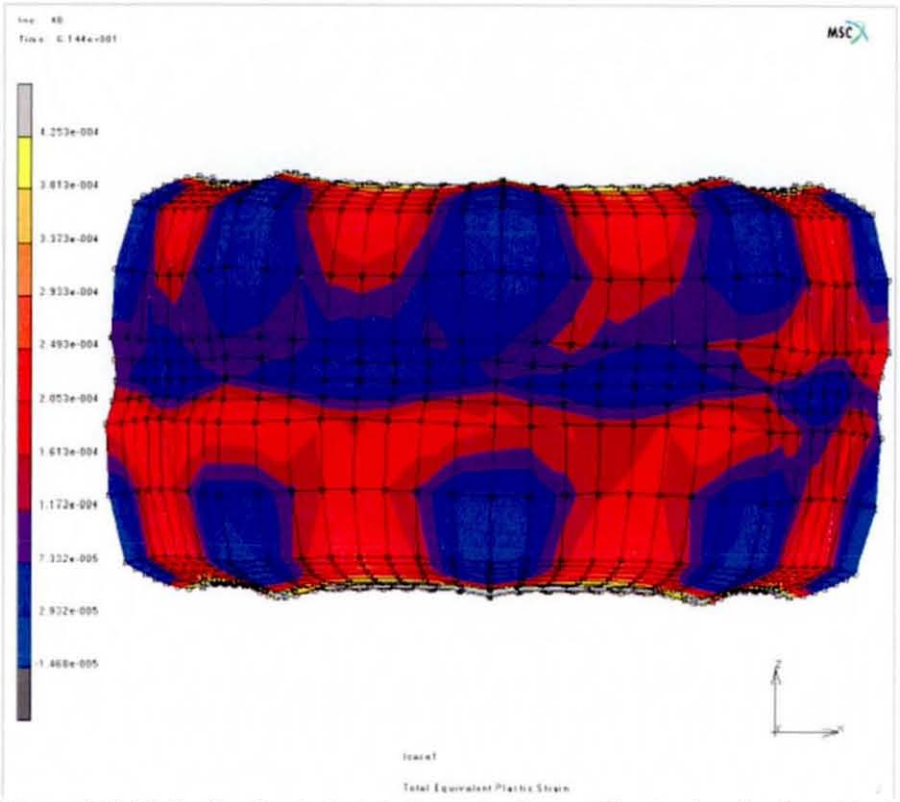


Figure 7.3.1.2d: *Profile of equivalent plastic strain of bump 4 The view is taken from AA taken using model R5*

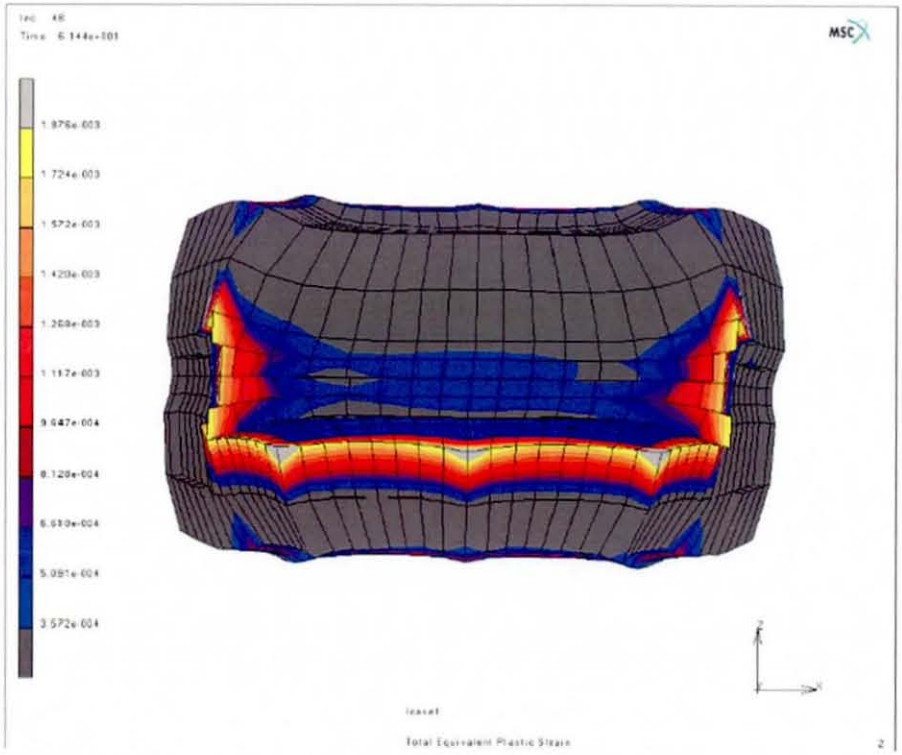


Figure 7.3.1.2e: *Sectioned view of bump 4 showing central plastic strain.*

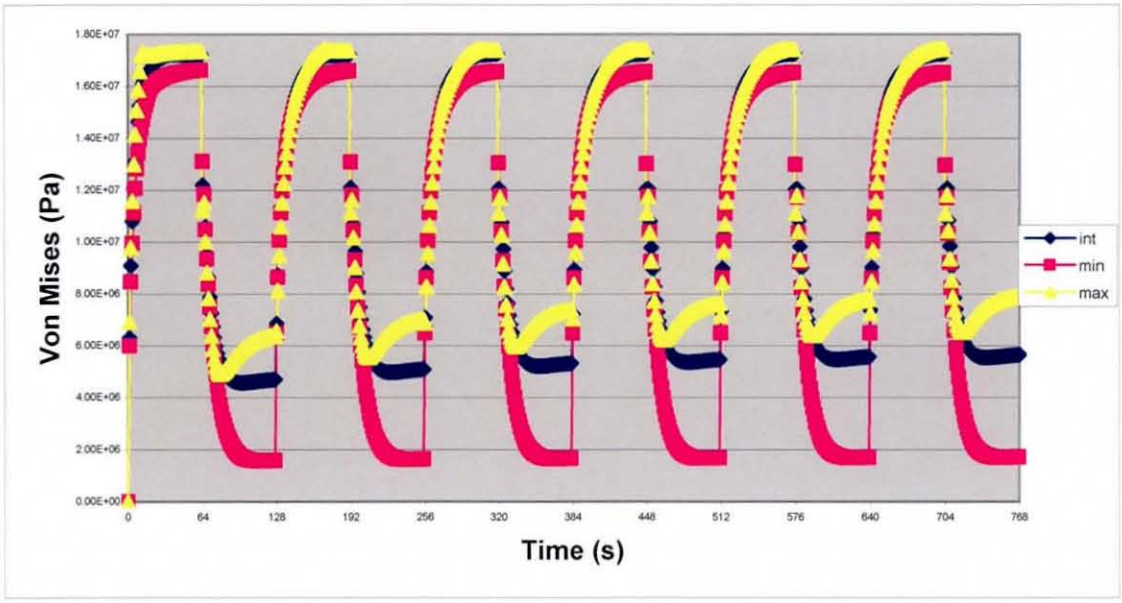


Figure 7.3.1.2f: *Non-linear Von Mises stress graphs of selected nodes from bump 4. The stress is expected to be a combination of shear due to the temperature difference between the heater and carrier chips, tensile due to the local CTE effects and the rise on the maximum node is the consequence of residual plastic deformation.*

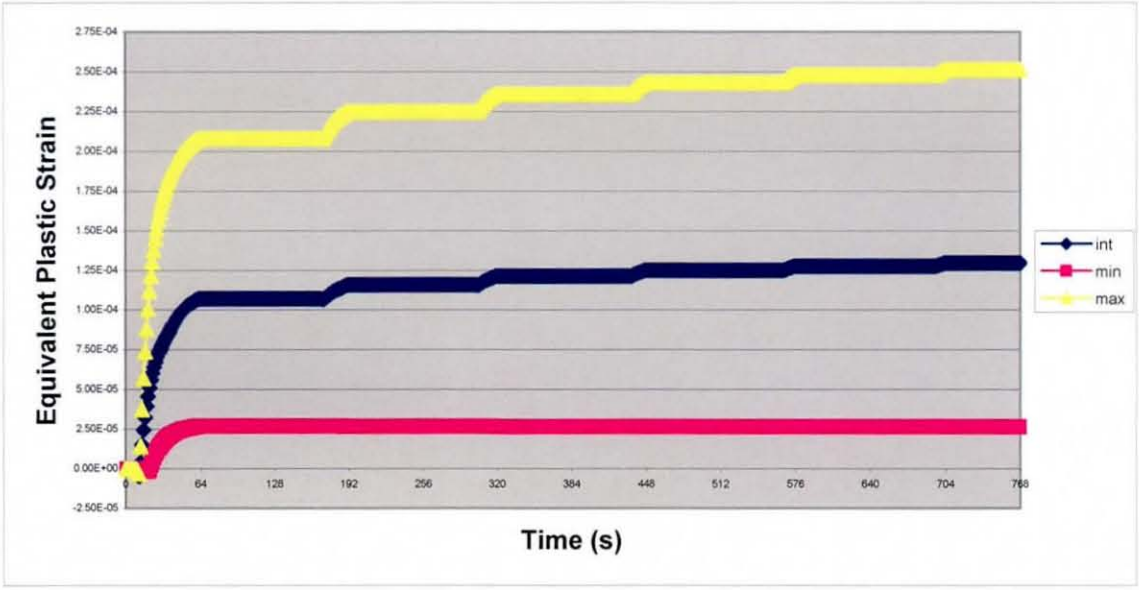


Figure 7.3.1.2g: *Equivalent plastic strain graph of the selected nodes from bump 4*

The Effect of Reducing the Power to 1.2 Watts

Von Mises Stress

Figure 7.3.1.2b shows the simulation results for the solder joints from the MCM mounted on FR4 power cycled with 1.2 Watts. It can be seen that the overall stress distribution in the joints is similar to that observed for the 1.4W cycled condition, though the average stress values reached are reduced accordingly to 14-15 MPa. The scalar in which the power level was reduced by is the same scalar in which the stress profile is reduced.

Equivalent Plastic Strain

When the MCM on FR4 substrate was power cycled with 1.2 Watts, the temperature of the solder joint was known to be 80°C. According to the material properties, the yield stress of the 63/37 SnPb solder at 80°C is approximately 19MPa (2,3) and this value was used in the model. As the maximum stress obtained in the simulation was 16MPa there was no plastic strain observed.

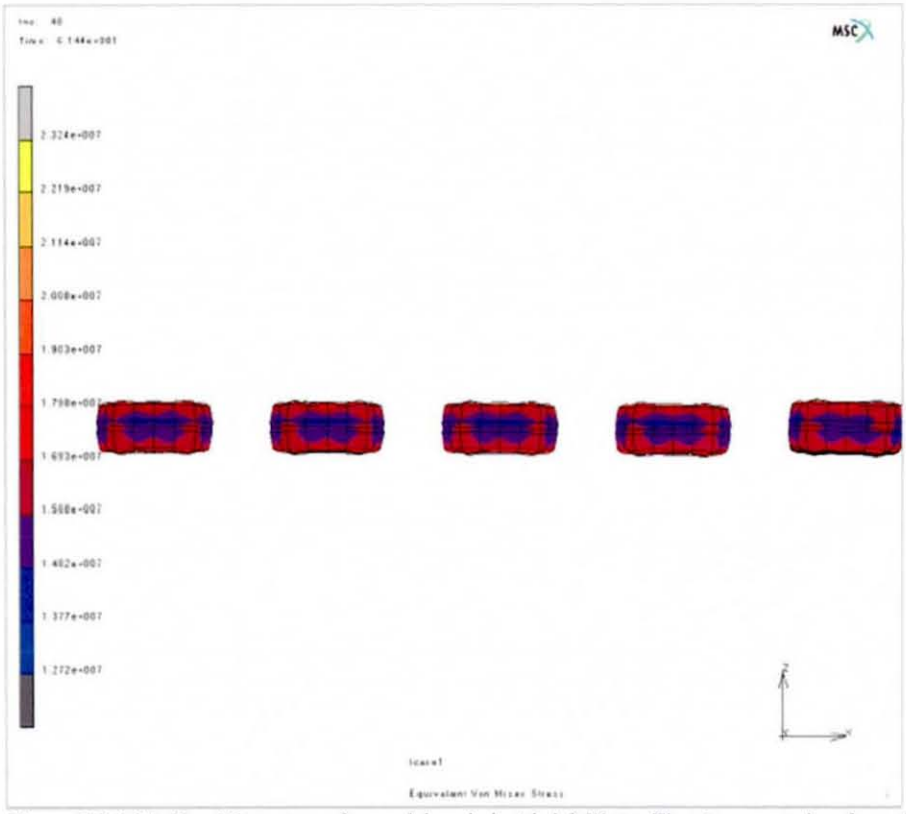


Figure 7.3.1.2d: *Von Mises stress for model cycled with 1.2 Watts. The view was taken from AA and model R1 was used*

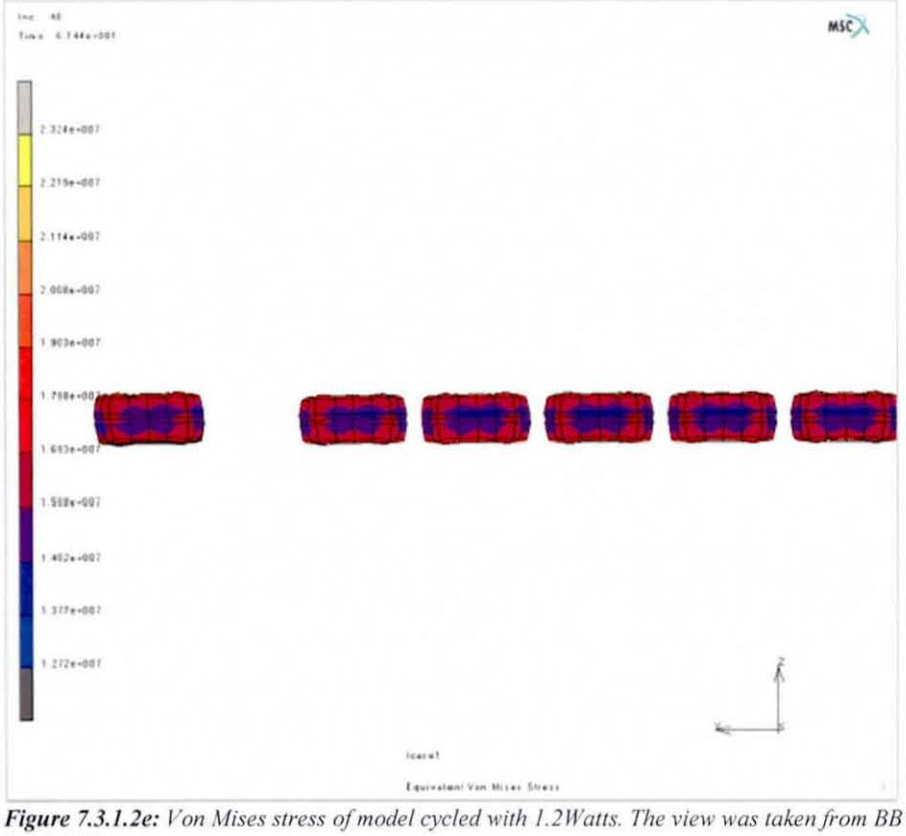


Figure 7.3.1.2e: *Von Mises stress of model cycled with 1.2Watts. The view was taken from BB (figure 7.2.3ai) (model R1)*

7.3.2: Assemblies on Copper Substrates

Stress 1.4 Watts

The distributions of the resultant stresses following the 1.4W power simulation are shown in figures 7.3.2a. The stress values over the joint can be clearly seen to vary by over 1MPa, with the profiles obtained exhibiting values between 6.3 and 7.5MPa. While all the solder joints exhibited similar stress values and profiles, from figure 7.3.2a, it can be seen that the location of the maximum/minimum stress on any join is very dependent on the location of the joint from the neutral point. From figure 7.3.2a, on joint 9 the minimum stress value can be seen on the centre of the joint, while on joint 5 the maximum has moved to the side of the joint consistent with the implications of model R1 in section 7.2.3.

Figure 7.3.2c shows the stress distribution of joint 4 from the inside and outside. The values of the stress at the lower region are significantly greater than those of the upper region and on the outside plot the upper region is greater. This was similar to the characteristics exhibited by the joints on FR4 but the effect is more profound when the copper substrate was used.

Plastic Strain

The temperature of the solder at the peak of the ON cycle was known to be 45°C. The yield stress of the solder at this temperature was 25MPa. The maximum stress detected in the solder joint was 8MPa; therefore no plastic strains were recorded.

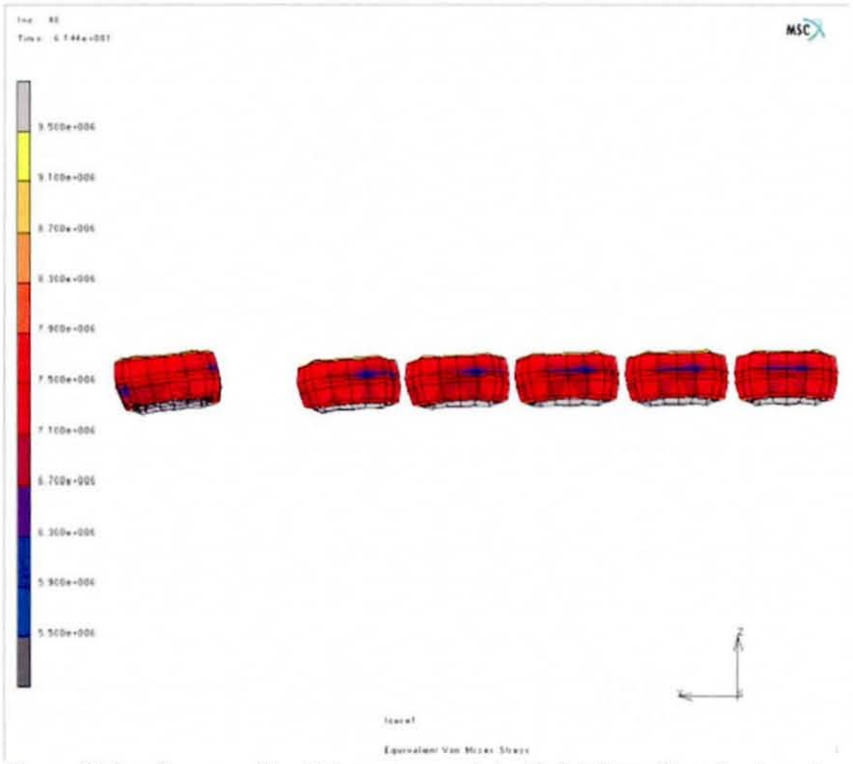


Figure 7.3.2a: *Stress profile of joints power cycled with 1.4 Watts. Note the clear change in the orientation of the minimum stress level as the bump position varies depending on the position of the bump with respect to the neutral point. The image was taken from BB (7.2.3ai) with model R1 used*

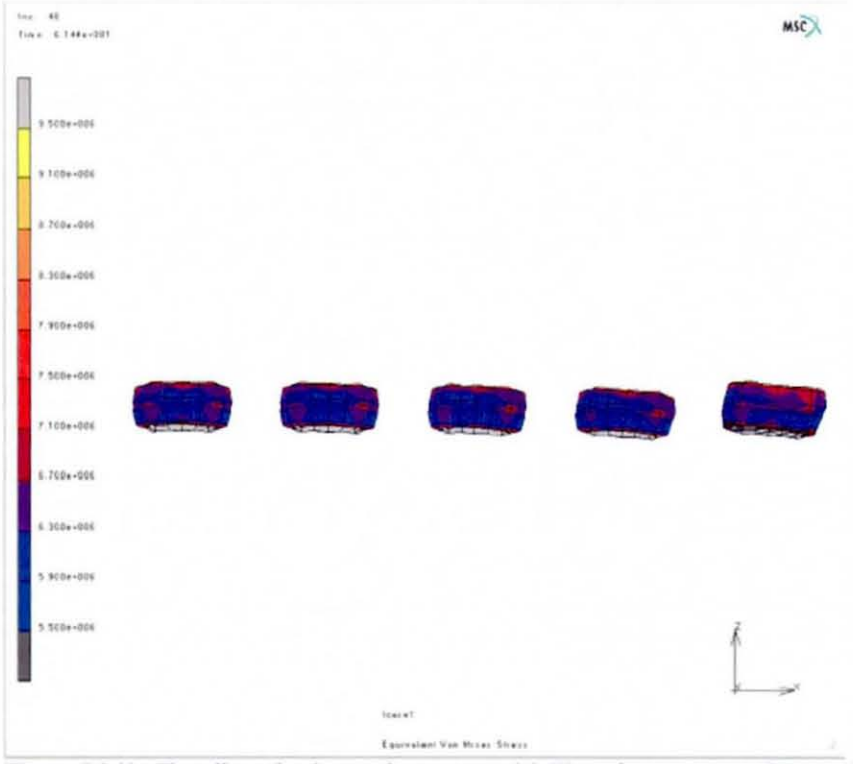


Figure 7.3.2b: *The effect of reducing the power to 1.2 Watts for a copper substrate. The observed stress levels have reduced proportionally with the power reduction. The view is taken from AA (figure 7.2.3ai) and the model used was R1*

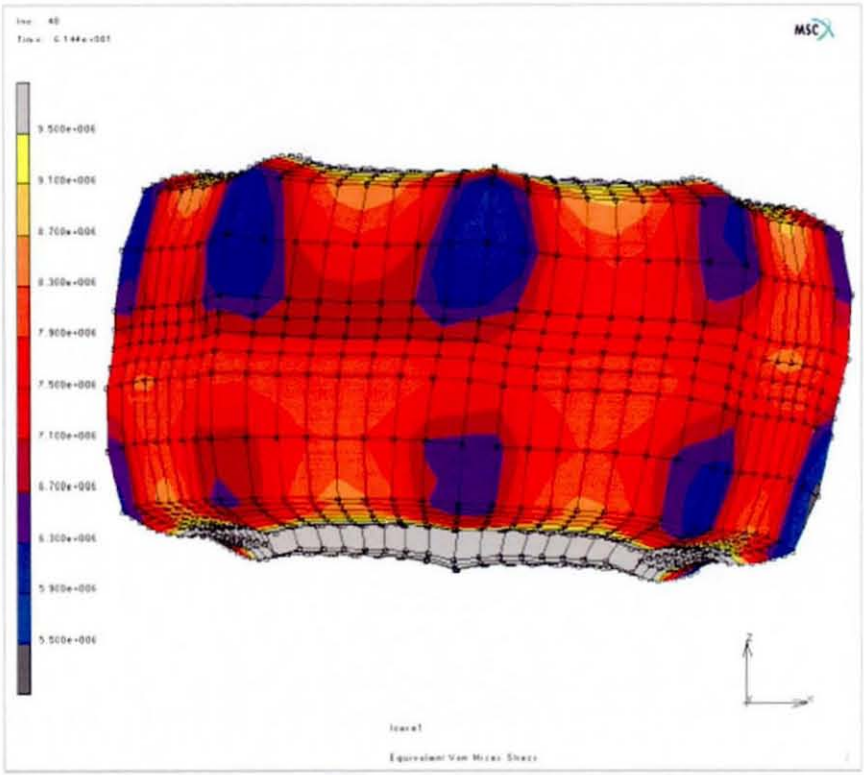


Figure 7.3.2c$_{i}$: *Von Mises stress of bump 4. Note the lower levels of stress in the upper region compared with the lower region. The rippling effect and stress unusual stress contours were due to the poor approximation of the solder joint (the octagonal shape) The view was taken from AA (from figure 7.2.3ai) and the model R5 was used.*

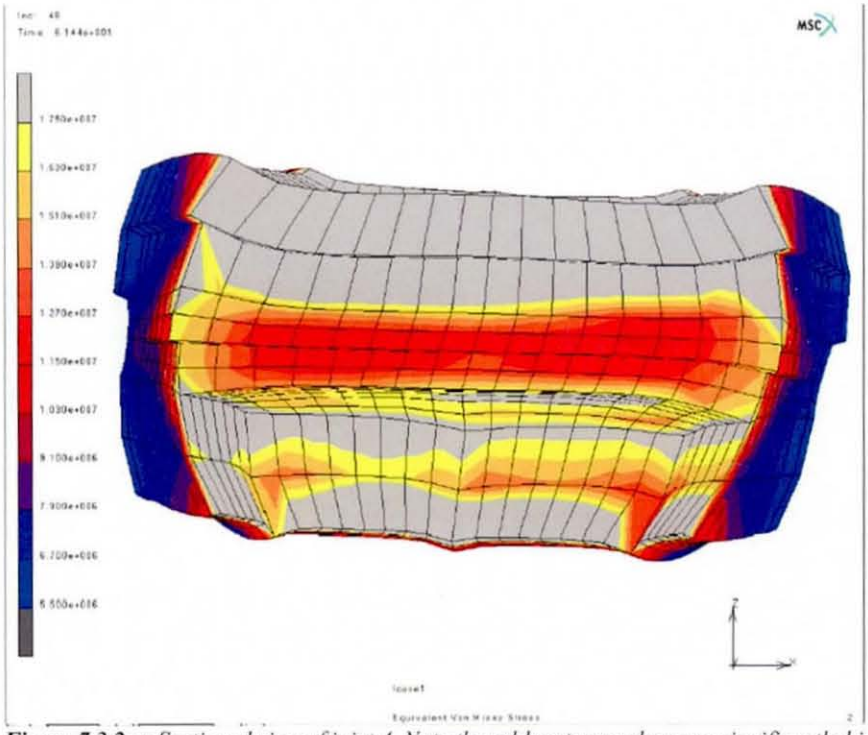


Figure 7.3.2c$_{ii}$: *Sectioned view of joint 4. Note the solder stress values are significantly higher in the central region, this was attributed to the small amount of solder present and the inevitable CTE mismatche between the UBM and the solder*

7.3.3: Influence of Mesh Density

The stress distribution of the solder joints using high and low density meshes are shown for the surface (linear properties) and for the section views (linear and non-linear) in figures 7.3.3a, b and c respectively. In figure 7.3.3a, it can be seen that there is general good agreement with the maximum value reached; however the high-density solder bump shows an unusual stress distribution on the surface with high and low regions, while low-density model exhibits more uniform stress contours. The varied stress patterns on the high density were attributed to the nature of the model: The cylindrical profile of the solder joints was approximated by the octagonal shape as described in section 5.4: When the FE model or contact bodies were subdivided it would have been desirable to manually reposition the nodes to obtain a body that was more circular; however time constraints dictated that this method was impractical to implement. Therefore the shape of all joints, including those that were densely meshed were modelled in the original octagonal shape. The inconsistencies of the stress distribution on the high-density mesh were attributed to the poor positioning of the nodes; it is well known that sharp corners act as stress raisers and this coupled with the higher mesh density resulted in stress concentrations at these corners. In reality it is known that the solder joint is cylindrical/spherical so such stress raisers would not exist.

From the sectioned views, it can be seen that both the high and low density elements show a much greater stress value in the centre of the joint; the lower mesh density predicts higher central stresses however this was attributed to the low mesh density in this area. Moreover, the stress values predicted are well above the yield stress of solder within the temperature ranges considered, so a better representation of the actual scenario was found when plastic deformation is considered as shown.

Figures 7.3.3d and e show the solder joint 4 taken from model R5 with FR4 powered with 1.4 Watts (figure 7.3.1.1) as well as the view from the opposite side. Comparing the values it can be seen that there is no significant difference in the stress distributions or max/min values.

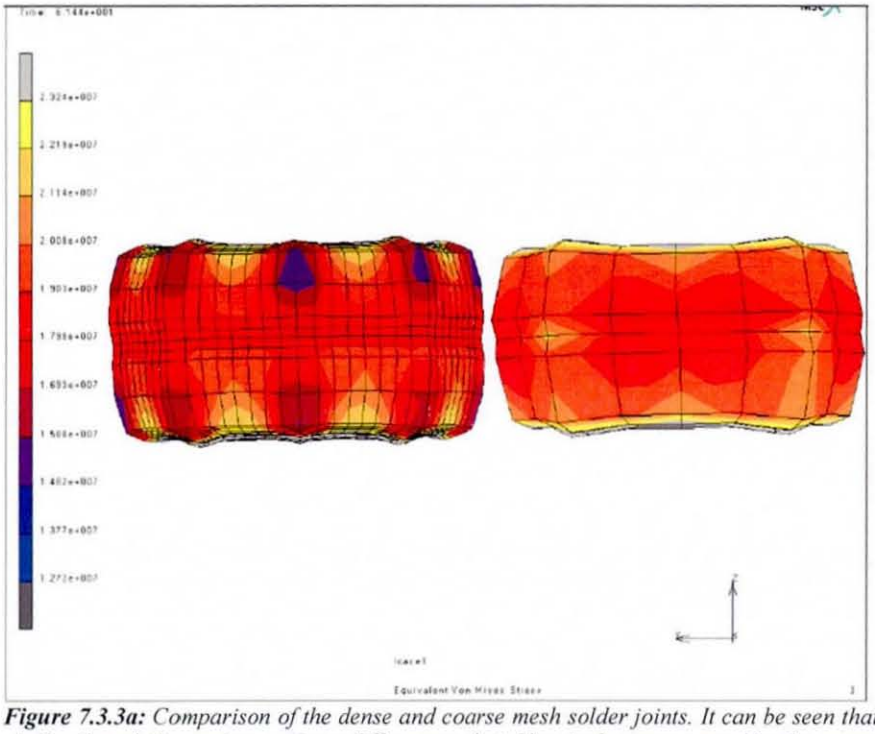


Figure 7.3.3a: *Comparison of the dense and coarse mesh solder joints. It can be seen that the nominal values are similar though the contour patterns differ somewhat. The ripples were caused by the octagonal approximation.*

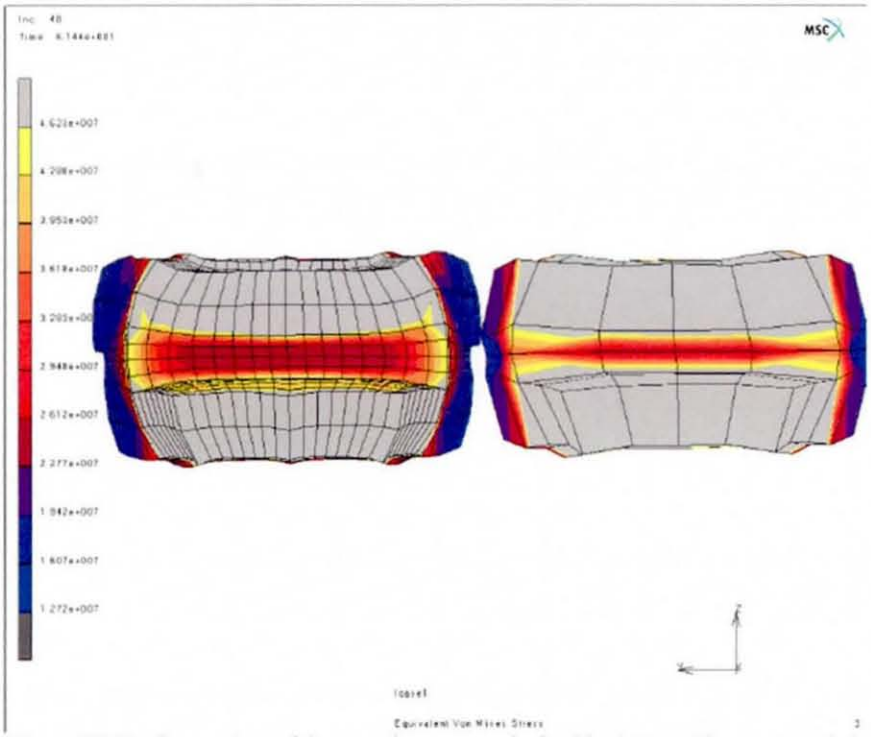


Figure 7.3.3b: *Comparison of dense and coarse meshed solder joints with a sectioned view. Both models correctly identify high stress concentrations around the solder UBM interface.*

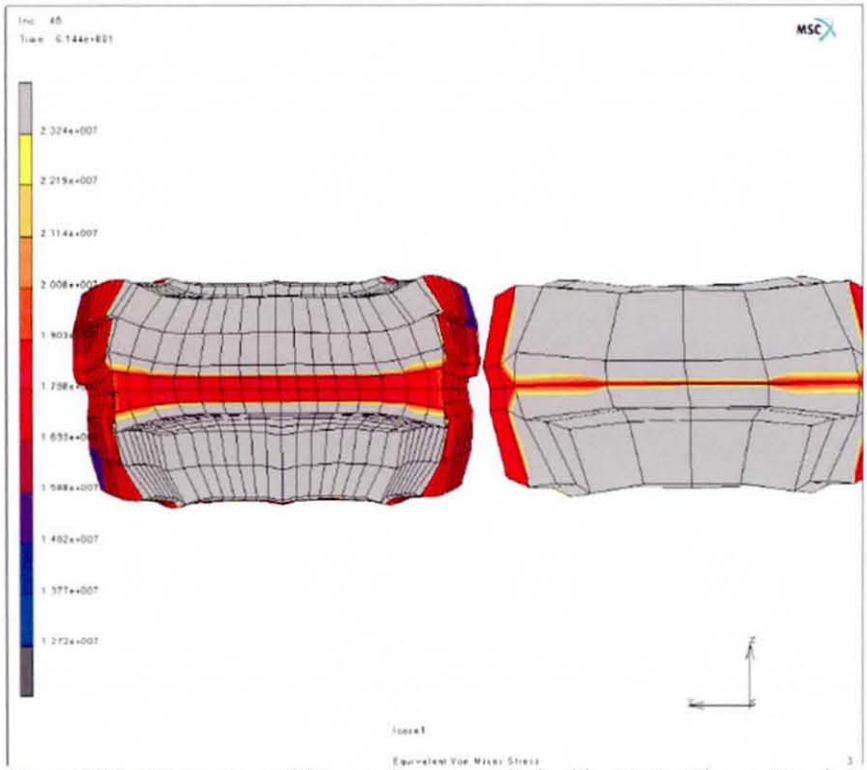


Figure 7.3.3c: *Comparison of dense and coarse meshed solder joints with a sectioned view for plastic deformation . Both models correctly identify high stress concentrations around the solder UBM interface.*

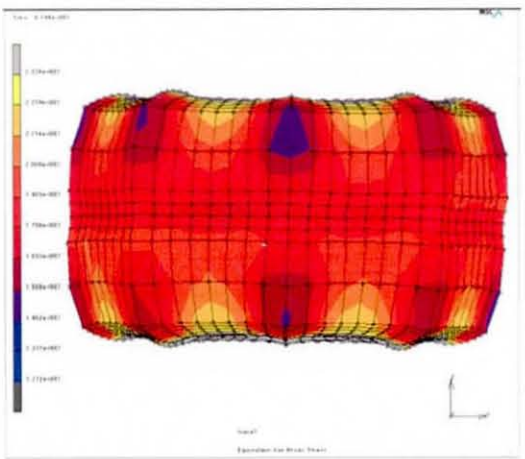


Figure 7.3.3d: *Linear Von Mises stress of bump 4. View is taken from AA. As shown in figure 7.3.1.1a. The ripple effect is caused by the octagonal approximation of the solder joint shape.*

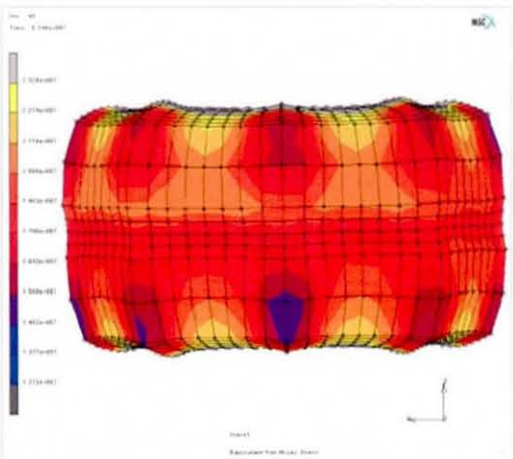


Figure 7.3.3e: *Complimentary view of figure 7.3.3d. The ripple effect is caused by the octagonal approximation of the solder bump.*

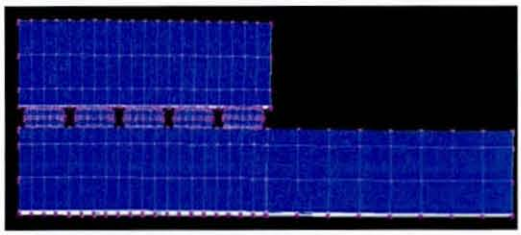
7.4: Discussion

The previous thermal studies have identified that for the different MCM assembly configurations considered, there was always the same temperature difference between the heater and carrier chip at the end of the ON part of the cycle. This was the observed situation regardless of whether FR4 or copper was used as the substrate (when 1.4W was used in power cycling, both FE models exhibited a temperature difference of 8°C between the chosen heater and carrier nodes). Therefore, if the shear stress caused by the CTE mismatch was the dominant factor then similar stress values were anticipated in the solder joints for both situations, despite the different final overall assembly temperatures.

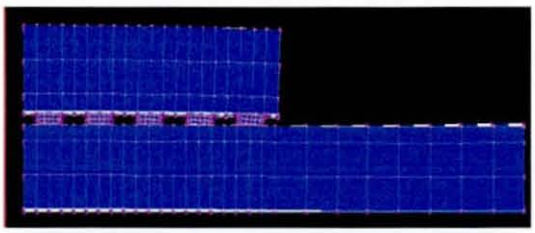
7.4.1: FR4 Vs Copper

The stress distributions and magnitudes within the solder joints were markedly different depending on whether the MCM was mounted on FR4 or copper substrate. The assemblies on the copper substrate exhibited much lower stress than those attached to FR4, though the temperature difference between the heater and the carrier chips was the same for both cases. The FR4 simulation showed an even stress distribution across all the solder joints while for the copper substrate models, a more obvious stress gradient was observed in the solder. From figures 7.4.1a, it can be seen that from the outside of the joint, the stress is large at the top of the joint but is low towards the bottom. Furthermore, when the model with the copper substrate was simulated, the orientation of the stress profile (the location of the maximum stress on any particular joint) was more dependent on the position of the joint with respect to the neutral point.

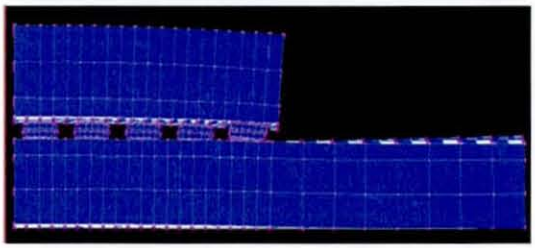
Figures 7.4.1a, b and c show the influence the substrate material properties had on the deformation of the MCM as a whole. It can be seen that shear effects caused by the different temperatures of the heater and carrier chip are more apparent when copper was used than with FR4. This was attributed to the fact that the temperature difference between the heater and carrier chip is a greater percentage of the overall temperature profile when copper was used as the substrate as opposed to FR4.



7.4.1a: MCM on FR4
Deformation magnified
2000 times



7.4.1b: MCM on copper
Deformation magnified 2000 times



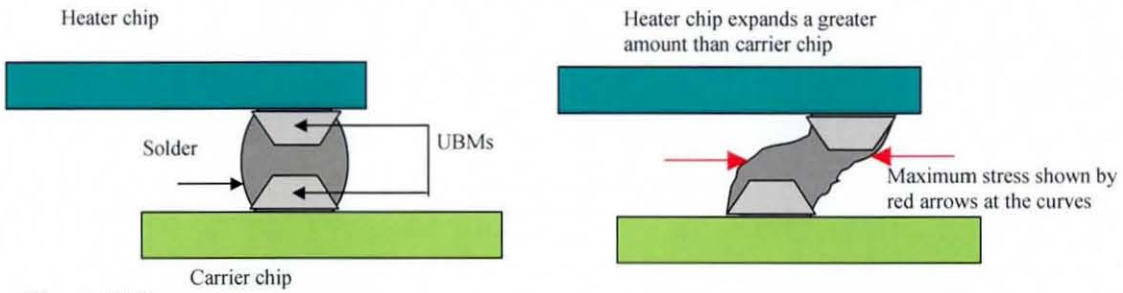
7.4.1c: MCM on copper Deformation magnified 5000 times

Figure 7.4.1:
Plots of the deformation effects the substrate had on the MCM

7.4.2: Generic Stress Characteristics at the Joints

Figure 7.3.2, shows high stress concentrations where the solder meets the UBM, indicating that this particular area is likely to exhibit a failure. However in "real" solder joints there is known to be a layer of intermetallics between the solder and the UBM, which are likely to exhibit different material properties to those of solder and Ni. Therefore the high stress concentrations depicted in these areas may not be a true representation of the characteristics of a "real" device; as they were omitted from the model as described in 7.3.

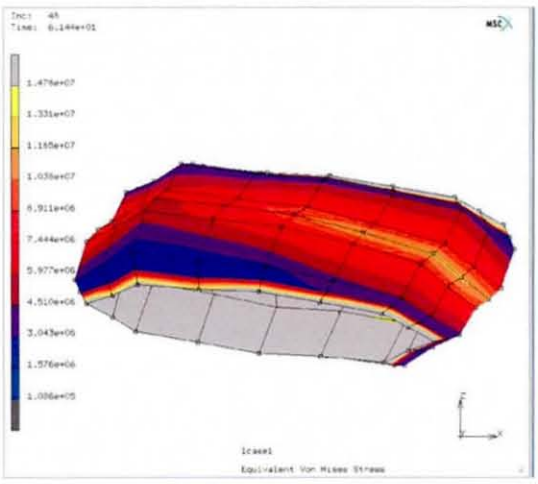
When examining the stress effects of the copper joints, a stress gradient was very prominent. From view AA, it can be seen that the stresses are marginally greater on the lower region of the solder joint, conversely for the outside plot greater stress is exhibited within the upper region. Figure 7.4.2 shows the anticipated characteristics of strains in the solder joint due to the CTE mismatch that justifies the particular profile exhibited, and is similar to the distributions predicted from other studies.



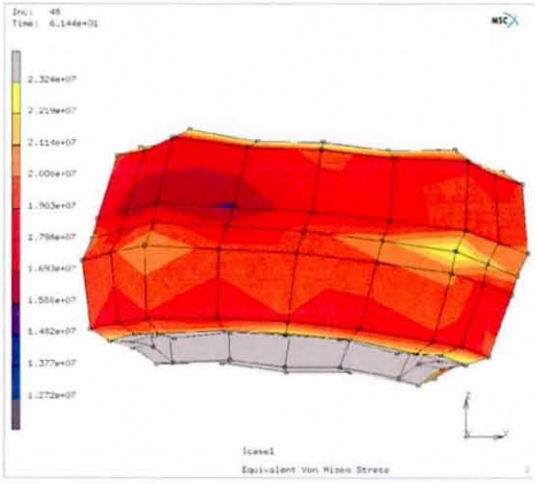
Figures 7.4.2:
Illustration of the shear influences on the solder joint

7.4.3: Influence of Local CTE Mismatch

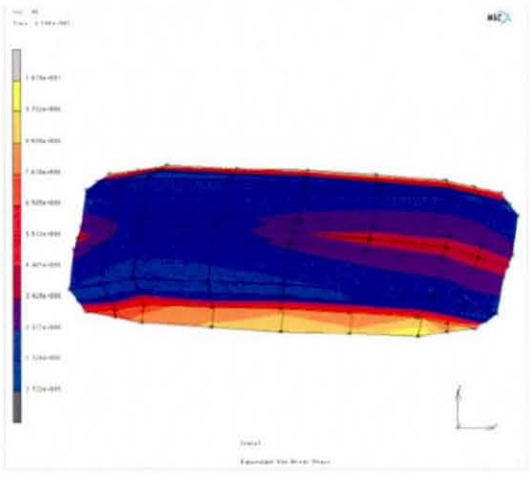
The CTE mismatch local to the interconnection joints (i.e. the CTE mismatch between the solder and the UBM) appears to be an important factor in the stress generation for these models, potentially justifying the difference in the recorded stress levels. As the solder and UBM are heated up, their different CTEs may generate internal stresses in an individual joint, irrespective of the shear effects caused by the heater and carrier chips. To investigate the influence of this, further simulations of the MCMs on the two substrates were performed; in this case the CTE rates of the solder, the UBM and the aluminium pads were all set to that of silicon. Figures 7.4.3b, i, ii, iii and iv show a stress plot of joint 4 viewed from direction AA (figure 7.2.3ai) shown for both when the CTE levels are correct and when all the CTEs are set to that of silicon. It can be seen that the stress values were reduced significantly for the case with FR4 and less so for the case of a copper substrate. This shows that the local CTE mismatch between the solder and nickel contributes significantly to the accumulated stress strain rates in the solder joints. From the deformation plots (figure 7.4.3bi), the shear influence of the temperature difference between the heater and the carrier chip can be seen where the misalignment between the upper and lower region is seen as well as the angle of the bottom part of the joint. The shear effect appears more prominent than when the correct CTE values are used in the joint as shown in 7.4.3.bii. Observing the stress values, it can be seen that the overall and maximum stress values have been reduced significantly, and high stress concentrations can be found at specific locations in the bump as opposed to the correctly modeled bump that exhibits a more uniform stress. Comparing the copper CTE matched models, it can be seen that the local CTE match has had a similar effect though to a lesser extent. This was attributed to the lower temperatures reached.



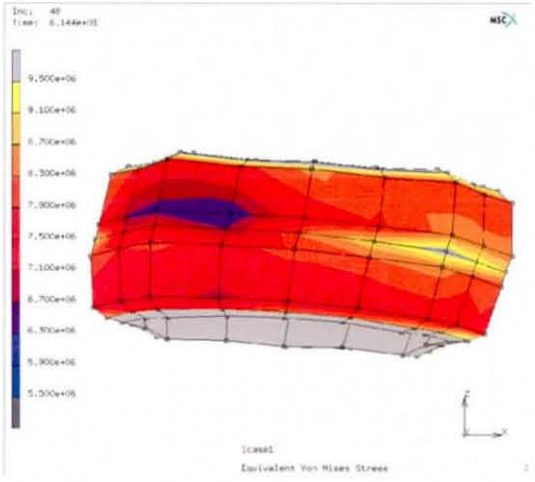
FR4 CTE matched



FR4 natural



Copper CTE matched



Copper natural

Figures 7.4.3b i, ii, iii, & iv. *Comparison of the CTE matched solder joints with the original models. The deformation is magnified 2000 times.*

7.4.4: Plastic Strain Profiles

The graph characterising the plastic strain was markedly different from the traditional hysteresis loops obtained for traditional thermal cycling. During conventional thermal cycling, large temperature amplitudes are often implemented (typically -50 to 125°C for example); these are sufficient to cause significant plastic strains during both the temperature increase and decrease. In addition, the long "hold" or "dwell" times are sufficient to allow adequate stress relaxation to occur such that the accumulated stresses are dissipated. There was no such dwell time implemented in the given experiments, therefore during the OFF part of the cycle the stresses are simply released as opposed to reversed.

As the yield stress of 63/37 SnPb solder is so sensitive to temperature, the amount of plastic deformation for any given power level is heavily dependent on the initial ambient temperature assigned in the model. Table 7.4.4 shows the plastic strain obtained from different ambient temperatures. The 1.4W FR4 model was simulated with an ambient temperature of 20°C and a maximum total plastic strain value of 2.078×10^{-4} was predicted. However, if the ambient temperature is reduced to 10°C then the yield stress of solder at 80°C (the maximum solder temperature) is 20MPa and there is no plastic deformation in the model. Conversely if the ambient temperature is increased to 30°C then the solder reaches 100°C and the yield stress value of solder at this temperature is reduced to 14MPa and the plastic strain increases. Likewise, for the 1.2W under the ambient temperature of 20°C, as implemented in the model is not sufficient to cause plastic strain though when the temperature is raised to 30°C, under the same conditions significant plastic strain can be obtained.

Power level	***Temperature***		
	30°C	**20°C**	**10°C**
1.2W	1.962×10^{-4}	nil	nil
1.4W	5.109×10^{-4}	2.078×10^{-4}	nil

Table 7.4.4: *Characterisation of the equivalent plastic deformation for the MCM on FR4 depending on the ambient temperature*

7.4.5: Creep Strain

In conventional thermal cycling, creep strain has been identified as a significant factor regarding accumulated non-elastic strains; however in these experiments with the short cycle times used, even under the harshest test conditions (1.4W, MCM on FR4) little creep was expected to occur. This is in agreement with the anticipation of Neal et al (4) who specifically performed fatigue cycle tests at 0.5Hz in order to minimize the creep strain effects. Grossmann (5) thermally cycled assemblies using dwell times of 30 and o minutes; after 4000 cycles the 30-minute dwell samples showed significant crack propagation while for the 0min trial only very small cracks were observed. A common law used to predict the accumulated creep strain in solder joints is the sinh law (5) or

$$\varepsilon_c = A\left(\frac{E}{T}\right)\left[\sinh\left(\frac{\alpha\sigma_{obs}}{E}\right)\right]^n \exp\left(\frac{-Q}{kT}\right)$$

Key
A=16.7x10^{-6} (63/37SnPb Eutectic constant)
T= Temperature (°K) = 363
α=material constant (866 for SnPb Eutectic)
E= Young's Modulus (21.231GPa for solder at 363°K)
σ_{obs}= observed stress
Q= Activation energy (0.548eV)
k= Boltzman's constant (8.617x10^{-5}eV)

In order to estimate the creep rate for the MCMs, values were substituted. Solving for MCM on FR4 powered with 1.4W

$$\varepsilon_c = 16.7x10^{-6}\left(\frac{21.231\times10^9}{363}\right)\left[\sinh\left(\frac{866\times(17\times10^6)}{21.231\times10^9}\right)\right]^{3.3}\exp\left(\frac{-0.548eV}{(8.617\times10^{-5}eV)\times363}\right)$$

$= 9.3217\text{x}10^{-6}$ /s

so the creep rate can be seen to be very low for a one minute cycle; much lower than the plastic strain rates observed. However the applicability of this particular equation is questionable as the sinh law predicts the creep strain rate during the secondary or "steady state" creep stage. With the short cycle times used in this experiment, it is debatable on whether the solder would have actually reached the secondary stage when this model for creep is valid (6). Of the previous published work most is based on thermal cycling with significant dwell times so the assembly can be assumed to be at secondary creep. From the previous thermal cycle profiles the assembly appears to have only just reached steady state so is likely to still be in the primary or transient creep stage. Unfortunately most modeling work seems to neglect the creep

accumulated in this stage as steady state creep assumed to be the dominant mode. Consequently, there was no suitable creep model found that could confidently model the creep strain in this case.

7.4.6: Lifetime Predictions

From the obtained mechanical data, it was possible to obtain a lifetime prediction for the joints based on the stress values obtained. There were no creep strains and owing to the an-isothermal temperature distribution as well as the very small plastic strains obtained in one example the only model that was applicable was the modified Coffin Manson equation or

$$\frac{\Delta\varepsilon_{tot}}{2} = \frac{\sigma_f}{E}(2N_f)^b + \varepsilon_f(2N_f)^c$$

Key
$\Delta\varepsilon_{tot}$= Total Strain amplitude
ε_{ft}= Fatigue Ductility coefficient (Typically true strain at fracture)
c= fatigue Ductility exponent (typically -0.7<c<-0.5)
σ_f= fatigue strength coefficient (typically true stress at fracture)
b =Basquin's Exponent (typically –0.12<b<-0.05) (5)
E= Young's Modulus (Pa)
N_f= Cycles to Failure

Evaluating the elastic component first, the part of the equation representing plastic strain contribution reduces to zero and the equation simplifies to the form shown below. The maximum stress value obtained from the figures was used to predict the lifetime. The results are shown in table 7.4.6, the table shows the solder stress values at fracture as obtained from (2,3). The Basquin's exponent value for eutectic solders was unknown so the upper limit (-0.05) and lower limit (-0.12) were used and the obtained values are shown.

$$\frac{\sigma_{obs}}{2} = \sigma_f(2N_f)^b$$

	Stress at fracture (@ temp (°C))	***Cycles to failure (b=-0.12)***	***Cycles to failure (b=-0.05)***
FR4 1.4W	20MPa (@ 90°C)	624	$1.352\text{x}10^{7}$
FR4 1.2W	22MPa (@ 80°C)	3923	$1.112\text{x}10^{9}$
Copper 1.4W	40MPa (@ 40°C)	$4.037\text{x}10^{7}$	$4.74\text{x}10^{18}$
Copper 1.2W	40MPa (@ 40°C)	$1.077\text{x}10^{8}$	$5\text{x}10^{19}$

Table 7.4.6: *summary of lifetime predictions using modified Coffin Manson relationship based on equivalent Von Mises stress.*

Where σ_{obs} is the observed stress (*stress= (Young's Modulus)/(elastic strain)*). While the actual number of cycles to failure cannot be determined due to the unknown true

value of Basquin's exponent, it can clearly be seen that just by reducing the power from 1.4 to 1.2W has a significant increase in the number of cycles to failure. A large contribution to this is the increased fracture strength of the solder when the temperature amplitude is decreased from 70 to 60°C, in addition to lower stress values. Furthermore there can be seen to be a huge increase in the number of cycles to failure when copper is used as a substrate for the same power levels. Again this is due to the increased fracture strength of solder given the smaller temperature amplitude endured in the power cycle.

Only the MCM on FR4 powered with 1.4W exhibited any plastic strain so the plastic strain was seen to be small and was put into the equation with the elastic part reduced to zero (i.e. original Coffin Manson law was used). The strain at fracture used was obtained from (1,2) and the fatigue ductility exponent was unknown but values are known to be between -0.5 and -0.7. Evaluating the Coffin Manson equation with these values resulted in cycles to failure between 49 207 and 3.71×10^6 cycles due to plastic strains.

7.4.7: Die Stresses

So far the stress results have only been concerned with the characteristics of the solder but what effect do the stresses have on the die? The plots of the die for when the MCM is mounted on FR4 and copper are shown in figures 7.4.7a and b respectively. For both models it can be seen that the stress levels are comparable across the majority of the silicon. Both models show the heater chip is under greater stress than the carrier chip, where the stress is greatest at the bottom of the chip and reduces towards the top. This was believed to be due to the heater chip and expanding a greater amount than the carrier chip, as implied by the nature of the contours. The greatest levels of stresses on each assembly were found at the silicon-joint interfaces where values of 6 and 2 MPa were found for FR4 and copper respectively. However, with the compressive yield stress of 120MPa for silicon these stresses were found to be of little consequence.

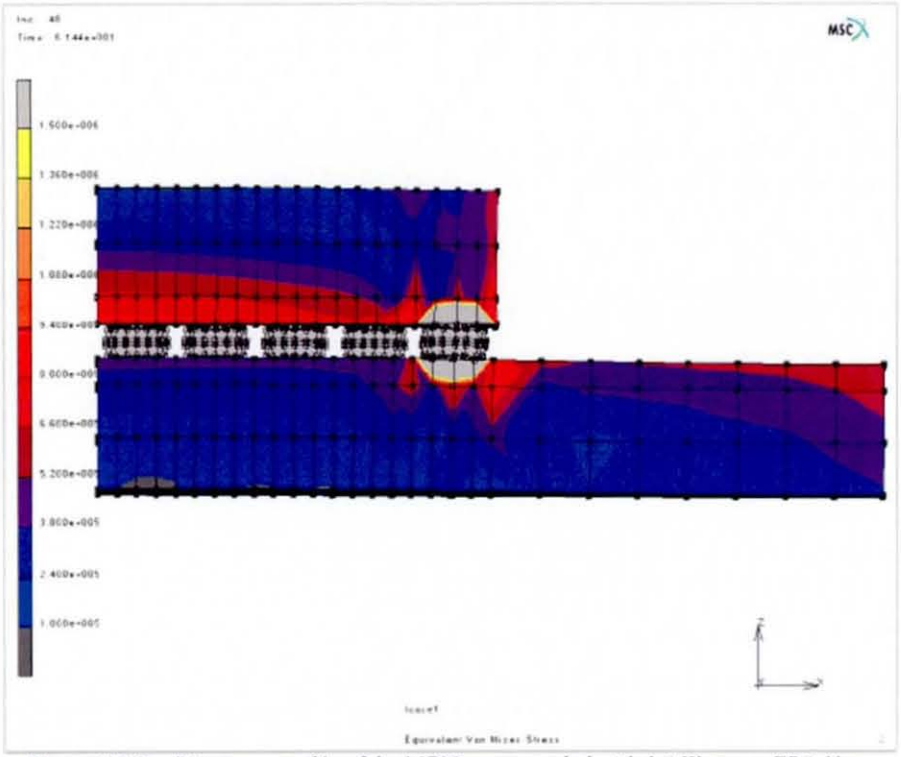


Figure 7.4.7a: *Die stress profile of the MCM power cycled with 1.4 Watts on FR4. Note the different contour ranges shown by the heater and carrier chips*

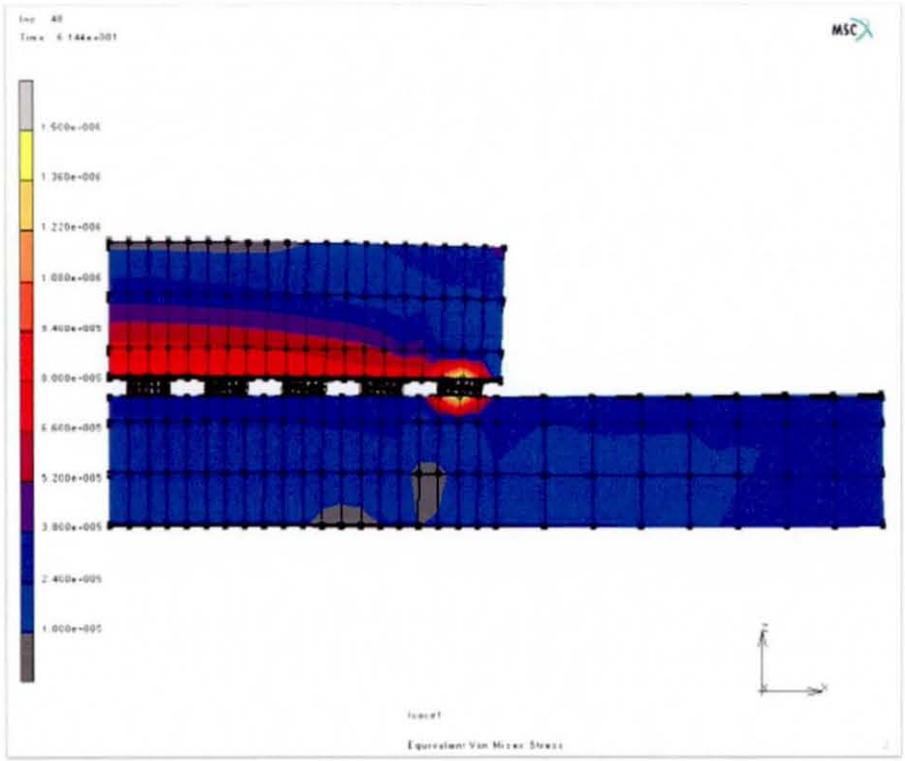


Figure 7.4.7b: *Die stress profile of the MCM power cycled with 1.4 Watts on FR4. Note the different contour ranges shown by the heater and carrier chips*

7.4.8: Comparison with Other Studies

The studies of design optimization by Tee (7) and Stoyanov et al (8) largely considered their optimization on the geometric influences of the model (e.g. pad diameter, board thickness etc.) as well as using a limited amount of thermo-mechanical properties. In this experiment, it was apparent that simply altering the substrate thermal properties has substantially reduced the stress levels in the solder by a factor of 2 with all other conditions kept the same. While it may be anticipated that increasing the thermal conductivity of the substrate would reduce the stress levels in the joint (by reducing the temperature amplitude for a given power level) the exact extent can now be quantified and the recommendation passed on to designers and manufacturers. The thermal properties of a material would be undetectable with conventional thermal cycling.

Sur and Turlik (9) conducted a power cycle of a flip chip assembly and reported a shearing action in the reverse direction for the first three seconds before the substrate heated up and *overtook* the die due to a higher CTE rate. This is very likely to be a ubiquitous scenario of a silicon die attached to a substrate such as FR4 though such a phenomenon was not present in this study. In the work presented here, the joints were also sheared in the opposite direction than that expected in conventional thermal cycling and during the ON cycle as the carrier chip is always 8°C cooler than the heater chip, this shearing in the opposite direction was maintained throughout.

The stress distributions of the solder joints obtained from the power cycle simulations appear to contradict typical findings from traditional thermal cycling results where the CTE mismatch between the chip and substrate had been persistently identified as the dominant failure mechanism. In addition, previous thermal cycling work had always predicted that corner joints exhibit significantly greater stresses than those closer to the neutral point. Therefore based on this logic, it was anticipated that the corner joints (4 or 5 as depicted in figure 7.2.3ai) would show evidence of much greater stresses than the joints halfway along the edge (joints 1 or 9 according to figure 7.2.3ai). In this work, all the joints showed very similar stress levels, therefore impeding the process of identifying the joint that is likely to be the first failure. If the

MCM test vehicles were subjected to a thermal cycle trial it is likely that (even if the nominal stress values are different) the solder joints would have very similar values (i.e. an "over stressed" joint would not be identifiable) in agreement with the findings of Trigg & Corless (10). However, if the carrier chip was made of a material of similar conductivity but a different CTE rate, then a joint exhibiting some significant "maximum stress" value may be obtained in the corner joints.

7.5: Conclusions

The stress values for the joints have been obtained. The octagonal approximation of the solder joints and the omission of the inevitable intermetallics were likely to compromise the accuracy of the nominal stress values obtained. However, it has been shown that the stress levels are much lower when a copper substrate is used as opposed to the FR4 for the same power/airflow conditions implemented. This was the observed scenario despite the equivalent temperature differences between the heater and carrier chip, which implied that the origins of the stress were also due to some other mechanism apart from the shear stress as a result of the temperature difference. It was observed that the CTE mismatch between the UBM and solder was a significant contribution to the stresses developed. The results correspond with a substantially longer predicted life times for the assemblies mounted on copper.

Plastic deformation contributes to an extent to some of the cases when the MCM is mounted on FR4, however the ambient temperature of the assembly is also an important factor when determining the extent to which plastic deformation occurs. It is also apparent that the solder joints are stressed in a different way to conventional thermal cycling, therefore different failure mechanisms are likely.

References

1) **Min He, Zhong Chen, Guojun Qi** "Solid State Interfacial Reaction Sn-37Pb and Sn-3.5Ag Solders with Ni-P Under Bump Metallisation" *Acta Materialia 52, 2004, pages 2047-56*

2) **John Lau** "Solder Joint Reliability of a Low Cost Chip Size Package" (Introductory Invited Paper) *Microelectronics Reliability 38, 1998, pages 1519-29*

3) **Chin C Lee** "Analysis and Study of Stress in the Structure of High Power Laser Chip Bond on a Metallic Joint", *Final Report 1999-2000 for MICRO 99069, 1999, pages 1-8*

4) **Neal F. Enke, Thomas J Klinski, Scott A Schroeder and Jon Lesniak** "Mechanical Behaviours or 60/40 Tin-Lead Solder Lap Joints" *IEEE Transactions on Components, Hybrids and Manufacturing Technology Volume 12 No 4, December 1989, pages 459-68*

5) **Gunther Grossmann and Ludger Weber** "Metallurgical Considerations for Accelerated Testing of Soft Solder Joints" *IEEE transactions on Components Packaging and Manufacturing Technology Part C Volume 20 No 3, July 1997, pages 213-8*

6) **Milton Ohring** "Reliability and Failure of Electronic Materials and Devices" *Academic Press, 1998, Chapter 9*

7) **Tong Yang Tee, Hun Shen Ng, and Zhaowei Zhong** "Design Optimisation of Wafer Level CSP Solder Joint Reliability" *EMAP2003 Proceedings Singapore, November 17-19 2003, pages 184-9*

8) **S, Stoyanov, C. Bailey, H Lu and M Cross** "Solder Joint Reliability Optimisation" *APACK 2001 Conference on Advances in Packaging, Singapore*

9) **Biswajit Sur and Iwona Turlik** "Power Cycling and Stress Variation in a Multi-Chip Module" *IEEE Transactions on Components Packaging and Manufacturing Technology part B Volume 10 no 2, May 1995, pages 388-395*

10) **A.D. Trigg & A.R. Corless** "Thermal Performance and Reliability Aspects of Silicon Hybrid Multi-Chip Modules" *40th Electronic Components and Technology Conference, 1990, pages 592-9*

Chapter 8: Power Cycling of the MCM Assemblies

8.1: Background and MCM Manufacture

Background

The thermal attributes of the MCM attached to the different substrates have been characterised in previous sections. FE models were created to give predictions of the maximum temperatures reached and the overall transient profiles for both the assembly configurations with specified power levels and airflow rates. Also, manufactured test vehicles were thermally profiled under equivalent conditions and found to be in good agreement with the results from the FE models. Following these studies, FE models were used to perform coupled analyses for two power levels such that predictions of the resultant stresses and strains in the solder joints were obtained. However, there was no apparatus available to measure the "real" stress/strain of a specific solder joint within the MCM. Therefore a large number of assemblies were manufactured and subsequently cycled to failure, as a correlation with the cycles to failure and the stress/strain endured was expected. As earlier modeling work had shown that the substrate upon which the MCMs were mounted had a significant influence on the level of stress, two batches of devices were cycled: one with FR4 substrates and the other with copper substrates.

MCM manufacture

For the power cycling trials, a total of 39 assemblies were manufactured in accordance with the procedure described in chapter 3. It was decided that it was important that each assembly could be identified later on and its performance tracked and evaluated. Initially, both the heater and carrier chips were numbered to permit a comprehensive history record for each assembly, however labelling the heater chips was difficult due to the small dimensions. Therefore only the carrier chip was numbered for the majority of the assemblies, though the numbering of the heater chips was implemented for the first batch. The MCMs were manufactured in small batches of no more than 6 at a time, so that if a fault was detected it could be corrected without compromising the remaining assemblies. Table 8.1 shows the batches that the

MCMs were manufactured in; also shown is the MCM mortality, referring to the number(s) of the MCMs that did not function following the manufacturing process. Following the reflow stage, the MCMs were checked to verify that the interconnection was correctly formed. This was done by measuring the resistance of each path with an ohmmeter and comparing the readings with the generic values shown in table 3.1. Afterwards, satisfactory MCMs were divided into two equal groups and secured to the relevant substrates, followed by the external wire attachment procedures. Once finished, all their resistance values were recorded again to check for any changes. Each manufacturing batch was split in two with approximately half from each batch used for FR4 substrate trials and the remaining samples used for copper substrate trials, to provide consistency throughout the trials.

Batch no	Assembly number	Mortality
A (letters for carrier chips, numbers for heater chips)	A3, B5, C1, D4 and E2	C1, E2
B	1-3, A0,A1	2 with samples A0, A1, 1 & 3 used for profiling
C	4-9	6,7,8
D	10-12	
E	13-15	
F	16-20	18
G	21-26	23
H	27-30	28 (reworked)
J	31-36	33
K	37-39	39

Table 8.1: *Batch name and assemblies manufactured in them*

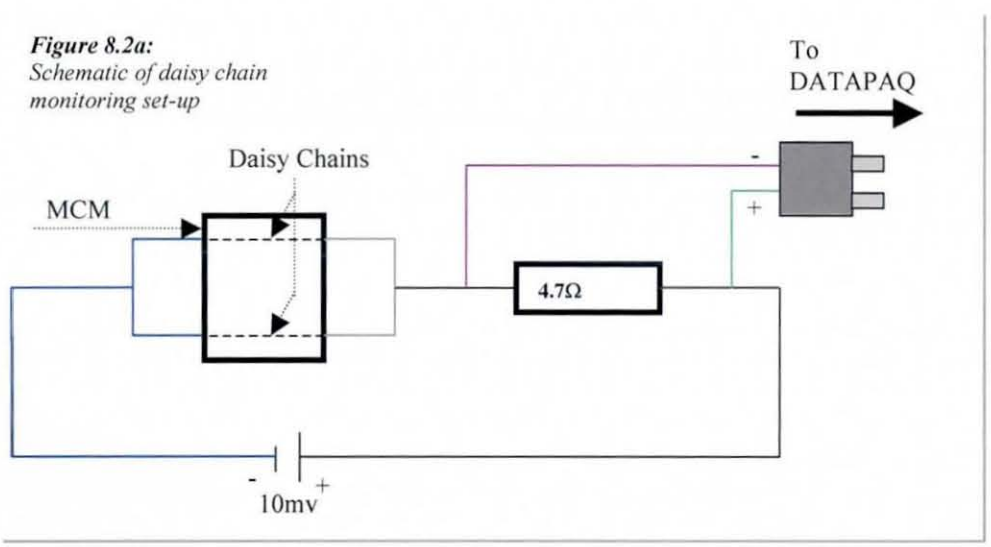
8.2: Power Cycling Apparatus and Procedure

The assemblies were power cycled in batches of 13 at a preset power level for a continuous period with a total cycle time of 128 seconds (64 seconds for the ON part and 64 seconds OFF). The timer described in section 4.1 controlled the cycle period, and the cycle chamber used was the wind tunnel described in section 4.1. The air rate was set to the maximum such that an average velocity of 10m/s was obtained over the designated specimen area.

The continuity/resistance of the daisy chains was constantly monitored using DATAPAQ data loggers. The two daisy chains on each MCM were connected in

parallel in order to maximise the number of chains that could be monitored simultaneously. The daisy chains were connected to the data loggers as shown in figure 8.2. Each pair of daisy chains on an MCM was connected in series with a 4.7Ω resistor and a potential of 10mV was applied from a dedicated power supply. The current flow through the daisy chains resulted in a small voltage being developed across the 4.7Ω resistors that was monitored by the DATAPAQ which sampled data every 4 seconds. The data loggers were secured to the wind tunnel body and they remained connected to the controlling PC for the duration of the power cycling. As the DATAPAQ devices were designed to monitor thermocouples, the voltage received was displayed as a temperature reading: a fully working MCM would display a temperature of approximately 140-150°C, an MCM with one failed daisy chain would be displayed as a temperature within the region of 100-120°C, a MCM with both chains failed would display a reading of room temperature.

Figure 8.2a:
Schematic of daisy chain monitoring set-up



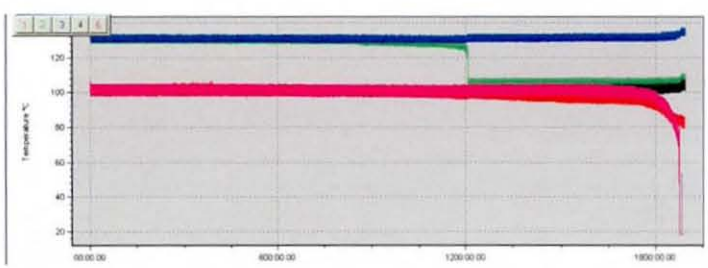


Figure 8.2.b:
Example of data-capture output.

- ***Blue line**: - A chip with both chains working*
- ***Green line**: - A chip where one chain fails during the data-capture*
- ***Magenta line**: - A chip with one chain working when data capture commences and fails at the end*

To maximise the amount of data, it was decided to cycle as many samples as possible at once. There were two DATAPAQs available for the experiment, each with 6 inputs. Each MCM had its respective daisy chains connected in parallel such that only one data input was used per assembly. As a result, this allowed 12 assemblies to be monitored at any given time. As well as those that were monitored, it was considered beneficial to have an additional *control* assembly that was to be cycled relentlessly for the entire cycling period; therefore 13 assemblies were power cycled in each batch. The assemblies were secured on a 500x230mm Perspex plate in the layout depicted in figure 8.2b and their positions recorded; the Perspex plate was subsequently positioned in the wind tunnel at the specimen viewing area. Once secured, the wires were also taped to the Perspex to minimize their obstruction to the air flow.

The power supply used to drive the heaters of the assemblies had a maximum output of 3 amps and 10 Volts. The heaters on the MCMs had a resistance of approximately 27Ω at room temperature, therefore considering the power supply and the number of chips cycled, the absolute maximum load possible per MCM was 6.2Volts at 230mA or 1.43W. As a result, the power supply was set to deliver this amount, however, a voltage drop of 0.5V occurred across the power cycling time controller (transistor internal resistance) and the final measurements at the MCM were 5.7V at 211mA which gave a power input for each assembly of 1.2W.

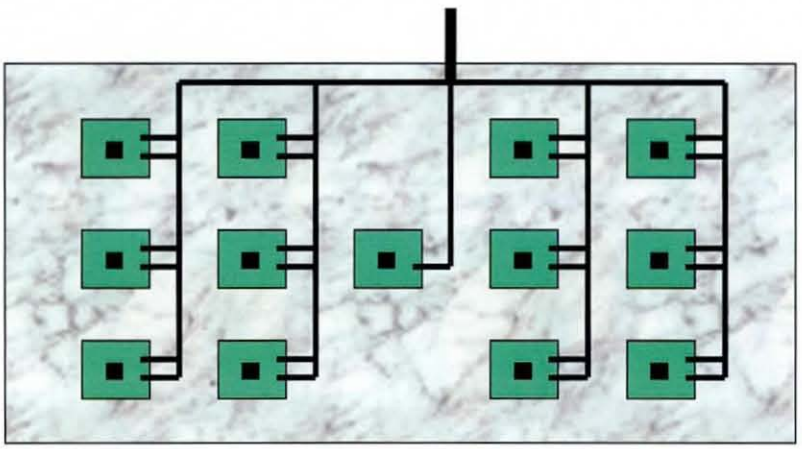


Figure 8.2b: *Specimen layout for simultaneous power cycling of components. The specimen in the centre was a control chip for which the temperature change and daisy chains were unmonitored.*

Before the assemblies were power cycled, the resistance values of the daisy chains and the heater were noted by probing the relevant connection pads. In addition, the 4-point resistant values were recorded using the micro-ohmmeter such that a record of each assembly at room temperature was made. The precise time the power cycling

started was logged and the samples were left to cycle. Every day it was possible to do so, the room temperature and airflow rates were checked; the data collected was also checked to see if any failure had occurred. When a single failure was noted, a log of the time of failure was made, in the case when total failure of an assembly was detected (i.e. both daisy chains failed) the assembly in question was disconnected from the main power supply to prevent further power cycling and "freeze" its condition and it would remain idle until the cycling had finished.

8.3: Specimen Preparation

Once the cycling was complete, the specimens were then sectioned so the joints could be analysed. This section explains the procedures used when preparing the samples including the mounting, potting and grinding procedures used. Different techniques were used depending on whether the samples were mounted on FR4 or copper.

8.3.1:Mounting and Potting Procedures

As soon as the samples had finished power cycling, the resistance values of both daisy chains from each MCM were checked and recorded. Based on the resistance values recorded, a decision was then made as to which daisy chain would be sectioned from a particular MCM (due to practical constraints, it was possible to only section one daisy chain per MCM).

When a daisy chain had been selected, the samples' external wires were disconnected from the substrate by resting a soldering iron on the external connection joint until the solder reflowed and the wire could be removed. There were two stages needed to encapsulate the MCM: namely coupling the heater and carrier chips, as well as the final potting stage. As underfill was not used in the assembly manufacture, the heater chip required mechanical coupling with the carrier chip to minimise potential damage during the subsequent mounting and grinding procedures. **EPOTHIN** was used for this purpose as it was designed to flow through narrow channels such as the standoff between the heater and carrier chips. The final encapsulant used was fast curing **SAMPL KWIK** such that the specimen could be gripped for the subsequent grinding

and polishing. In total there were three different procedures that were used to try and successfully prepare the MCM specimens. The methods were changed whenever a specific fault was noted in the procedure and they are explained below.

Method 1: Secure the MCM to the board: The overall procedure used initially is described in figure 8.3.1a. The **EPOTHIN** was deposited onto the MCM while it was still attached to the substrate and left for the required curing time. Once fully cured, the excessive substrate material was then removed using a **Draper hand-held multi-tool** with a 2.5mm diameter ceramic blade attached. The initial cutting procedure was applied to leave the remaining substrate approximately 20x20mm such that it could be positioned in the **Buehlers low-speed diamond saw**. The saw removed excess substrate and left a smooth edge approximately 1mm away from the start of the MCM. The samples were then mounted in a clip with the side of interest facing downwards and were then encapsulated with **SAMPL KWIK** potting compound. Following this the samples were ground and polished to reveal the solder joints using the method described later on in table 8.3.1.

This procedure was used to mount all the MCMs that were on FR4 substrate, however, there were very few successfully sectioned assemblies using this method. It was noted that this method required many cutting procedures that would have been omitted had the MCM been removed from the substrate in the first instance. It was felt that these cutting procedures (the hand saw and the diamond cutter) might have damaged the assembly through vibration. The decision was then made to attempt to remove the MCMs from the substrate before encapsulating them with EPOTHIN and this was used for the samples mounted on copper substrates.

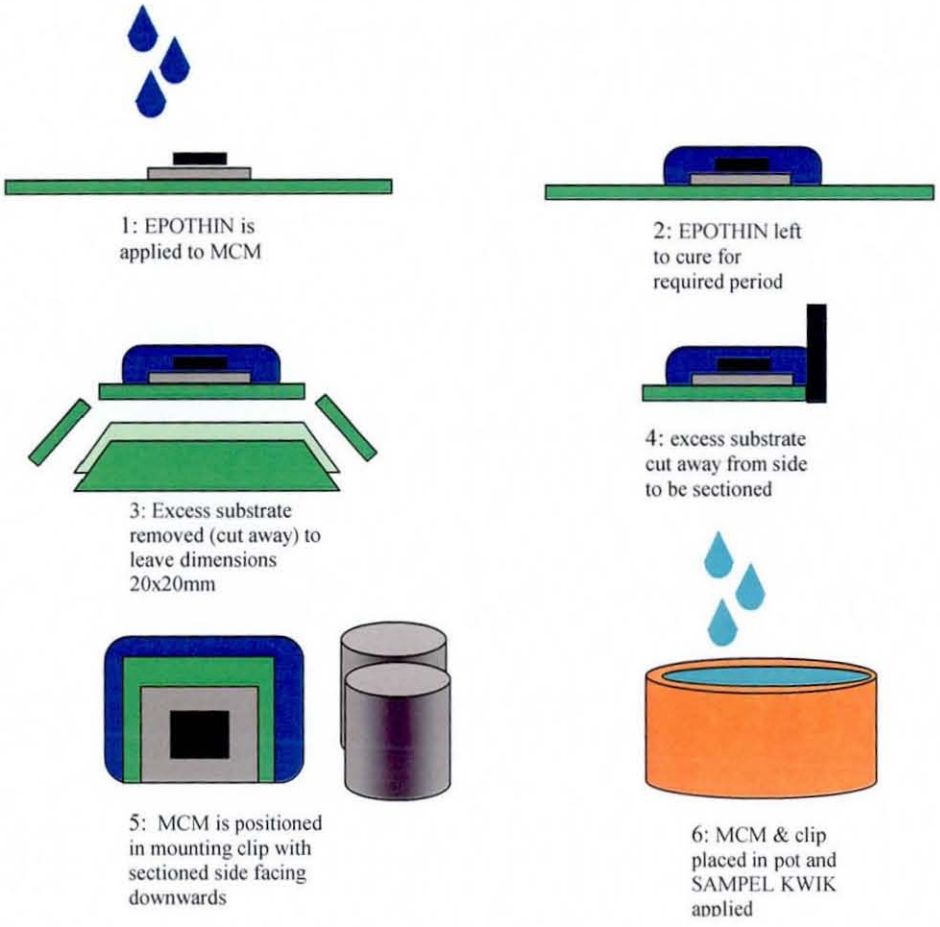


Figure 8.3.1a: *Implementation of the first mounting & potting procedure*

Method 2: Mechanically secure the MCM to the mounting clip: The **Chomerics** adhesive pad allowed the chip to be removed from the substrate by simply prizing apart with a thin blade, however before this could be done, the thin wires connecting the MCM to the substrate were severed using a sharp instrument such as a knife. The MCM was aligned and vertically mounted in a steel clip and the EPOTHIN® was subsequently applied. A metal clip was used in order to increase the electrical conductivity of the potted sample to enhance the images obtained from the SEM. The assemblies that were mounted in this way were all originally attached to copper substrates and were numbers 11, 19, 25, 26 and 30. However, again, this method was not found to be successful in producing good samples for cross-sectioning, largely due to the force exerted by the steel-mounting clip on the unsupported heater chip that damaged the solder joints.

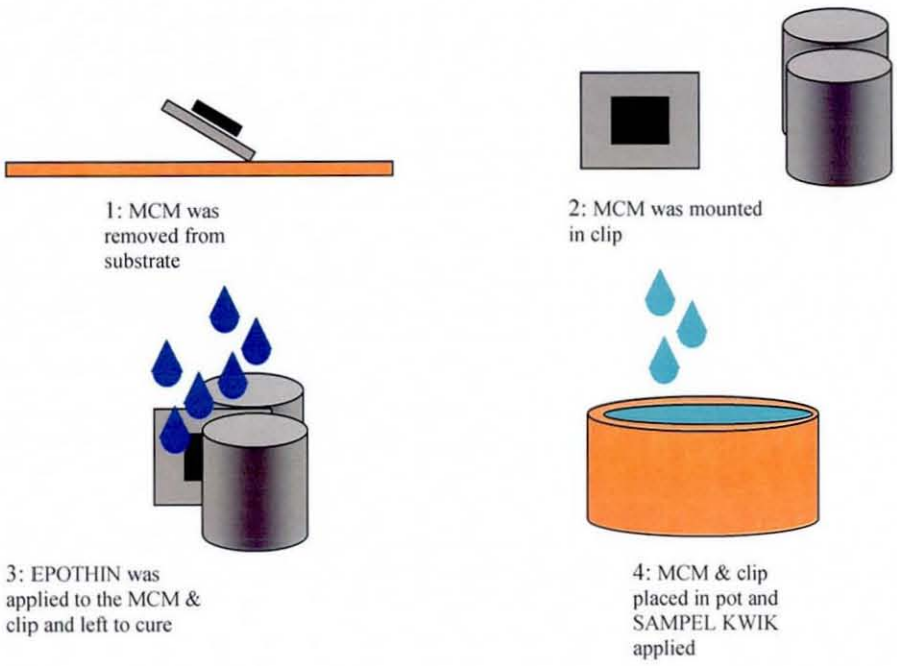


Figure 8.3.1b: *Implementation of the second mounting & potting procedure*

3) Mechanically secure the MCM only: The most successful mounting procedure involved the application of EPOTHIN® to the MCM before mounting it in a plastic clip. Plastic clips were used as they were significantly softer than steel. The remaining copper MCMs were all mounted in this way, namely samples 12, 15, 29, 35, 36, 37, 38 and B5

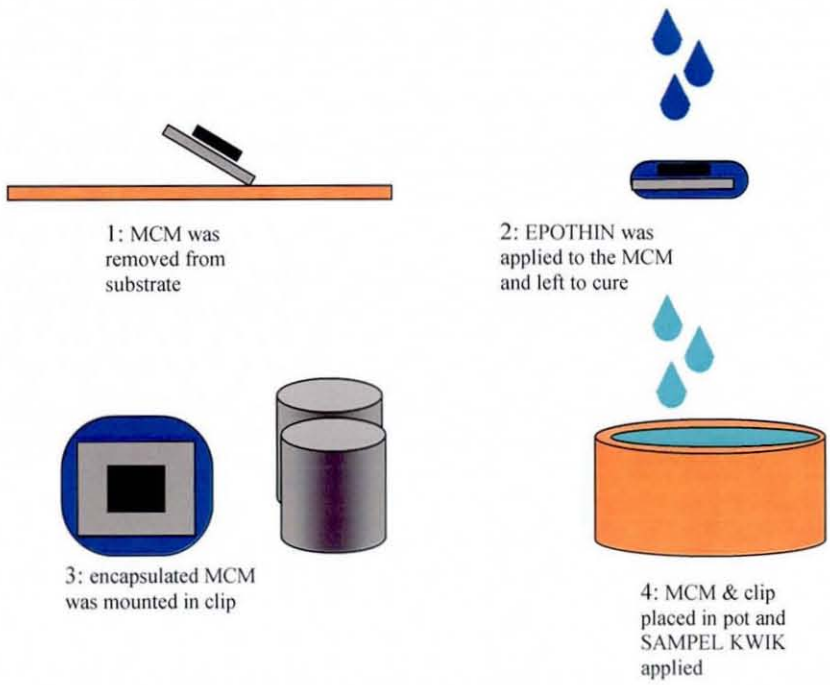


Figure 8.3.1c: *Implementation of the final mounting & potting procedure*

Table 8.3.1 shows the grinding and polishing procedures as well as the grit sizes used. All the specimen preparation surfaces available were for general/multi-purpose use (i.e. they were not specifically used for preparing soft solder surfaces), however considerable effort was made to ensure the surfaces were free from foreign debris. They were rinsed with water and if necessary a fresh surface was laid. For SEM analysis, the samples were gold coated using an **Edwards vacuum coating machine** in order to produce an electrically conductive surface that was suitable for imaging.

Stage	Grit size	Comments
1	120 emery (Rotating at 250rpm)	Was performed until heater chip was exposed.
2	240 emery (by hand)	Until first bump was visible
3	320 emery (by hand)	10 strokes
4	400 (by hand)	10 strokes
5	600 (by hand)	10 strokes
6	6μm diamond polishing (300)	2 minutes
7	1μm diamond polishing	3 minutes
8	0.5μm alumina polishing	3 minutes

***Table 8.3.1:** The grit sizes and times used when grinding and polishing samples*

8.4: Results

The results are first presented for FR4 and then copper substrates. For each section, the results first show the estimated temperature change for each assembly during various periods of the power cycling. Following this, the results of the failure times and the final heater and daisy-chain resistance values are shown.

8.4.1: FR4 Substrate Results

The positions of the assemblies on FR4 substrate is shown in figure 8.4.1a They were power cycled continuously under the conditions described in 8.2 for a total of 22 362 cycles or 33 days 3 hours and 5 minutes. During this time the ambient temperature was known to vary from 20 to 26°C. Within this duration there were a total of 11 daisy chain failures in 7 MCMs.

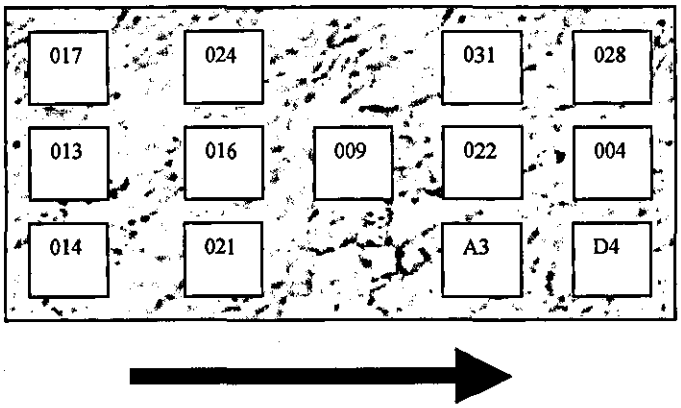


Figure 8.4.1 Schematic of the positioning of chips on FR4 substrate, the arrow indicates airflow direction.

Thermal Performance

The thermal data captured is displayed in tables 8.4.1a to c. Table 8.4.1a shows the records of the 4-point resistance values for each assembly at the end of the ON (peak) cycle and the OFF (low) cycle, the readings were taken at various points in the trial. T1 denotes readings that were taken early on (after approximately 100 cycles), T2 represents records after approximately 15 000 cycles while T3 shows the readings taken after approximately 22 300 cycles. Table 8.4.1b shows the change in resistance at these times as well as the estimated temperature change based on the calibration described in section 4.2.2. The changes in resistance are also displayed as a bar chart in figure 8.4.1a. Table 8.4.1 summarises the data from tables 8.4.1a and b as means and standard deviations. The average change in resistance observed for the assemblies at the start of the trial was 1.3Ω which corresponded to a temperature change of approximately 60°C, in good agreement with the previous thermal profiling and FE results. From the data it can be seen that many samples failed before T2 in addition at T2 and T3 the temperature difference between ON and OFF cycles increased. The resistance difference increased by approximately 15% leading to a corresponding temperature rise estimated to be 70°C. This temperature increase was thought to be as a consequence of disconnecting "finished" (failed) samples from the power supply to prevent further cycling and thereby reducing the voltage drop across the transistor of the power cycling timer so that the remaining assemblies were subjected to a greater power level during the latter part of the trial.

MCM number	Initial reading Ω	T1 Peak Ω	T1 LowΩ	T2 Peak Ω	T2 Low Ω	T3 Peak Ω	T3 Low Ω
004a	5.940	7.341	6.051				
004b	5.980	7.399	6.093				
009a	N/A	N/A	N/A				
009b	N/A	N/A	N/A				
013a	5.790	7.070	5.878				
013b	5.714	6.970	5.798				
014a	6.507	8.120	6.656				
014b	6.828	8.440	6.924				
016a	6.150	7.569	6.252	7.796	6.271	7.813	6.273
016b	6.164	7.593	6.260	7.820	6.285	7.832	6.286
017a	6.073	7.442	6.165	7.631	6.173	7.658	6.191
017b	6.109	7.496	6.204	7.685	6.215	7.708	6.228
021a	6	7.390	6.110				
021b	5.942	7.320	6.048				
022a	6.029	7.381	6.125				
022b	6.018	7.368	6.111	7.580	6.119	7.587	6.119
024a	5.858	7.259	5.936				
024b	5.932	7.341	6.009				
028a	6.018	7.428	6.129				
028b	5.950	7.356	6.056				
031a	5.955	7.198	6.017				
031b	5.778	6.978	5.829				
A3a	5.811	7.164	5.897				
A3b	26.56	29.82	27.61				
D4a	6.690	8.251	6.808				
D4b	7.209	8.889	7.332				

Table 8.4.1a: *4-point resistance records for FR4 samples. a and b correspond to the two aluminium tracks (a is path 3344, b is path 5566 from section 3.2.1). T1 is after 200 cycles, T2 is after 15 000 cycles and T3 is the readings just before the trial ceased.*

MCM number	T1 Resistance change / Ω	T1 Temperature rise / °C	T2 Resistance change / Ω	T2 Temperature rise / °C	T3 Resistance change / Ω	T3 Temperature rise / °C
004a	1.29	60				
004b	1.306	60.8				
009a						
009b						
013a	1.192	55.5				
013b	1.172	54.6				
014a	1.464	68.2				
014b	1.516	70.6				
016a	1.317	61.3	1.525	71	1.54	71.5
016b	1.333	62.1	1.535	71.5	1.546	72
017a	1.277	59.4	1.458	67	1.467	68
017b	1.292	60.1	1.47	68.5	1.48	68.9
021a	1.28	59.5				
021b	1.272	59.2				
022a	1.256	58.5				
022b	1.257	58.6	1.461	68	1.468	68
024a	1.323	61.6				
024b	1.332	62.1				
028a	1.299	60.5				
028b	1.3	60.6				
031a	1.181	55.0				
031b	1.149	53.5				
A3a	1.267	59.0				
A3b	2.21	102.9				
D4a	1.443	67.2				
D4b	1.557	72.				

Table 8.4.1b: *Four point resistance changes from table 8.4.1a and corresponding temperature changes. a and b correspond to the two four point resistance aluminium tracks (a is path 3344, b is path 5566 from section 3.2.1). T1 is after 200 cycles, T2 is after 15 000 cycles and T3 is the readings just before the trial ceased.*

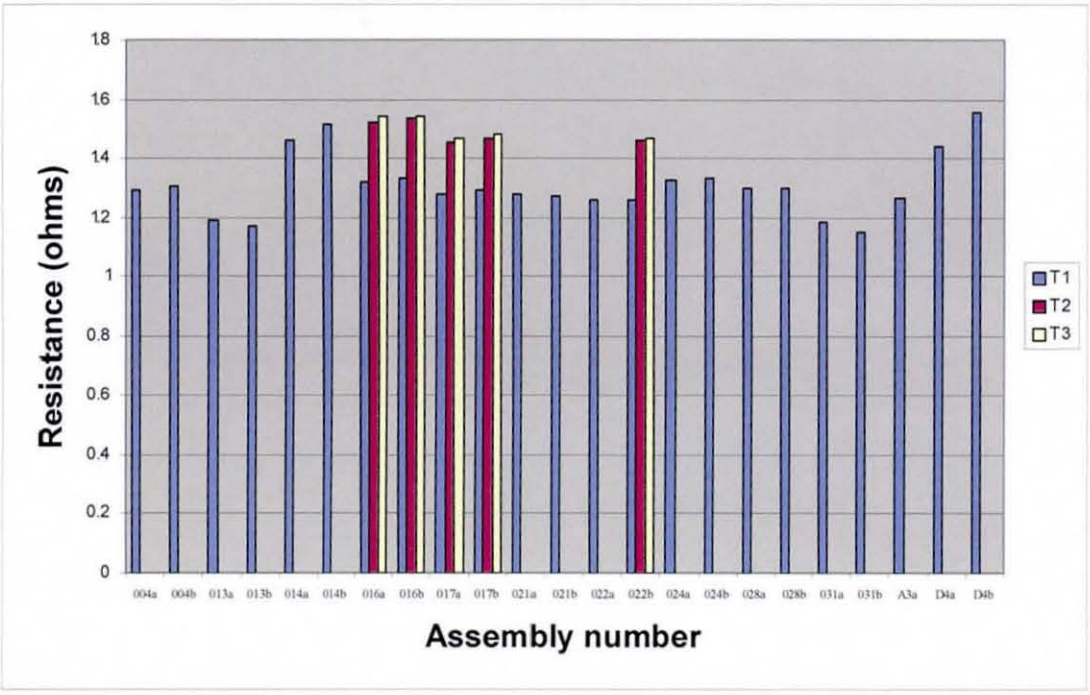


Figure 8.4.1: *Bar chart of resistance changes from table 8.4.1.b*

Chip number	Rest (Ω)	Peak (Ω)	Low (Ω)	Change in resistance (Ω)	Corresponding temperature (°C)
Table A	6.068, 0.26	7.468, 0.347	6.169, 0.275	1.299, 0.08	60.48, 3.763
Table B (Int)	6.147, 0.436	7.559, 0.559	6.242, 0.446	1.317, 0.118	61.29, 5.4
Start (c & d)	6.103, 0.053	7.494, 0.082	6.198, 0.06	1.310, 0.02	61.1,1.3
Table C	N/A	7.702, 0.09	6.213, 0.06	1.49, 0.03	68. 1.75
Table d	N/A	7.720, 0.09	6.219, 0.06	1.50, 0.03	69.2,1.75

Table 8.4.1c: *Obtained mean and standard deviation of data from tables 8.4.1a &b*

Specimen Lifetime and Fatigue

Table 8.4.1d displays the failure times of the relevant specimens in chronological order. It could be seen that all the recorded failures occurred relatively early on in the cycling time; below 3500 cycles, yet 3 assemblies were observed to have survived the entire power cycle duration with all their daisy chains showing no obvious change in resistance. Therefore there was a period of approximately 20 000 cycles between the last recorded failure and the termination of the power cycling; the cause of so many apparently premature failures is not known though it was anticipated that some of the daisy chains could contain faulty solder joints, or other typical "burn-in" problems. It should be highlighted that a daisy chain only needed to fail at one connection and this would be registered as a failure.

Sample	Failed after
028 *(first chain)*	136 cycles
028 *(second chain)*	169
014 *(first chain)*	172 cycles**
024 *(first chain)*	195 cycles
021 *(first chain)*	643
A3 *(first chain)*	890
013 *(first chain)*	1000
D4 *(first chain)*	1270
031 *(first chain)*	1871
004 *(first chain)*	2185
021 *(second chain)*	2209
031 *(second chain)*	2618
024 *(second chain)*	2948
013 *(second chain)*	3240
016	>22 362
017	>22 362
022	>22 362

Table 8.4.1d: *Number of cycles to failure for each daisy chain*

Table 8.4.1e shows the daisy chain resistance readings for each assembly recorded immediately before and after the power cycling. It was assumed that samples that displayed a dramatic increase in resistance had failed at one or possibly two joints, though there was still a small amount of contact (touching) that allowed a continuity signal to be detected when the samples were checked. According to tables 8.4.1d and e it can be seen that sample 14 was recorded to have failed, yet on the post cycling resistance measurement, finite values could be obtained. It was later found that some of the connections from the MCM to the substrate were not adequately attached and consequently became loose during the power cycling.

Chip number	Thick (Ω)	Daisy Chain 1 Before cycling (Ω)	Daisy Chain 2 Before cycling (Ω)	Daisy Chain 1 Post cycling (Ω)	Daisy Chain 2 Post cycling (Ω)
004	26.9	4.3	4.2	∞	∞
009	27.0	N/a	N/a	4.3	27.1
013	26.8	4.4	4.1	∞	∞
014	27.0	4.4	4.1	4.5	4.3
016	27.0	4.3	4.1	4.6	4.2
017	26.9	4.2	3.9	4.2	3.9
021	27	4.3	4.0	∞	∞
022	27	4.2	3.9	4.3	4.0
024	27	4.3	4.0	∞	∞
028	26.9	4.2	4.0	∞	∞
031	26.8	4.3	4.0	8.6	9.2
A3	26.9	4.5	4.3	4.8	∞
D4	26.9	4.6	4.3	∞	8.5

Table 8.4.1e: *Recorded resistance values of assemblies before and after cycling*

Failure Analysis of Samples on FR4 Substrate

As well as the failures recorded on the daisy chain logged in table 8.4.1d, there were additional failures that were noted (i.e. joints failed that were not on the daisy chains). Failures at the heater (thick) path were noted in assemblies A3, 024 and 028. The heater on 028 was found to fail shortly after the second daisy chain had failed. The temperature change in 024 was observed to be only half that of the original change at the time of the second daisy chain failure, presumably because the resistance of one of the heater connection joints had increased. For A3, the heater failed before the second daisy chain had failed. In addition, samples 004, 021,024 and 028 all lost the ability to read the 4-point resistance measurements, and subsequent probing found that they had all failed at a single joint on their respective 4-point resistance parts. Unfortunately as failures at these locations were not catered for in the original failure monitor set-up there were no means of recording when these failures occurred.

Due to a defective potting and grinding procedure, there were no clear cross-section images of failed solder joints for any of the samples that were monitored in this batch. The joints obtained from the surviving assembly 022 are shown at the end of the section. It can be seen that good wetting had occurred on both the UBMs, and the pads had aligned well. The interconnection layer exhibited many of the typical characteristics of well-formed joints therefore it was anticipated that this should survive the power cycling.

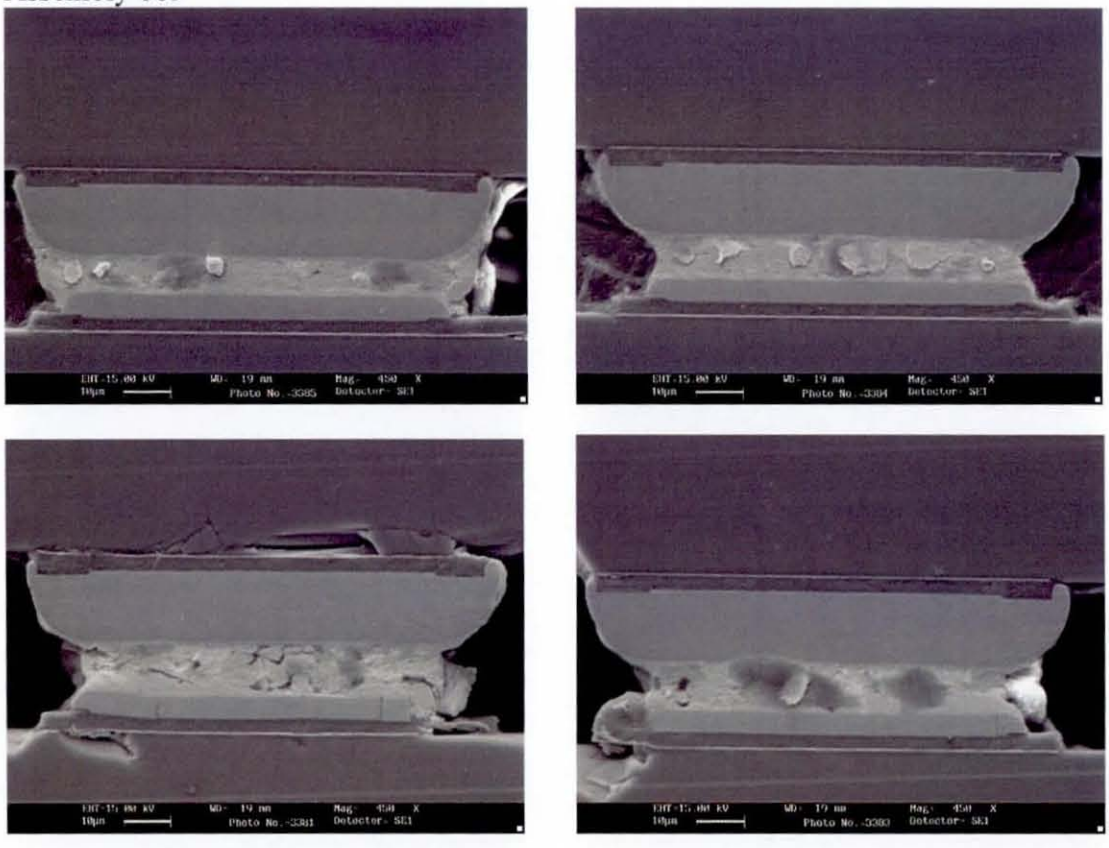
Also included in the trial was assembly 009, though this was unmonitored. According to table 8.4.1e it can be seen that there was a failure on one of the daisy chains (the higher resistance). From images of the section, it can be seen that there was very poor solder wetting on the UBM pad on the heater chip and the actual connection was only sufficient to secure the assembly and carry a signal. These left a comparatively high standoff height between the heater and carrier chips, and the badly formed joints may have acted as stress raisers.

Though many of the FR4 MCMs were destroyed during the mounting and preparation stages, individual records for each assembly were kept regarding the manufacture and its power cycling performance. Some of these records imply that some assemblies

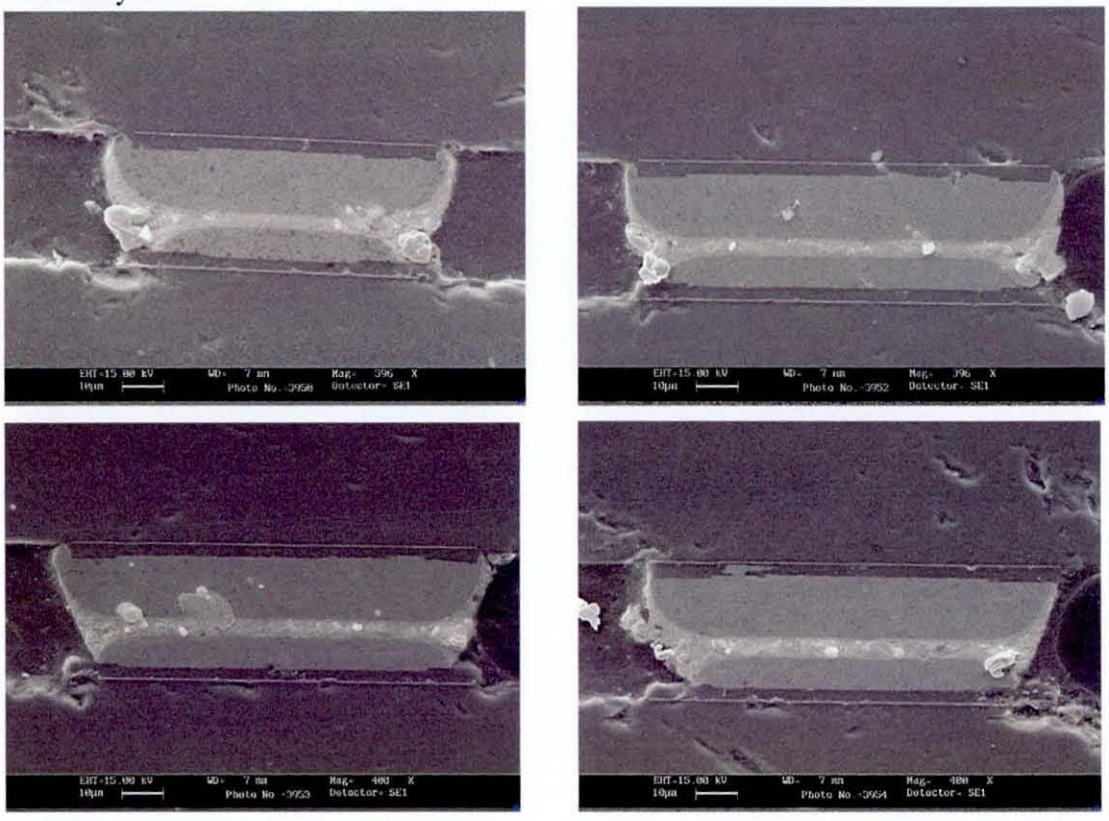
may have been badly manufactured, therefore defective interconnection layers may have been present (there was no way of checking the quality of the daisy chains apart from the resistance values). For example sample number 028 failed on both daisy chains early on, as well as the heater failing some short time after the second failure. This assembly was subject to re-work due to failure to reflow the first time it was assembled. The assembly had been cleaned ultrasonically in IPA and all the fluxing, dipping, cleaning and assembly procedures were performed a second time and upon inspection, the interconnection was found to be satisfactory. It may be possible that the second solder dipping process had increased the layer of intermetallics present between the UBM and the solder, therefore making all the joints very brittle.

In addition, both daisy chains on sample 024 failed, though one chain lasted over 15 times longer than the other. The heater was observed to fail over a period of time similar to when the second chain failed. It was thought that too little flux was applied to this assembly and the oxides on the solder caps were not correctly removed, so the interconnection layer was not formed properly.

Assembly 009



Assembly 022



8.4.2: Copper Substrate Results

8.4.2.1: Power Cycling under Standard Conditions

A batch of MCM samples on copper substrates were initially power cycled with the same power levels as those on FR4 (standard conditions). On completion of these trials some of these samples were power cycled at an increased power level (high power) and the results of these will be discussed later.

The positions of each assembly in the wind tunnel are shown in figure 8.4.2a. They were power cycled continuously under the same conditions as for the FR4 samples for a total of 15 553 cycles. During this time the ambient temperature was known to stay at 22°C during the daytime though the temperature may have dropped significantly in the evening. Within this test duration there were a total of 5 daisy chain failures from 4 assemblies.

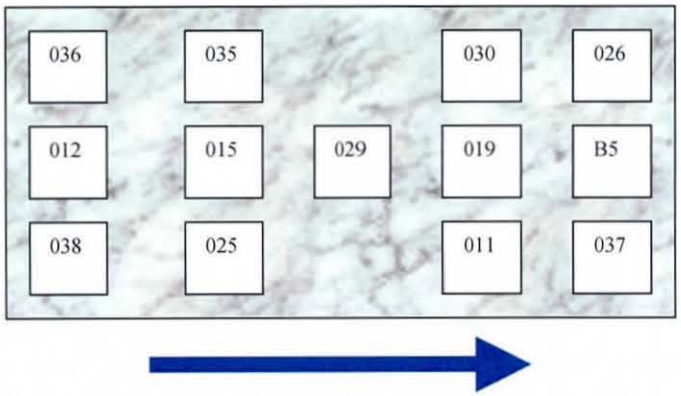


Figure 8.4.2.1 *Layout of the MCMs on copper substrate in the wind tunnel, the arrow indicates airflow direction.*

Thermal Results

Due to difficulty in successfully soldering the connection from the MCM to the substrate on these assemblies, it was not possible to utilise all the 4-point resistance connections on each and every sample, so for many of them, only one 4-point resistance measurement was used. Table 8.4.2a shows the 4-point resistance measurements taken at two time periods: the first after 200 cycles (T1) and after approximately 15500 cycles (T2); while table 8.4.2.1b shows the corresponding resistance/temperature changes and graph 8.4.2.1 shows the readings

It can be seen that one sample had a much greater temperature reading, however this was considered to be due to a poor connection or damaged aluminium track and the sample was omitted from the averages. Comparing the two tables, it can be seen that the temperature amplitude at the end of the cycling was marginally higher for each sample. As with the case for FR4 samples, one sample was disconnected before the end of the power cycling so the voltage drop across the transistor was reduced.

MCM number	Initial reading / Ω	T1 Peak / Ω	T1 Low / Ω	T2 Peak / Ω	T2 Low / Ω
011a	6.213	6.983	6.279	6.983	6.275
012b	6.047	6.732	6.123	6.743	6.122
015	5.943	6.554	6.008	6.569	6.008
019b	6.314	7.035	6.423	7.028	6.416
025a	6.104	6.859	6.202		
026b	6.076	6.740	6.142	6.766	6.162
030b	5.990	6.621	6.060	6.624	6.061
035b	5.594	6.223	5.658	6.231	5.658
036b	38.41	40.05	38.59	40.05	38.59
037b	5.811	6.444	5.876	6.443	5.856
038b	6.330	7.044	6.397	7.044	6.397
B5	6.753	7.520	6.838	7.566	6.830

Table 8.4.2.1a: *4-point resistance records for MCMs on Copper substrate. T1 refers to the readings after 200 cycles while T2 refers to the records after 15500 cycles.*

MCM number	T1 Resistance change / Ω	T1 Estimated temperature change / °C	T2 Change in resistance / Ω	T2 Estimated temperature rise / °C
011a	0.704	32.7	0.708	33
012b	0.609	28.3	0.621	29
015	0.546	25.4	0.561	26.1
019b	0.612	28.5	0.612	28.5
025a	0.657	30.6		
026b	0.598	27.9	0.604	28.1
030b	0.561	26.1	0.563	26.2
035b	0.565	26.3	0.573	26.8
036b	1.46	68	1.46	68
037b	0.568	26.4	0.587	27.3
038b	0.647	30	0.647	30.1
B5	0.682	31.7	0.736	34.2

Table 8.4.2.1b: *Corresponding resistance and temperature change for table 8.4.2.1a. T1 refers to the readings after 200 cycles while T2 refers to the records after 15500 cycles.*

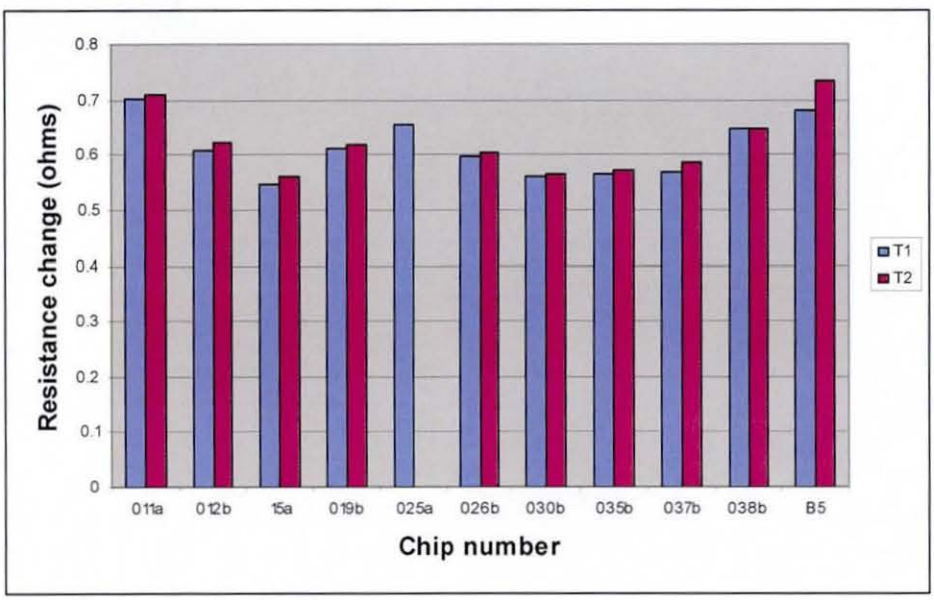


Figure 8.4.2.1a: *Graph showing the change in resistance from T1 & T2*

Chip number	Initial reading	Peak	Low	Change in resistance	Corresponding temperature rise
Start (Y=25)	6.107, 0.289	6.802, 0.33	6.182, 0.29	0.614, 0.050	28.53, 2.326
Start (N=25)	6.107, 0.303	6.790, 0.35	6.180, 0.31	0.609, 0.051	28.33, 2.3422
End	N/A	6.799, 0.36	6.179, 0.31	0.621, 0.057	28.93, 2.6313

Table 8.4.2.1c: *Summary (mean & standard deviation) of tables 8.4.2.1a (with & without failures) and b*

Lifetime and Fatigue Results

Table 8.4.2.1.d shows the resistance measurements of each assembly pre and post power cycling, while 8.4.2.1e shows the times of the recorded failures. With the exception of sample 038, it can be seen that there were no recorded resistance changes in any of the assemblies apart from those that were known to have failed (i.e. the data loggers indicated that a failure had occurred). The increase in resistance for sample 038 implies some joint fatigue and it was anticipated that had the power cycling continued, then 038 would have failed shortly afterwards. According to table 8.4.2.1e, it can be seen that the number of failures spans over half the duration of the power cycling period.

MCM number	Thick (heater) track	Daisy chain 1 (before cycling)	Daisy chain 2 (before cycling)	Daisy chain 1 (post cycling)	Daisy chain 2 (post cycling)
011	26.9	4.3	4.0	4.3*	4.0*
012	26.9	4.3	4.0	4.3*	4.0*
015	26.8	4.3	4.0	∞	4.2
019	27.0	4.3	4.0	4.3*	4.0*
025	27.0	4.3	4.0	16.7	8.6
026	27.0	4.3	4.0	43.7	4.1
029	27	4.5	4.0	N/A*	N/A*
030	27	4.3	4.0	∞	4.1
035	26.8	4.6	4.1	4.6*	4.1*
036	27.9	11.0	4.8	11*	4.8*
037	27.0	4.3	4.0	4.3*	4.0*
038	26.9	4.3	4.0	10	5.2
B5	27.0	4.5	4.3	4.5*	4.3*

Table 8.4.2.1d: *Resistance records of daisy chains before and after power cycling. The stars indicate the assemblies that were used for the subsequent high power cycling.*

MCM number	Number of Cycles to Failure	
026	Instant	
015	647	
030	6891	
025	7643	
025	8150	

Table 8.4.2.1e: *Record of sample failures during "standard" power cycling*

Failure Analysis of Samples on Copper Substrate

Sample 026 exhibited an instantaneous failure, however after the power cycling was completed, it was found that the daisy chain in question still showed continuity, but with a high resistance connection. This implied a cracked or poorly formed solder joint that provided intermittent "touch" contact. Such a failure was thought to be due to a defect occurring during the MCM assembly process and may have been related to a bump that was not adequately capped with solder following the solder dipping phase.

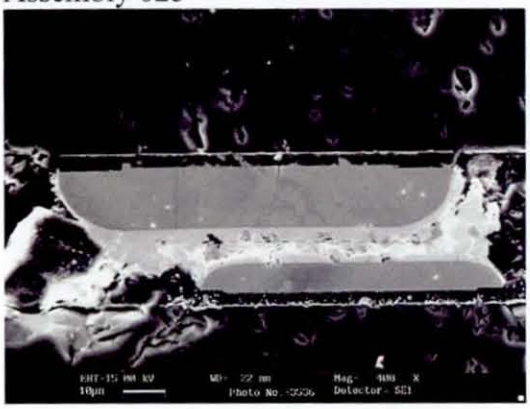
SEM images of solder joint cross-sections from specimens 15, 25, and 38 are shown at the end of this section. Sample 038 appears to have aligned well; wetting the sides of the UBM appears adequate although some areas of high contact angle are evident that may have acted as stress raisers and contributed to the failure of a joint, leading to the increased resistance measurements shown in table 8.4.2.1. No cracks were evident in the joints studied.

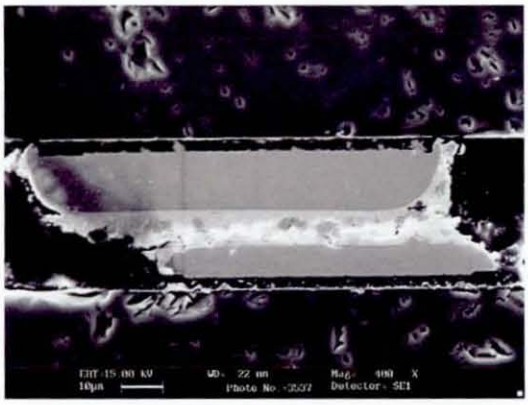
One of the daisy chains of MCM 015 failed during the power cycling, however the cross-section of the joints showed good wetting of the solder to the UBM with a rounded shape.

Two daisy chain failures were noted for MCM 025 and they both occurred at similar times. From the figures, it can be seen that the assembly was badly aligned and that the solder had failed to completely wet the UBM on the heater chip. It is therefore not surprising that this sample experienced an early failure.

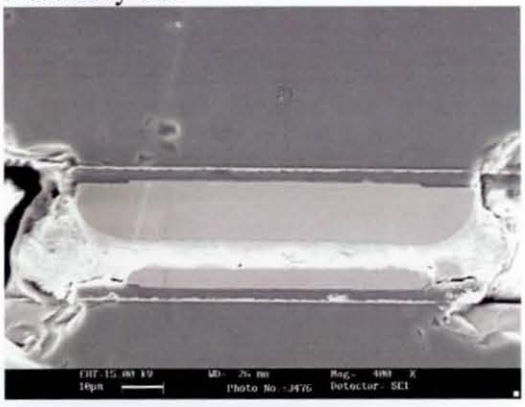
There were no additional failures other than those in the daisy chains of these samples. The resistance of all the heaters and the 4-point resistance measurements were identical to those values prior to cycling.

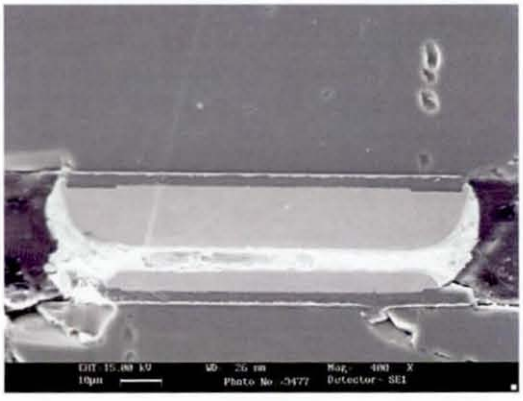
Assembly 025

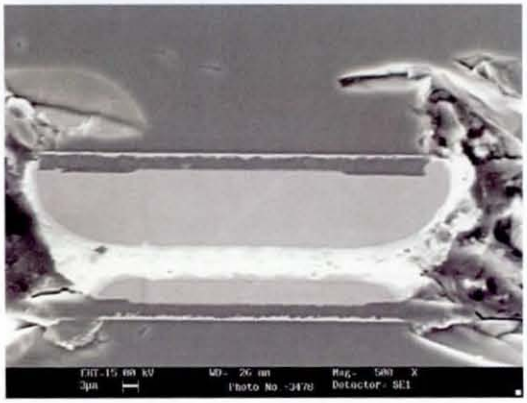




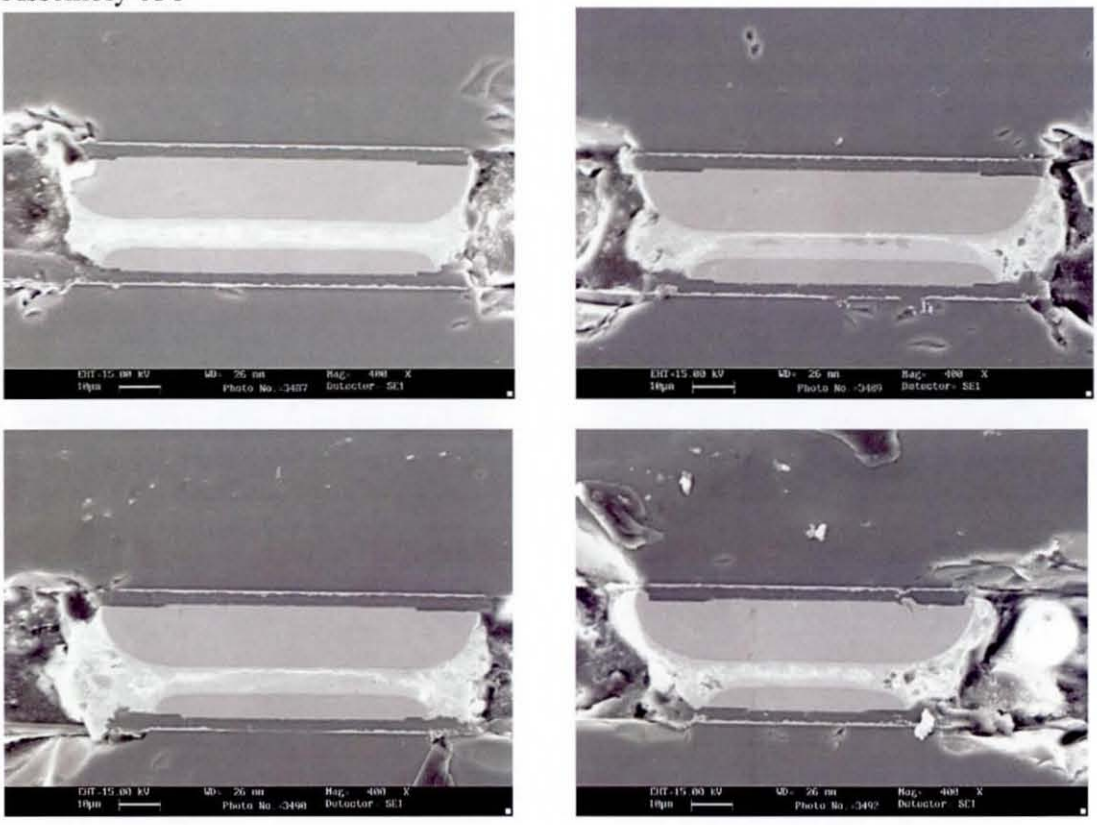
Assembly 015







Assembly 038



8.4.2.2: High Power Cycling of MCM Assemblies on Copper Substrates

As the MCM on copper substrate was known to reach a lower temperature than on an FR4 equivalent it was decided to investigate the effects of increasing the power input to some of the copper assemblies. The previous FE analysis estimated the resultant stress for the conditions to be low with respect to the yield stress, so the fatigue impact was anticipated to be minimal. As there were no further devices available, previously cycled assemblies were chosen to power cycle with a higher power level of 3.75W.

Following the initial standard power cycling, the resistance values for each assembly were checked and recorded as shown in table 8.4.2.1a The specimens that showed no increase in resistance were considered the best examples to use for the subsequent high power cycling (3.75watts or 10.06 volts at 373mA) for a period of 5662 cycles. In total there were 8 assemblies used: specimen numbers 11, 12, 19, 29, 35, 36, 37 and B5.

Thermal Results

Due to the short trial time, only one record of 4-point resistance measurements was made for the experiment after approximately 100 cycles and these results are shown with their equivalent readings from the end of the previous trial in table 8.4.2.2a. Table 8.4.2.2b shows the resistance change and the estimated temperature change. The average values are summarised in table 8.4.2.2.c

MCM number	Peak (1.2W) / Ω	Low (1.2W) / Ω	Peak (3.75W) / Ω	Low (3.75W) / Ω
011	6.983	6.275	8.469	6.248
012 b	6.743	6.122	8.007	6.086
019b	7.028	6.416	8.301	6.370
035b	6.231	5.658	7.382	5.631
036	40.05	38.59	43.12	38.51
037b	6.443	5.856	7.653	5.844
B5	7.566	6.830	8.942	6.791

Figure 8.4.2.2a: *Table of recorded 4 point resistance changes. The high powered (3.75W) results are compared with T2 results from the end of the previous standard power trial*

MCM number	Resistance Difference (1.2W) / Ω	Corresponding temp (1.2W) / °C	Resistance Difference (3.75W) / Ω	Corresponding temp (3.75W) / °C
011	0.708	33	2.221	103.4
012 b	0.621	29	1.921	89.5
019b	0.612	28.5	1.931	89.9
035b	0.573	26.8	1.751	81.6
036	1.46	68	4.61	214.7
037b	0.587	27.3	1.809	84.2
B5	0.736	34.2	2.151	100.2

Figure 8.4.2.2b*: The change in 4-point resistance and the corresponding temperature from the data in table 8.4.2.2a.*

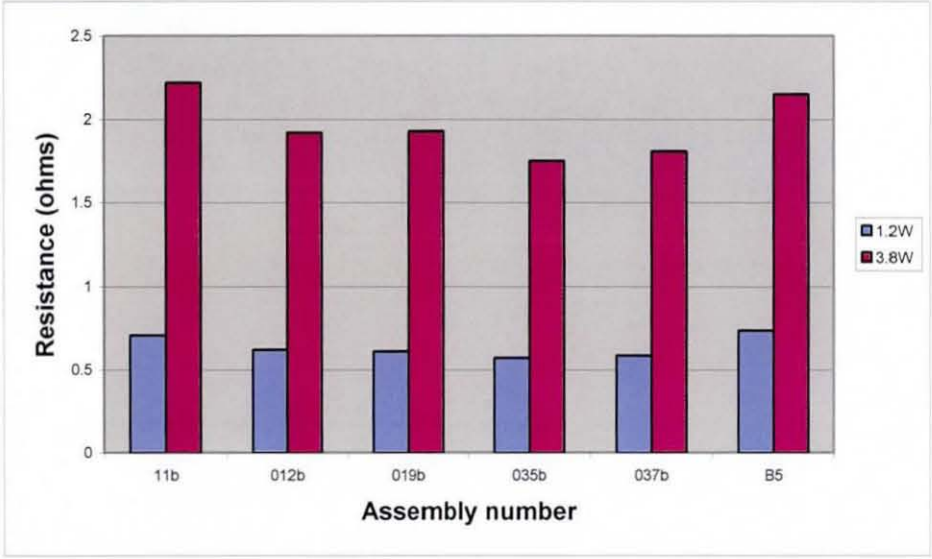


Figure 8.4.2.2a: *Chart showing the change in 4-point resistance as the power is increased from 1.2 Watts to 3.8 Watts. Sample 036 was omitted from this analysis.*

Trial	Peak / Ω	Low / Ω	Difference / Ω	Corresponding temp / °C
Original "Standard"	6.822, 0.422	6.199, 0.38	0.623, 0.053	28.986, 2.438
"High" Powered	8.126, 0.517	6.162, 0.37	1.964, 0.1699	91.47, 7.9

Table 8.4.2.2c: *Summarised average data from tables 8.4.2.1 and relevant data from 8.4.2.2 a.*

Lifetime

The final resistance readings and the number of cycles to failure are listed in tables 8.4.2.2d and e respectively. It can be seen that the failures were evenly spread out through out the trial period and in many cases the resistances of the daisy chains increased.

MCM number	Daisy chain 1 (before cycling) / Ω	Daisy chain 2 (before cycling) / Ω	Daisy chain 1 (post cycling) / Ω	Daisy chain 2 (post cycling) / Ω
011	4.3	4.0	∞	∞
012	4.3	4.0	5.9	∞
019	4.3	4.0	4.6	∞
029	4.5	4.0	4.9	4.3
035	4.6	4.1	∞	15.1
036	11.0	4.8	∞	10
037	4.3	4.0	6.1	4.4*
B5	4.5	4.3	4.8	9.3

Table 8.4.2.2d: *Resistance of daisy chains before and after high-powered cycling*

MCM number	Cycles to failure
019	669
012	838
035	1244
035	1400
012	1619
036	1643 (10 cycles)
037	2911 (161 cycles)
011	3621
011	4059

Table 8.4.2.2e: *Log of failure times*

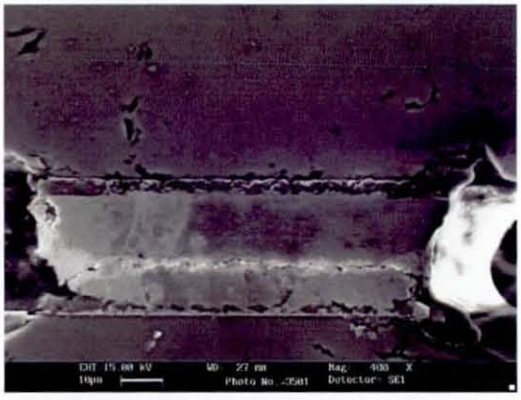
Assembly 011, both cycles failed at similar times as well. It can be seen that on each of the joints shown, one side of the UBM on the heater chip has wet adequately but the other side has not, crack propagation can be seen in the joints though they do not appear to have propagated from the apparent stress raisers. The chip alignment looks very good.

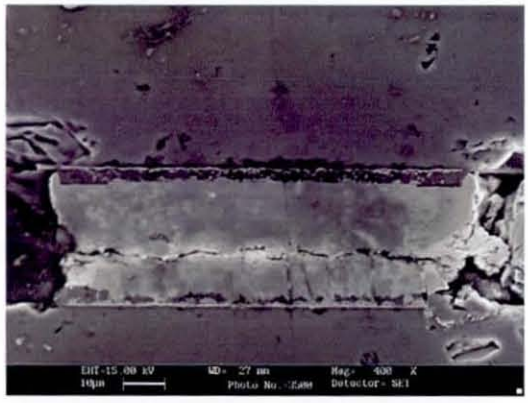
Long failures: - 36 & 37 took 10 cycles and 161 cycles to fail their interconnection joints,

Comparison of Power levels

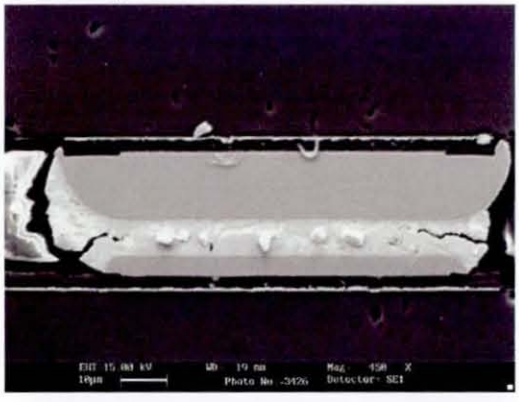
The high power temperature changes were 3.15 times those of the moderate power increase. The standard deviation is also in good agreement with this. The minimum temperature is the same.

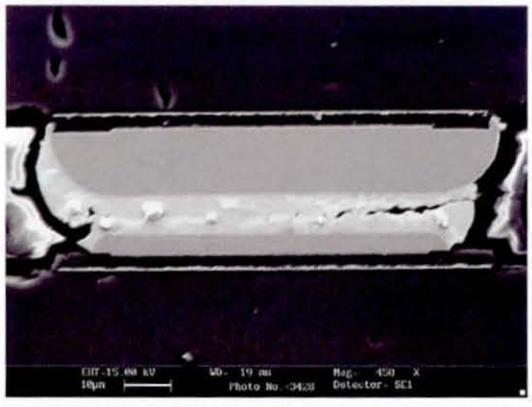
Assembly 037

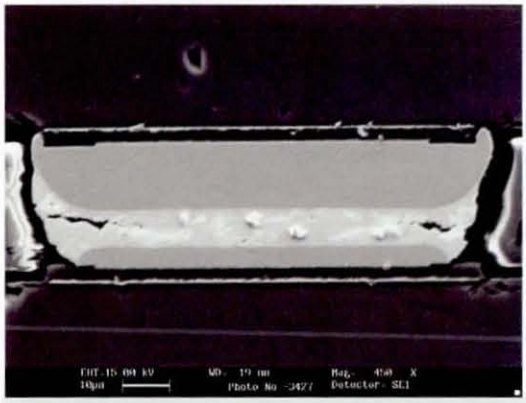




Assembly 011







8.5: Discussion

8.5.1: Comparison of MCMs on FR4 and Copper Substrates

The overall recorded failures for all scenarios are shown in figure 8.5.1. The assemblies on FR4 exhibited a large number of failures early on; 11 failures or 42% of all the daisy chains failed before 3500 cycles or within 15% of the allocated cycle time. For the remaining cycle period, there were no recorded failures and the surviving assemblies 016, 017 and 022 showed no change in resistance. In contrast, the assemblies on copper substrates had 5 failures or 19% of the daisy chains failed within the designated time and the failures were more spread out over the time period. In addition, the post cycle checks implied that a further assembly was likely to fail in a short while afterwards, had the initial power cycle trial been allowed to continue.

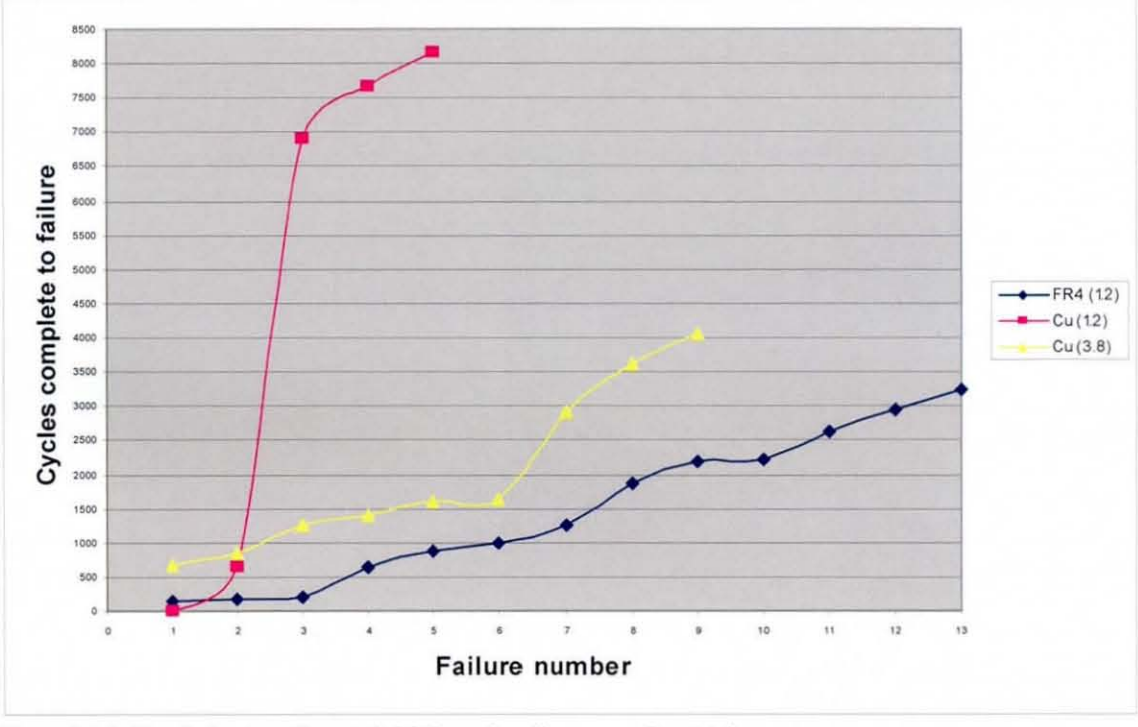


Figure 8.5.1: *Graph showing all recorded failures for all power cycling trials*

From the results, it can be seen that the assemblies on FR4 failed much sooner than those on copper, in good agreement with the FE results where a much greater stress value was predicted on the assemblies on FR4 than those on copper. However as some assemblies survived the power cycling trial it is likely that some external factor was contributing to the early failures. A possible theory was that the assemblies on the

FR4 substrate were faulty, however, table 8.5.1a shows the 4-point resistance data for all assemblies used for both FR4 and copper substrates. Also included is the resistance of the daisy chains used on each assembly. It can be seen that there is very little difference between the two as the mean values are in good agreement. These results imply that the samples used were of equivalent quality, so if the early failures on the FR4 substrates were due to faulty connections then the same number of assemblies failing prematurely should be found in the copper substrate batch. Therefore if the bad assemblies were due to inconsistencies in manufacturing quality, there is more likely to be a correspondence with the batch an assembly was originally manufactured in and its subsequent performance. Table 8.5.1b shows the assemblies original batch identifier and the recorded cycles to failure; from closer examination it can be seen that the most successful assemblies were manufactured in batch F where all three (16,17 and 19) performing exceptionally well in their respective groups. With the exception of assembly 022, batch G performed the poorest with all the other assemblies performing badly in their groups (021 & 024 lost both their daisy chains in the FR4 power cycling, as did 025 in the copper run. Also 026 exhibited the instantaneous failure)[1]

As the airflow rate was not constant across the specimen area in the wind tunnel, the assemblies positioned at the rear, may have received a poorer convection heat transfer rate and consequently reached marginally hotter temperatures than those at the front. The amplitude of the resistance change was checked for each assembly and there was no correlation between the maximum temperature reached and the assembly position in the tunnel, hence a sample was no more prone to failure depending on its position. These imply that the early failure were due to badly manufactured batches. The larger percentage of early failures on the FR4 batch was attributed to the fact that that the assemblies on FR4 were subjected to greater stress levels, therefore the stress had a faster fatigue effect on the bad joints on the FR4 batch than on the copper. Had the initial copper trial been allowed to continue for long enough, it was anticipated that the same number of failures from defective assemblies would be observed before better manufactured examples would seemingly cycle indefinitely.

[1] Number 022 was the only assembly to function correctly first time after reflow. All other assemblies were reworked and reflowed a second time.

Batch	4-point (Ω)	Heater (Ω)	Chain 1 (Ω)	Chain 2 (Ω)
FR4 (all)	6.106, 0.357	26.93, 0.072	4.33, 0.118	4.075,0.132
Copper (ecxl 036)	6.107,0.289	26.94, 0.075	4.35, 0.104	4.03, 0.084

Table 8.5.1b: *averages resistance values of the samples used, the first copper averages omits the defective assembly(036).*

Flaws in the manufacturing method were noted that may have compromised the quality of the interconnection layer. Procedures such as applying the flux before the placement were cumbersome as the flux was applied by hand using a small paintbrush. The small standoff height of the solder caps required a precise amount of flux applied in order for the interconnection layer to correctly reflow. Only eye judgement was used to determine if there was a correct amount used on each chip and the high viscosity of the flux further impeded this process. Once the flux was applied, they then had to be moved in a specimen tray to the location of the placement machine. Once correctly placed, the assemblies then had to be moved a significant distance again to the location of the reflow oven. If the flux did become less tacky then it is possible that an assembly may have become dislodged during the transit.

Assembly	Manufacturing batch identifier	Substrate	First failure (cycles)	Second failure (cycles)
A3	A	FR4	890	N/A
B5	A	Copper	>5662	>5662
D4	A	FR4	1270	N/A
4	C	FR4	2185	N/A
9	C	FR4 (cont)	N/A	N/A
11	D	Copper	3621	4059
12	D	Copper	838	1619
13	E	FR4	1000	N/A
14	E	FR4	172	N/A
15	E	Copper	647	>15 553
16	F	FR4	>22 362	>22 362
17	F	FR4	>22 362	>22 362
19	F	Copper	669	>5662
21	G	FR4	643	2209
22	G	FR4	>22 362	>22 362
24	G	FR4	195	2948
25	G	Copper	7643	8150
26	G	Copper	instant	>15 553
28	H	FR4	136	169
29	H	Copper (Cont)	N/A	>5662
30	H	Copper	6891	>15 553
31	J	FR4	1871	2618
35	J	Copper	1400	>5662
36	J	Copper	1643	>5662
37	K	Copper	2911	>5662
38	K	Copper	>15 553	>15 553

Table 8.5.1b: Illustration of assemblies' performance against the batch it was manufactured in. Assemblies 9 and 29 were the control samples therefore there was no record of cycles to failure if indeed a failure had occurred. A second failure could not be logged for some of the assemblies where the heater had failed or where one of the external connections had failed. The cycles in blue text indicate that the sample endured the initial 15 553 cycles for the standard copper trial in addition to those stated.

8.5.2: Comparison of FE Models and the Power Cycling Results

Stress Distribution: The FE models predicted even stress results over all the joints, i.e. there was no joint that was significantly overstressed. Therefore all joints were considered equally likely to fail as others. This was especially highlighted during the MCM on FR4 trials where the interconnection layer was observed to fail at other areas not on the daisy-chain (i.e. the heater or 4-point resistance measurements). For each and every solder joint for the FE model on copper substrate, the stress profile

showed high concentrations around the areas of the solder that were either nearest to or furthest away from the neutral point, where the shear effects were most apparent. From the SEM images of the joints on assembly 011, it can be seen that cracks are present in either end of the joints, though they have not fully propagated through the joint. This scenario was observed on three of the joints indicating that all the joints were subjected to similar stress levels and no particular joint was over stressed. Sample 037 also exhibited more than one crack within the interconnection layer.

Stress levels: The FE models predicted significantly higher stress/strain levels for the assemblies on FR4 substrate and the subsequent lifetime predictions implied that the assemblies on copper substrate would last a far greater number of cycles compared with those on FR4. Assuming that there was an even number of defective samples in the FR4 and copper trials, it can be seen that a greater number of samples failed during the FR4 trial than on the copper. In addition, all the failures from the FR4 batch occurred relatively early on during the trial as opposed to the case with the copper substrates where the samples were seen to fail over a greater number of cycles. The logical explanation is that the MCMs on FR4 were subjected to a greater stress level; and with the same temperature difference between the heater and carrier chips, this greater temperature difference is likely to be as a result of the local CTE mismatch as described in 7.4.3.

Ambient influences: It should be noted that the FR4 trials were started at the end of an unusually warm time period, and the FE models have shown that there is a strong dependence on ambient temperature as to the stress/strain profiles the interconnection layer will endure. For a given power level, the temperature change (ΔT) endured by the solder will be the same, however, due to the decrease in yield strength (according to graph 7.3.1.2) at higher temperatures the solder joint may exhibit significant plastic strain depending on the ambient temperature. As all the failures were recorded early on within the trial, the warmer ambient temperature may have contributed to plastic strain on the more vulnerable joints, hence the early failures observed in the trial. Had the start of the FR4 trial been postponed to a time when the temperature was cooler, it is possible that the early failures detected would have occurred over a longer time period.

From the FE results, it was noted that most of the plastic deformation occurs on the first cycle. While the recorded plastic strain on the model was very small, the strain will inevitably be larger on joints that did not form correctly. An extreme example of when this might have been the case was observed on sample 026 where a chain was seen to fail instantaneously. It was anticipated that the strain endured was sufficient to cause immediate crack propagation and therefore fracture.

8.5.3: Comparison with Other Studies

Most previous power cycling trials consisted of power cycling a single assembly and observing the cycles to failure for that device, therefore impeding comparisons with this study in terms of failure rates and distributions; however Lenkkeri and Jaakola (1) did compare performances of a given die on different substrates. While a direct comparison in terms of power levels against temperature change/cycles to failure cannot be made, the results imply that the die would perform better (i.e. smaller temperature amplitude and more cycles to failure) on the more conductive substrate in good agreement with the results of the MCM power cycling trials.

Zhang & Baldwin (2) managed to capture the stages of failure within a joint by carefully monitoring its resistance changes. They successfully managed to achieve this as the monitoring system employed allowed each joint to be probed individually such that its fine resistance changes could be detected. In addition this design also enabled them to identify which joint was the first to fail. As the monitor setup in these power cycle trials only permitted a complete daisy chain to be monitored, it was not possible to measure fine resistance changes of a single joint and a change in the daisy chain resistance was only noted when there were only a few cycles till complete failure. Furthermore, the original design of the MCM did not permit individual probing of joints, therefore it was not possible to identify the location of the failed joint until the samples had been ground and polished.

8.6: Conclusion

The power cycle trials have been completed for a number of cycles at specified power levels for the MCMs on both FR4 and copper substrates. There were several failures noted in the early part of the FR4 trial and then no subsequent casualties for the majority of the trial. For the same power level the copper batch showed fewer failures and were spread over a greater cycle period. Assuming that the quality of the assemblies was the same for both batches, this is in good agreement with the results from the FE modeling where the samples on FR4 endured a much higher stress than those on copper; and the higher stress corresponds with a shorter cycles to failure time. In addition, selected samples from the copper batch that may or may not have sustained damage from the initial power cycling trial were power cycled at a much higher level where several failures were observed as would be plausibly anticipated.

The early failures of the assemblies give rise to the theory that the local CTE mismatch between solder and UBM is a contributing factor as the FR4 batch endured a higher stress, despite the same temperature difference between the heater and carrier chips shown. In addition, the warm ambient temperature of the FR4 trial may have contributed to the early failures observed where the temperature amplitude may have been significant to instigate plastic strain on some joints, particularly where the joint quality may be defective. The information obtained of the failures was somewhat limited as only the time of failure could be recorded. As the design of the MCM did not permit individual joints to be checked it was not possible to identify the location of the failed joints so a direct comparison with the stress distributions cannot be made. However, the additional failures (on the heater and 4-point resistance) detected in the FR4 batch imply that the stress distribution is even resulting in random joints failing.

The wide range of failure times observed imply that the quality of the MCMs may not have been consistent throughout the trial. The best batch was seen to perform exceptionally well in both the FR4 and copper trials; conversely the "worst" batch had assemblies that performed poorly in both the FR4 and copper trials. As the manufacturing process required many procedures to be performed manually (by hand)

and logistics also acted as an impeding factor, it was difficult to ensure consistent quality of the assemblies.

References

1) **Jaakko Lenkkeri and Tuomo Jaakola,** "Rapid Power Cycling of Flip Chip and CSP Components on Ceramic Substrate" *Microelectronics Reliability 41, May 2001, pages 661-8*
2) **Jian Zhang and Daniel F Baldwin** "Reliability Assessment of Flip Chip on Organic Board using Power Cycling Techniques" *2002 Electronic Components and Technology Conference (IEEE), 28-31 May 2002, pages 1402-10*

Chapter 9: Conclusion

The aim of this work was to power cycle a flip chip assembly and to investigate how substrate material properties influence the thermo-mechanical attributes of the assembly for a given power level, airflow rate and cycle time. The flip chip scenario considered was a silicon on silicon multi-chip module that was mounted on either an FR4 or a copper substrate. The investigation was conducted by manufacturing a batch of assemblies and thermally profiling them and later building and simulating equivalent FE models such that the maximum temperatures, transient profiles, stress levels and the subsequent cycles to failure could all be compared.

It was first required to derive a suitable process for the chip assembly. This was implemented by first considering the necessary manufacturing processes and then evaluating different ways in which the necessary manufacturing process could be implemented. The final manufacturing process employed was considered to be unsuitable for producing a large number of assemblies, but adequate for manufacturing of the small batch required for the experiments. Following the successful manufacturing of the assemblies, some samples were chosen for thermal profiling on either FR4 or copper substrates and two different methods were used to capture the data. Using the individual techniques in isolation was found to provide limited data, however a combination of 4-point resistance measurements and thermal imaging cameras was adequate in identifying the thermal properties required. The thermal profiling of the assemblies demonstrated the importance the substrate conductivity plays in terms of the maximum temperature reached by the assembly for a specific power level.

The FE model was created by considering generic modelling criteria such as the necessary mesh density and refinement, as well as specific issues that were fundamental to correctly simulating the power cycle of a device such as the convection heat transfer. With a combination of the aforementioned subject matter and adequate material properties, it was then possible to construct the FE model, which would permit the material properties to be easily changed when deemed necessary. The FE model created allowed the power cycling of a flip chip assembly to

be simulated for given airflow conditions, power input and material properties. The model provided predictions of the steady state and transient temperatures expected to be reached by the assembly, which were validated against real devices that were tested under identical conditions. The development of an FE model means that material or geometric parameters can be readily varied and the results may be used as guidelines that may be passed on to manufacturers for recommended design improvements. Furthermore, the models have demonstrated that the transient behaviour of an assembly is governed by the mass and thermal properties of the substrate, noted from the thermal profiles of the MCM on FR4 and copper substrate. The assemblies mounted on the copper substrate had a much lower maximum temperature compared to those on the FR4, and the copper endured a much slower temperature change compared with the FR4 case. In addition, the substrates exhibited different thermal profile due to their very different conductivity's; the FR4 was only heated close to where the MCM was mounted while the copper substrate showed a uniform temperature.

Though both models showed a similar temperature difference between the heater and carrier chips for the same power level, the models on FR4 showed much greater stress levels than when the copper substrate was used. This scenario was echoed in the subsequent power cycling trials where the assemblies on copper substrate were found to have lower failure rates than those on FR4. A suitable contribution to stress developed in the joints was considered to be due to a local CTE mismatch (i.e. between the solder and the UBM/pad); potentially an important factor that is often overlooked in flip chip reliability evaluation where the dominant cause of failure is normally considered the mismatch between the die and the substrate. In the more traditional flip chip configuration his may well be the case when a silicon die is mounted on FR4 substrate and the assembly is thermal cycled, however when a silicon on silicon assembly is power cycled, the low CTE of silicon combined with the marginal temperature differences exhibited in this case places a stronger emphasis on the local CTE mismatches. In addition, for the case when an assembly is power cycled, the exact temperature of the environment can play a critical role even when large temperature changes are expected. As the yield strength of the eutectic solder is heavily temperature dependent, plastic strain may or may not occur for a given power level depending on the "start" temperature. This was demonstrated for the stress seen

in the power simulation of the 1.2W MCM assembly on FR4 cycle, when the start temperature was varied. The initial ambient temperature of 20°C was not sufficient to cause plastic deformation, however when the temperature was increased to 30°C significant plastic deformation could be observed in the solder. Conversely, for the 1.4 W power level, the initial temperature of 20°C caused plastic deformation, but when this was reduced to 10°C there was no plastic deformation detected. Therefore the performance of these assemblies was heavily dependent on the ambient conditions.

Unlike conventional thermal cycling plastic strain profiles, where the strains are continually reversing, the plastic strain history during power cycling showed that the solder joints exhibit a large amount of plastic strain on the first cycle and the accumulated strain on subsequent cycles was reduced and appeared to approach a strain limit. When power cycling "real" devices where manufacturing defects could be present it was found that there were many early failures (especially MCMs on FR4) that were thought to be caused by this initial high strain level.

The thesis has shown that power cycling a device can lead to an-isothermal temperature distributions that can cause stress to build up in assemblies that under isothermal cycling conditions would not normally be present. Further investigation is necessary to understand the effects of creep, which are known to be heavily time dependent and would require trials with longer cycle times. With the current drive following environmental legislation to eliminate lead based solders from consumer products it would also be interesting to study the effect of lead free solders on the stress distributions within the assemblies and subsequent reliability. Finally, the extension of the modelling and experimental work to the investigation of area array devices would also provide more information on the future implementation of flip chip technology for high power devices.