**Computational Fracture Mechanics of Solder Joints in an
Application Specific Integrated Circuit**

by

Brock Bolton

B.Eng., Mechanical, Carleton University

A thesis submitted to

the Faculty of Graduate Studies and Research

in partial fulfilment of the requirements for the degree of

Master of Applied Science

Ottawa-Carleton Institute for Mechanical and Aerospace Engineering

Department of Mechanical and Aerospace Engineering

Carleton University

Ottawa, Ontario

Canada

September 2002

© Copyright

2002 Brock Bolton

National Library
of Canada

Acquisitions and
Bibliographic Services

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale
du Canada

Acquisitions et
services bibliographiques

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-79691-4

Canada

**The undersigned recommend to
the Faculty of Graduate Studies and Research
acceptance of the thesis**

**Computational Fracture Mechanics of Solder Joints in an
Application Specific Integrated Circuit**

submitted by

**Brock Bolton
B.Eng., Mechanical, Carleton University
in partial fulfilment of the requirements for
the degree of Master of Applied Science**

Thesis Supervisor

Chair, Dept. of Mechanical and Aerospace Engineering

**Carleton University
September 2002**

Abstract

An application specific integrated circuit (ASIC) is analyzed in three-dimensions using FEM techniques for thermal and stress loading, crack initiation and growth.

A multigrid technique is used to analyze the entire ASIC, and analyze one “critical” solder ball by the Coffin-Manson equation. This solder ball has contact elements placed along the cracking plane with special stiffness formulation. A stress analysis is performed, and the results used as Dirichlet boundary conditions on the boundary of a “crack object” embedded within the solder ball. It is analyzed using J-integral software, and an equivalent stress intensity factor is calculated. A Paris-Erdogan equation used to predict crack prorogation.

For a 0.0001% probability of failure, the Coffin-Manson equation predicts that the first solder ball will fail after approximately 3475 cycles, while a crack is predicted to take 286 cycles to propagate from the edge of the solder ball to the center.

To my parents, for their undying support.

Acknowledgments

I would like to thank my thesis supervisor, Prof. John Goldak for help I received when performing my research for this thesis, and allowing me to complete my master's studies here at Carleton.

I would also like to thank the entire MMO team of programmers and software developers, for without their help, this thesis would surely not be possible. Special thanks goes out to Adrian Vasilescu, Stanislav Tchernov, Arif Baig, and Shaodong Wang for the incredible patience and skills you brought to this thesis.

The financial support of Carleton University, Nortel, and OGS is also gratefully acknowledged.

Of course, no acknowledgment section would be complete without a heartfelt thank you to my family and friends, who had the misfortune of having to put up with my ranting and raving over the past 2 years of my life while this project was completed.

To everyone mentioned here, I am in debt to all of you.

Contents

Acceptance Sheet	ii
Abstract	iii
Dedication	iv
Acknowledgments	v
Table of Contents	vi
List of Figures	xi
List of Tables	xix
1 Introduction to Electronic Packaging	1
1.1 Introduction	1
1.2 Packaging Techniques	2
1.2.1 First Level Packaging	2
1.2.2 Second and Third Level Packaging	4
1.3 Surface Mount Technology Issues	6

1.3.1	Solder Joint Formation and Failures	7
1.3.2	Overview - Solder Property Theory	12
1.3.3	Solder Fatigue Life Theory	14
2	FEM Thermal Analysis Theory	20
2.1	Heat Transfer Methods	20
2.1.1	Conductive Heat Transfer	20
2.1.2	Convective Heat Transfer	21
2.1.3	Radiation Heat Transfer	23
2.1.4	Differential Equation For Heat Conduction	24
2.2	Initial and Boundary Conditions	27
2.3	Discretization	29
2.4	General Finite Element Formulation - Energy Solver	31
2.4.1	Bubnov-Galerkin Finite Element Method	31
3	Plasticity Theory	38
3.1	Stress Strain Theory - Elastic and Plastic Deformations	38
3.1.1	Stress Space and the Yield Function	42
3.1.2	Plastic Flow Theory	47
3.2	Hardening	50
3.2.1	Isotropic Hardening	52

3.2.2	Kinematic Hardening	54
3.2.3	Viscous Effects	55
4	FEM Stress Analysis Theory	61
4.1	Finite Element Method for Stress	61
4.1.1	Continuum Mechanics Review	61
4.1.2	Virtual Work Equation	66
4.1.3	Elastic Stress Analysis	68
4.1.4	Elastic-Plastic Stress Analysis	70
4.1.5	General Plasticity Theory - FEM Formulation	71
4.1.6	Infinitesimal Strain J_2 Flow Theory - FEM Formulation . . .	76
5	Theory of Fracture	83
5.1	Deformation Mechanisms	83
5.1.1	Brief Dislocation Definition	83
5.1.2	Elastic Collapse	84
5.1.3	Low Temperature Plasticity	84
5.1.4	High Temperature Plasticity	85
5.2	Microscopic Aspects of Fracture	88
5.2.1	Brittle Cleavage Fracture	90
5.2.2	Brittle Intergranular Fracture	90

5.2.3	Ductile Fracture	91
5.2.4	Shear Fracture	94
5.2.5	Ductile-Brittle Transition	95
5.2.6	Fatigue Cracking	95
5.3	Fracture Mechanics Review	96
5.3.1	Fully Elastic Treatment	97
5.3.2	Elastic-Plastic Treatment	102
5.3.3	J-integral derivation	103
5.3.4	Evaluation of J-integral terms	112
5.3.5	T* Integral	114
6	Software Development	116
6.1	Introduction	116
6.2	J-integral Software Testing	117
6.2.1	Testing Requirements	117
6.2.2	Testing Methods	117
6.2.3	Testing Results	124
6.2.4	Figures for J-integral Testing	128
6.3	Rigid Body Rotation Software Testing	133
6.3.1	Motivation	133
6.3.2	Proof of Stress Rotation	134

6.3.3	Expected Results - A to B rotation	136
6.3.4	Expected Results - B to A rotation	137
6.3.5	Test Procedure	139
6.3.6	Test Results	142
7	Thermal, Stress, and Cracking Analysis of Dos 100 ASIC	150
7.1	Special FEM Methods	150
7.1.1	A General Multigrid Method	150
7.2	Description of Domain and Problem	153
7.3	Thermal Analysis	158
7.3.1	Procedure and Results	158
7.4	ASIC Stress Analysis	160
7.4.1	Procedure	160
7.4.2	Material Properties	162
7.4.3	Results	166
7.5	Solder Ball Stress Analysis	170
7.5.1	Procedure	170
7.5.2	Results	172
7.6	Solder Ball Cracking Analysis	174
7.6.1	Procedure	174
7.6.2	Results	189

8	Conclusions and Discussion	197
8.1	Discussion of Results	197
8.2	Suggested Future Work	204
	Bibliography	206

List of Figures

5.1	Notation for the T^* integral, from [6, pg. 538].	114
6.1	Infinite plate with collinear cracks (from [5, pg. 82]).	119
6.2	The “infinite” plate mesh on the left, with embedded refined cracked mesh on the right.	121
6.3	Coordinate system for mode I cracking [40, pg. 31].	122
6.4	Embedded refined mesh showing the elements taken in the first (red and green), third (dark blue and yellow), and fifth (smaller purple and light blue mesh) contour for J-integral evaluation, along with half of the finest embedded refined cracked mesh (larger purple and light blue mesh).	124
6.5	Close-up view of the through 41.67cm crack with displacement magni- fied $500\times$ for emphasis.	128
6.6	Z-displacement of the plate mesh, showing that the entire width is in plane strain.	129

6.7	Asymptotic solution for an infinite plate, (6.8), extrapolated to $r \rightarrow 0$ for 8.33cm crack.	129
6.8	Asymptotic solution for an infinite plate, (6.8), extrapolated to $r \rightarrow 0$ for 25.0cm crack.	130
6.9	Asymptotic solution for an infinite plate, (6.8), extrapolated to $r \rightarrow 0$ for 41.67cm crack.	130
6.10	Comparison of three solutions for 8.33cm crack.	131
6.11	Comparison of three solutions for 25.0cm crack.	131
6.12	Comparison of three solutions for 41.67cm crack.	132
6.13	Rotation from configuration A (cube on left) to configuration B (cube on right) showing the magnitude of displacement.	144
6.14	Rotation from configuration A (cube on left) to configuration B (cube on right) showing X-displacement.	144
6.15	Rotation from configuration A (cube in front) to configuration B (cube behind) showing Y-displacement.	144
6.16	Rotation from configuration A (cube on left) to configuration B (cube on right) showing Z-displacement.	144
6.17	Original plate mesh in natural crack coordinates.	148
6.18	Plate mesh rotated using rotation matrix $\mathbf{R}$	148
7.1	The ASIC mesh used in this analysis.	155

7.2	Detail of the pedestal underneath the heat sink.	155
7.3	Detail of the underlying layers underneath the heat sink and pedestal.	155
7.4	Detail of the solder ball layer in the coarse ASIC mesh.	155
7.5	ASIC with part numbers labeled.	156
7.6	The temperature distribution in the ASIC after 90 seconds of heating.	159
7.7	The temperature cycle used for the stress analysis (actual temperatures shown are for the maximum temperatures in the solder balls). This is similar to a prescribed furnace temperature pattern for an accelerated test. This distribution was obtained at the Gauss point within the solder ball with the maximum effective plastic strain amplitude. . . .	160
7.8	Side view of the package after one temperature cycle. Note warping in the package (magnified 500 times for emphasis).	161
7.9	Viscosity ($Pa \cdot s$) values versus temperature (K) for the solder used in the analysis.	163
7.10	Yield stress (Pa) values versus temperature (K) for the solder used in the analysis.	163
7.11	Young's modulus (Pa) values versus temperature (K) for the solder used in the analysis.	164
7.12	CTE ($1/K$) values versus temperature (K) for the PCB material FR4 used in the analysis.	164

7.13 CTE ($1/K$) values versus temperature (K) for the substrate material PTFE used in the analysis.	165
7.14 Failure distribution predicted by the Coffin-Manson equation. Note red shows lowest CTF (most likely to fail), blue shows highest CTF (least likely to fail).	167
7.15 Domain chosen for analysis of critical solder ball. Shown is the PCB, solder ball array, and substrate from the entire ASIC (shown in purple, light blue and yellow, respectively), with the critical solder ball mesh location.	169
7.16 Refined solder ball mesh.	169
7.17 Critical solder ball showing location of maximum effective plastic strain (location is furthest from package center, shown in red, at solder ball and PCB interface). Displacements magnified 500 times for emphasis.	170
7.18 Effective plastic strain plot at each timestep for the finest refined sol- derball. This plot was obtained at the Gauss point within the solder ball with the maximum effective plastic strain amplitude.	172
7.19 Effective strain plot at each timestep for the finest refined solderball. This plot was obtained at the Gauss point within the solder ball with the maximum effective plastic strain amplitude.	173

7.20	Effective stress plot at each timestep for the finest refined solderball. This plot was obtained at the Gauss point within the solder ball with the maximum effective plastic strain amplitude.	173
7.21	The solder ball mesh with contact elements between part numbers 6 (yellow) and 7 (purple).	175
7.22	Photo of an actual solder ball with cracking near the substrate and solder ball interface [51].	175
7.23	Specified stiffness function for uncracked portion of solderball with contact elements (see Table 7.5 for uncracked parameter values). . . .	177
7.24	Resulting force between contact elements with uncracked stiffness for- mula specified (see Table 7.5 for uncracked parameter values).	178
7.25	Specified stiffness function for cracked portion of solderball with con- tact elements (see Table 7.5 for cracked parameter values).	178
7.26	Resulting force between contact elements with cracked stiffness formula specified (see Table 7.5 for cracked parameter values).	179
7.27	Geometry for contact cracking.	180
7.28	Cracking surface on part number 6.	183
7.29	Closeup of the $20\mu m$ contact crack created in the contact solder ball mesh.	185

7.30	Closeup of the $100\mu m$ contact crack created in the contact solder ball	
	mesh.	185
7.31	Closeup of the $50\mu m$ contact crack created in the contact solder ball	
	mesh.	185
7.32	Closeup of the $200\mu m$ contact crack created in the contact solder ball	
	mesh.	185
7.33	Half of contact solder ball with full $20\mu m$ crack object embedded within	
	it, showing orientation.	186
7.34	Half of contact solder ball with full $100\mu m$ crack object embedded	
	within it, showing orientation.	186
7.35	Half of contact solder ball with full $50\mu m$ crack object embedded within	
	it, showing orientation.	186
7.36	Half of contact solder ball with full $200\mu m$ crack object embedded	
	within it, showing orientation.	186
7.37	$20\mu m$ crack object mesh, in unrotated coordinates (note coordinate	
	axes).	188
7.38	$20\mu m$ crack object mesh, in rotated coordinates (note coordinate axes).	188
7.39	$20\mu m$ crack object mesh, in unrotated coordinates (note coordinate	
	axes).	188
7.40	$20\mu m$ crack object mesh, in rotated coordinates (note coordinate axes).	188

7.41	J-integral data plotted as equivalent stress intensity factor K , for each timestep for the $20\mu m$ crack length (blue dots are data points, black line is a line of best fit).	194
7.42	J-integral data plotted as equivalent stress intensity factor K , for each timestep for the $50\mu m$ crack length (blue dots are data points, black line is a line of best fit).	194
7.43	J-integral data plotted as equivalent stress intensity factor K , for each timestep for the $100\mu m$ crack length (blue dots are data points, black line is a line of best fit).	195
7.44	J-integral data plotted as equivalent stress intensity factor K , for each timestep for the $200\mu m$ crack length (blue dots are data points, black line is a line of best fit).	195
7.45	ΔK values for varying normalized crack lengths (blue dots are data points, black line is a line of best fit).	196
8.1	Effective stress at the bottom of the solder ball, where cracking analysis was to take place.	198
8.2	Effective stress at the bottom of the solder ball, with a contact crack length of $200\mu m$	199

8.3	Varying the Weibull exponent in the modified Coffin Manson equation for 0.01% probability and 0.0001% probability. Data shown is for the finest solder ball mesh.	201
8.4	Cycles to failure predicted by the Coffin Manson equation for coarse mesh and fine mesh.	202

List of Tables

6.1	Aluminum 380 Material Properties	118
6.2	J-integral values - Time step 1 (per unit length)	126
6.3	J-integral values - Time step 2 (per unit length)	126
6.4	J-integral values - Time step 3 (per unit length)	127
6.5	K values - Time step 1 (All K values in $Pa\sqrt{m}$)	127
6.6	K values - Time step 2 (All K values in $Pa\sqrt{m}$)	127
6.7	K values - Time step 3 (All K values in $Pa\sqrt{m}$)	127
6.8	J-integral values (per unit length)	148
7.1	ASIC part numbers and part types	157
7.2	Solder Material Properties	162
7.3	Coffin Manson CTF Results (Coarse Mesh)	168
7.4	Coffin Manson CTF Results (Coarse Mesh)	171
7.5	Contact stiffness parameter values	177
7.6	Paris-Erdogan constants for solder from [27]	190

7.7	η and ΔK values for varying crack lengths	192
-----	--	-----

Chapter 1

Introduction to Electronic Packaging

1.1 Introduction

In recent years, the explosive boom in electronics has caused an ever increasing demand on the packaging to house this technology. This has caused engineering in the packaging industry to devise new methods to package these new technologies. With this comes new problems and the need for a way to predict performance of such components. As electronics become smaller and their performance increases, the need to manage the heat these packages create becomes ever more crucial to their reliability.

Essentially, electronic packaging in its simplest form allows a central processing unit (CPU) to communicate with memory and input/output (I/O) units, such as monitors, keyboards, and the like [46]. There are three main levels of packaging:

First level packaging: individual chips or integrated circuit (IC) bonded to a substrate, as part of single chip modules (SCMs) or multichip modules (MCMs),

Second level packaging: components (such as SCMs, MCMs, connectors, discrete components such as resistors or capacitors) which when connected to a printed circuit board (PCB) comprise the second level,

Third level packaging: PCB assemblies, associated cables, cooling systems, cases, and such, when interconnected comprise the third level.

1.2 Packaging Techniques

1.2.1 First Level Packaging

Three common ways by which ICs are connected to PCBs are wire bonding, flip chip attach, and tape automated bonding. The oldest technology of the three is wire bonding, in which the silicon die is bonded to the substrate (which in turn is bonded to the PCB) backwards using an organic adhesive or a metal to metal bond, with its leads facing away from the board. The bonding wire is then attached to this

outer surface and connected to the substrate, with typical wire sizes in the order of 0.025mm to 0.05mm [46]. Typically these assemblies are encapsulated in an epoxy for environmental protection. Modern assembly technologies can place each IC and bonding wire within 0.1mm of one another.

Flip chip attachment simply implies that the leads of the IC silicon die are flipped and face the substrate. This technology allows the highest interconnection density for a given area with a specified performance. The solder bumps on the flip chip are placed on the wettable metal terminals, and the flip chip is aligned to the substrate, where there is a matching “footprint” of solder wettable terminals. The solder is then reflowed (meaning it is heated again above the melting point) and the connection made [23].

Ball grid arrays, or BGAs, are used when a high pin count density is needed, and is typically used in environments that require robustness, such as modern personal electronics. The assembly of a BGA is similar to that of flip chip, but on a larger scale. The use of a BGA allows a designer design flexibility, enhanced performance, a low amount of assembly defects, and a reduced cost [46]. BGAs can have ceramic or plastic carriers, and can be denoted in literature as CBGAs or PBGAs, respectively.

Using this technology is dependant upon the substrate material’s metallurgical properties since the solder reflowing when attaching requires the substrate to withstand these temperatures. Organic substrates (such as polyamide or epoxy-glass FR4)

have difficulty with this requirement, and recently low temperature flip chip technology is being aggressively developed. One such concept allows the use of a eutectic solder on the substrate side to provide “counterbumps” for the solder bumps on the chip [46]. This way, the counterbumps are reflowed and create a joint with the solder on the chip side. There are many other concepts and ideas being introduced on a rapid basis with emphasis on low cost and high density of packing.

Tape automated bonding is similar to wire bonding, in the sense that instead of wires flat metal fingers provide the means of connection between IC and substrate. It also allows the option of using a flip chip concept, with the leads facing towards the substrate. It should be mentioned that with flip chip setups, it is possible to implement heat sinks above the chips to dissipate the heat created from them.

1.2.2 Second and Third Level Packaging

Conventional technology for packaging has the ICs mounted to the PCB via pin-through-hole (PTH) technology, where as the name implies, the IC is mounted to the board via a wire which is inserted into pre-drilled holes and soldered in place. Recently, the demand for higher density packaging has given way to surface mount technologies (SMT), where the IC is directly mounted onto the PCB.

A typical PCB consists of alternating layers of copper and resin-impregnated glass cloth, with the layers interconnected with plated through-hole vias. These vias

aid in the electrical interconnection of components mounted on the PCB. There are three main types of vias, plated-through-holes (PTH), blind vias, and buried vias.

PTH vias simply go through the entire PCB, connecting the top and bottom layer, and is attached to one or more of the copper layers within the PCB. These are used when connecting PTH components to the PCB, as the name would imply, and have a typical diameter of about 30-42mm. Buried vias are not accessible from the external surfaces of the PCB, and are mainly used to interconnect the layers of copper within the board. Blind vias are used to connect top or bottom surface layer to the next copper layer below the surface. There are many chances for defects to occur in manufacturing in the PCB due to plating deficiencies.

The lines on the PCB, or traces, are used to distribute the power or signals though the board, and can be produced using a subtractive etching process or a additive plating process. Essentially, the etching process builds the layers of the PCB up and then etches away areas where traces are not needed, whereas the additive process uses a pattern of the traces on the PCB to allow for copper traces to be plated into this pattern.

Once the traces are established, a solder mask is applied to these exposed copper areas everywhere except where solder is to be applied. The exposed copper where solder is to be applied must be protected from oxidation due to the fact that oxidized copper does not allow good wetting of solder [46]. A common practice is to allow a

thin coating of an organic material or flux to prevent oxidation that melts away upon soldering to expose the unoxidized copper to the flowing solder.

Once the first level packaging is made, and the PCB prepared, next is the issue of physically connecting the two together. Sockets are used where it is desirable to detach components from the PCB for a particular reason (i.e., maintenance, upgrading, etc.). Electrical contact is a concern with sockets, since variation of physical clamping force and oxidation of contact surfaces may degrade the electrical properties of the connection. However, if the component to be mounted is not to be replaced more than once, then soldering becomes a cost effective alternative. The trade off with the solder joint is the increased risk of solder joint unreliability and the heat effects of the reflowing of the solder, especially with newer multi-pin components. As well, soldering may eliminate special heat sink mounting, if a heat sink is to be used in assembly.

1.3 Surface Mount Technology Issues

SMT packages are becoming ever more popular, as designers strive for higher density packaging, increasing the power of packages, while decreasing their size. Thus, with this increase in popularity comes the importance of the solder joint between substrate and PCB. Engineers designing such SMT packages must ensure that these joints perform reliably for as long as the product is to be used. Some critical problems with

SMT are the coefficient of thermal expansion (CTE) mismatch between the substrate and the PCB (or die if such is the case), the environment which the package will be exposed to in service, and the acceptable failure probability of the package. Since the solder provides both a electrical and a physical connection for the package, it is of utmost importance to be able to model its behaviour.

1.3.1 Solder Joint Formation and Failures

1.3.1.1 Solder Paste

In SMT packages, the common method of creating the solder joints is creating a print with solder paste, placing the component in the desired location, and then reflowing the solder at an elevated temperature. Typically, in high density SMT applications, this is applied with a squeegee over a stencil of the desired solder pattern, under controlled conditions. Pastes that resist solder balling are required as density of packaging becomes ever higher. Solder balling occurs when tiny bits of solder scatter from the intended joint and stick to surrounding surfaces.

Solder paste is composed of three main parts, the solder powder, the flux, and the vehicle [18]. The only component performing the actual bonding is the solder powder. The vehicle acts as a carrier for the powder, whereas the flux cleans the powder and the parts to be joined so that a good metallurgical bond can be achieved.

Fluxes can come in many forms, but the key purpose is the same with any flux,

to clean the surfaces which are to be bonded. Flux can be incorporated into the solder paste, as described above, or applied directly as a liquid to the component to be soldered.

1.3.1.2 Soldering Methods

Formation of solder joints can take on many methods [31]. There are three main types, conduction, convection, and radiation processes.

Conduction processes, such as a soldering iron or hot bar, require the user to physically contact the solder with a heating device to reflow the solder. Typically this method is used for repair or low volume assembly, since it requires a skilled human input.

Convection processes can use hot gases, vapour phase soldering, or wave soldering can be employed. Vapour phase soldering involves the use of a saturated heated vapour, and the parts to be soldered are passed into this vapour. The vapour condenses on the relatively cooler parts, and reflows the solder via heat transfer. Precise temperatures can be achieved with this method. Wave soldering involves the use of a liquid solder flux applied to where the joints are to be formed, usually by a conveyor. The parts are preheated to avoid thermal shock, and then conveyed to the solder wave, where the soldering is formed by direct contact of the part with the liquid solder. The flux allows the solder to bond only on the surfaces with flux applied.

Convection processes are mainly used when a uniform temperature profile is desired.

Radiation processes use emitted infrared (IR) energy, and fall into three main categories. The first is lamp IR systems, which use large amounts of electricity and have high emitter temperatures, which can cause overheating of parts being soldered, and thermal shadowing can occur. The second is panel natural convection IR systems. These use natural convection by the heated air in the chamber and absorption of radiant energy from the emitters. Panel natural convection processes can eliminate the shadowing problem with the use of natural convection. Finally there is forced convection IR, which is similar to panel natural convection IR processes, but at the name implies, uses forced convection instead of natural convection. The air used in the forced convection can have its temperature regulated so it can cool the parts being soldered, preventing damage. Radiation processes are the most versatile of the three methods mentioned above, combining the best volume, versatility, and lowest cost.

Another more precise method is that of laser soldering, where a tiny laser of high intensity light is focused for short periods of time on the solder to reflow it. This has the advantage of being very precise with application of the heat, and with its short application times, tends to be less damaging for surrounding parts. This application is gaining in popularity with the coming of higher density packages which require more precise soldering, although the cost is quite high. Due to this high cost factor, it is typically only used in specialized applications where no other process can

deliver the required quality.

1.3.1.3 Solder Voids

A common defect introduced in solder joints is that of voids. Voids are a result of gaseous bubbles being released into the solder during wetting of the solderable surfaces. Once the bubbles are trapped upon solidification, they are now voids. Flux breakdown, residues, or surface contaminants can be sources of these bubbles.

Solder pastes contain by volume up to 65% of volatile compounds in the form of fluxes and solvents [46]. Since these compounds decompose upon reflowing, these are a major source of voids. Reference [46] also states that metal content in the paste, reflow process temperatures, the solvent boiling point, and the reflow environment all have a role in void formation. The common trend is a higher metal content in the paste tends to yield more voids, due the increase in oxides, and the resultant gaseous bubbles upon reflowing. As well, the metal particles retard the bubbles from escaping the solder, compounding the problem. Also, if a reflow temperature is low, solvents with higher evaporation temperatures will not evaporate, and tend to be retained as bubbles in the solid solder.

In BGAs specifically, such as the package analyzed in this thesis, voids in the solder ball itself are a major concern, since it is both the electrical and mechanical connection. Recall that in leaded packages the lead is the main electrical connection.

The effect of voiding in a plastic BGA array has shown that voids in up to 24% of the pad area have little or no effect on joint reliability. It also has been shown that the joints with voids actually retard crack propagation [46], so there actually may be an increase in reliability with the inclusion of voids in BGAs.

1.3.1.4 Intermetallics

Intermetallics are an intermediate phase of stoichiometric composition formed by the reaction of two or more metallic elements with physical and chemical properties different from those of the constituent elements [46, pg. 201]. These form at the interface of the solder and the wetted surface the joint is to be formed on. Constituents diffuse from one surface to the other and react with each other. Intermetallics play a crucial role in the reliability of solder joint formation, and form the metallurgical bond between surfaces to be joined.

The strength of a solder joint is dependent on a delicate balance of the intermetallic layer. Too thin a layer will create a weak joint, whereas too thick a layer will also create a weak joint. Joints made to copper pads tend to have a decrease in tensile strength by a thick intermetallic formation. If a wetted surface gets excess oxides upon it, it will tend to have a thinner intermetallic layer than one with less oxides. This is due to the oxide layer retarding the intermetallic reaction. As well, many studies [46, pg. 205], [45] show that cracking in a solder joint tends to propagate

along the interface of this intermetallic layer and the bulk solder, starting at a high stress location. This is shown in many studies, where a cross section of the cracked surface shows a thin layer of solder still attached to the wetted surface, separate from the bulk solder.

Reference [45] notes that during typical temperature cycles, the intermetallic thickness can change. They also found that a smooth intermetallic layer can lead to a shorter joint life versus a more uneven intermetallic layer. This is due to the poor ability to resist shear, and the difficulty in crack propagation. Typical thicknesses for intermetallics from eutectic or near-eutectic solder and a copper pad can range from less than one micron ($1\mu m$) to about 10 microns.

1.3.2 Overview - Solder Property Theory

The solder joints in SMT packages are among the most researched topics in electronic packaging. Their behaviour is among the most critical in package reliability, yet it is among the most complex materials to understand. At typical operating temperatures in the range of 50-100 degrees Celsius, most solders are at high temperatures relative to their melting point (i.e., high homologous temperatures). Due to thermal cycling of power cycles within the package, the solder joint can experience plastic deformation as well as creep due to the high temperatures.

Commonly, Sn-Pb eutectic solder (63Wt%Sn-37Wt%Pb) is used to bond com-

ponents to the PCB, be it using SMT or PTH technology, and quite often, many failures occur at this level due to soldering practices. One convenient property of using eutectic solder is its low melting point. For 63/37 solder, the melting point is 183 degrees Celsius. Many reasons, such as fluxes used in soldering, residual stresses due to solder reflow, absorption of cleaning fluids by PCB resins can all contribute to failures [46]. Ball grid arrays (BGA) are one type of SMT that implement the same concept as flip chip ICs to attach the substrate to the PCB. These solder balls are typically larger than the ones used in the flip chip assembly, but are typically attached using similar methods. Like the solder joints in the flip chip, these solder balls provide both physical and an electrical contact between the substrate and the PCB.

At room temperature, eutectic or near-eutectic solder is a very ductile material. It has been shown that a 60Wt%Sn-40Wt%Pb wire can be stretched at room temperature 60% before breaking [24]. Thus, when stresses are applied to a solder joint, they can yield, and do so in a way to reduce the stress. If a strain is constant, yet the stress decreases, the term “stress relaxation” is used. When the stress is constant, and the strain changes during a time period, the term “creep” is used. In both cases plastic flow is present, and on a microstructure scale, the processes are the same [24]. In field use, there is typically a time-dependent change in both strains and stresses simultaneously, making it a combined process of “creep-relaxation”.

Thus, mechanical behaviour of typical Pb-Sn solder joints under thermal cycling can be very difficult to characterize. Depending on the temperature it is operating at, the mode of deformation can be elastic, plastic, creep, or as discussed above, a simultaneous “creep-relaxation” can occur. Creep is especially true, since at room temperature, 63Wt%Sn-37Wt%Pb is above half its absolute melting point, and creep becomes dominant at temperatures above half the melting temperature ($T \geq T_m/2$).

1.3.3 Solder Fatigue Life Theory

In the majority of fatigue failures of solder joints, the primary cause is low cycle fatigue. This failure typically occurs with cracks within the solder joint, caused by thermal cycling. As well, since the typical operating temperature of a solder joint is at a high homologous temperature, creep cracking can also play a role in solder joint life. Where these cracks occur and propagate is determined by the stress and strain distribution within the joint, and for different classes of joints (BGA, flip chip, peripheral leadless, gull-wing leaded, etc.) cracks tend to form and propagate in similar locations and directions. However, many studies have been made to try to predict the life of a solder joint.

1.3.3.1 Reliability Testing and Statistics

Testing for reliability of electronic packaging should, in theory conduct tests in as-used conditions. However, with increasing demand comes the need for accelerated testing, with higher magnitudes of force than would be seen in service in an attempt to determine by statistics the probability of failure under service loading. As well, the number of samples for testing produced should be such that they represent a good approximation of the actual failure distribution in service, without making costly excess samples. Finally, the modes of failure in the accelerated testing should be the same as those in service, if the data collected is to have any meaning or accuracy. A reliable acceleration test for solders has yet to be established [23]. In all the above testing factors, it is a general trade off between cost (in the form of testing time and sample creation) and accuracy.

For life prediction, some form of a statistical distribution should be used to characterize the life distribution. The most common one used in electronic packaging is a Weibull distribution. This distribution is characterized by the following equations. The two-parameter (in this case θ and β) Weibull probability density function (PDF) is:

$$f(x) = (\beta/x)(x/\theta)^\beta \exp(-(x/\theta)^\beta) \quad (1.1)$$

The Weibull cumulative density function is:

$$F(x) = 1 - \exp(-(x/\theta)^\beta) \quad (1.2)$$

The Weibull reliability, or fraction of survivors, function is:

$$R(x) = 1 - F(x) = \exp(-(x/\theta)^\beta) \quad (1.3)$$

The Weibull failure rate function is:

$$h(x) = \frac{f(x)}{R(x)} = (\beta/x)(x/\theta)^{\beta-1} \quad (1.4)$$

where:

β is the shape parameter, the Weibull curve slope,

θ is the characteristic life at which a predetermined percentage of the population has failed,

x random variable, in this case, the life of a part in question,

Generally, it is found that two-parameter distributions tend to be conservative [25]. Thus, the above equations can be used to analyze failure phenomena. The parameter β plays a crucial role in the failure rate distribution. For $\beta \leq 1$, the Weibull PDF approaches infinity as time approaches zero, and approaches zero as time increases. The failure rate also behaves the same way. When $\beta = 1$, the failure

rate is constant, and when $\beta = 2$, the failure rate increases linearly. For $3 \leq \beta \leq 4$ the failure rate increases rapidly.

1.3.3.2 Strain Life Approach

The strain life approach uses a Coffin-Manson type equation for cyclic fatigue life of surface mount solder joints. As the name implies, it relates the cyclic plastic strain range to the mean cyclic fatigue life. The typical form of the equation is:

$$N_f = \frac{1}{2} \left[\left(\frac{\Delta \varepsilon_p}{2} \right) \frac{1}{\varepsilon'_p} \right]^{1/c} \quad (1.5)$$

where:

N_f is the number of cycles to failure, defined as an initiation of a crack,

$\Delta \varepsilon_p$ the effective cyclic plastic strain range,

$\Delta \varepsilon_p / 2$ the effective cyclic plastic strain amplitude,

ε'_p is the fatigue ductility coefficient,

c is the fatigue ductility exponent.

Reference [10] suggests that a relation where cyclic plastic strain range is used, as well as taking into account temperature and time in the equation itself. The following equation was proposed for 63Wt%Sn-37Wt%Pb solder joints undergoing cyclic loading:

$$N_f(x\%) = \frac{1}{2} \left[\frac{\Delta\varepsilon_p}{2\varepsilon'_p} \right]^{1/c} \left[\frac{\ln(1 - 0.01x)}{\ln(0.5)} \right]^{1/b} \quad (1.6)$$

where:

$N_f(x\%)$ are the number of cycles to x% failure probability,

$\Delta\varepsilon_p$ the cyclic effective plastic strain range,

$2\varepsilon'_p$ is defined as:

$$2\varepsilon'_p = 0.65 \quad (1.7)$$

b is the slope of the Weibull probability plot, defined as “4 for stiff leadless attachments, and 2 for compliant leaded attachments” [10],

c is defined as:

$$c = -0.442 - 6 \times 10^{-4}T_{avg} + 1.74 \times 10^{-2}\ln\left(1 + \frac{360}{t_D}\right) \quad (1.8)$$

T_{avg} is the mean cyclic solder joint temperature,

t_D is the half-cycle dwell time in minutes, or the average time available for the stress to relax.

Recall, the definition of a failure here is taken as the initiation of a crack.

For an accurate approximation of the fatigue life, one should use the strain data from cyclic steady state. This may be achieved after several cycles of loading, which

can take up time in an FEM analysis as well as disk space. Thus we can approximate the fatigue life after the first few cycles, or even the first cycle, knowing that this will over-estimate the fatigue life, since the cycle has not achieved cyclic steady state as of yet. It has been shown that the difference between a fatigue life analysis using the first few cycles versus data after many cycles does not make a very large difference [50].

Chapter 2

FEM Thermal Analysis Theory

2.1 Heat Transfer Methods

2.1.1 Conductive Heat Transfer

Conductive heat transfer occurs in a gas or liquid with no bulk fluid motion (i.e., no mean velocity), or in solids. The heat transfer occurs as a result of random molecular activity within the medium in question. Thus, in the presence of a temperature gradient the energy from the molecules with higher energy is passed onto those with lower energy, and thus as logic would imply, the energy transfer occurs in direction of decreasing temperature. This phenomena is also known as the diffusion of energy [21]. In isotropic materials, the heat flux direction is normal to isotherms. Thus, we can define the following constitutive or rate equation known as Fourier's law:

$$\mathbf{q}_{\text{cond}}(\mathbf{r}, t) = -k \nabla T(\mathbf{r}, t) \quad (2.1)$$

T is the temperature,

k is the thermal conductivity tensor (second order), a material property, (units W/mK),

$\mathbf{q}_{\text{cond}}(\mathbf{r}, t)$ is the heat flux vector normal to an isotropic surface (Joules per unit time, per unit area),

$\mathbf{r}$ is the notation for an arbitrary point in three dimensions, (x, y, z) .

If we describe the heat flux as a vector quantity, and thermal conductivity constant in (x, y, z) directions, we can say that:

$$-k \nabla T(\mathbf{r}, t) = -k \left(\mathbf{i} \frac{\partial T}{\partial x} + \mathbf{j} \frac{\partial T}{\partial y} + \mathbf{k} \frac{\partial T}{\partial z} \right) \quad (2.2)$$

2.1.2 Convective Heat Transfer

Convective heat transfer occurs when there are essentially two mechanisms of heat transfer present, both diffusion by random molecular motion and bulk fluid motion. Convection essentially takes into account the superposition effect of both of these phenomena occurring simultaneously. In practice, it is common to use the term convection when referring to both of these cumulative forms of energy transfer, and the

term advection when referring to simply the energy transfer due to the bulk fluid motion.

If we consider a heated surface with a cooler fluid flowing over it, we introduce boundary layer phenomena. The velocity boundary layer occurs since we assume a no-slip condition at the solid surface, and thus zero velocity there. The gradient of velocity gradually decreases as one moves away from the surface until the velocity profile is uniform and of the value of a finite velocity value, say u_∞ . Similarly, the thermal boundary layer occurs in a similar fashion, where the surface is at some temperature T_s , and the ambient temperature of the fluid at some temperature T_∞ . In between these two temperatures there will be a varying gradient, which is the thermal boundary layer.

There are two main types of convective heat transfer, forced convection and free convection. The difference between the two is simple. Forced convection occurs when the fluid flow is caused by an external means, such as a fan or pump, whereas free convection occurs due to buoyancy forces arising from the heating of the fluid by the surface causing changes in density (assuming $T_\infty < T_s$).

The particular rate equation associated with convective heat transfer is [21] (assuming $T_\infty < T_s$) :

$$q''_{conv} = h(T_s - T_\infty) \quad (2.3)$$

q''_{conv} is the convective heat flux (units W/m^2),

h is the convection coefficient, (units W/m^2K), which depends upon factors such as surface geometry, nature of fluid motion, and other properties.

2.1.3 Radiation Heat Transfer

Radiation heat transfer occurs due to an energy emission from matter at a finite temperature. Primarily, this is accounted for from solid surfaces, but it can occur from solid, liquid or gases. This transport of energy occurs via electromagnetic waves, and occurs most readily in a vacuum. Thus, it is critical when performing heat transfer analysis for space-bound applications. However, for typical terrestrial applications, this mode of heat transfer is negligible at lower temperatures, but it is worth mentioning.

Thus, the rate equation for the maximum radiation heat flux between a surface surrounded by a vast space is given by [21]:

$$\begin{aligned} q''_{rad} &= \epsilon\sigma(T_s^4 - T_\infty^4) \\ &= \left\{ \epsilon\sigma \left[(T_s^2 + T_\infty^2)(T_s + T_\infty) \right] \right\} (T_s - T_\infty) \\ &= h_{rad}(T_s - T_\infty) \end{aligned} \tag{2.4}$$

q''_{rad} is the radiation heat flux (units W/m^2),

ϵ is the emissivity of a surface ($0 \leq \epsilon < 1$),

σ is the Stefan-Boltzmann constant, equal to $5.67 \times 10^{-8} W/m^2 K^4$.

h_{rad} a constant term, created by using values from the previous time step or iteration.

The term ϵ takes into account if the body emitting radiative energy is a black or grey body. If it is a black body, then $\epsilon = 1$, which an idealized case, but rarely happens in reality. Grey bodies have values in the range from $0 \leq \epsilon < 1$. Thus, this value ϵ determines how much heat a body radiates (or emits) and absorbs.

2.1.4 Differential Equation For Heat Conduction

We can introduce the differential form of the conduction equation as:

$$\frac{\partial}{\partial x} \left(k_x \frac{\partial T(\mathbf{r}, t)}{\partial x} \right) + \frac{\partial}{\partial y} \left(k_y \frac{\partial T(\mathbf{r}, t)}{\partial y} \right) + \frac{\partial}{\partial z} \left(k_z \frac{\partial T(\mathbf{r}, t)}{\partial z} \right) + q'''(\mathbf{r}, t) = \rho C_p \frac{\partial T(\mathbf{r}, t)}{\partial t} \quad (2.5)$$

where:

$q'''(\mathbf{r}, t)$ is rate of energy generation per unit volume, per unit time at the point (x, y, z)

$T(\mathbf{r}, t)$ is the temperature at a point (x, y, z) at a time t .

If we assume that conductivity is constant in (x, y, z) , then we can reduce equation (2.5) to:

$$k \left(\frac{\partial^2 T(\mathbf{r}, t)}{\partial x^2} + \frac{\partial^2 T(\mathbf{r}, t)}{\partial y^2} + \frac{\partial^2 T(\mathbf{r}, t)}{\partial z^2} \right) + q'''(\mathbf{r}, t) = \rho C_p \frac{\partial T(\mathbf{r}, t)}{\partial t} \quad (2.6)$$

$$k \nabla^2 T(\mathbf{r}, t) + q'''(\mathbf{r}, t) = \rho C_p \frac{\partial T(\mathbf{r}, t)}{\partial t} \quad (2.7)$$

If we examine equation (2.7) we can get a better physical meaning of the terms within it. First, if we consider taking the integral of (2.7) over a volume, V , (2.7) can be transformed using equation (2.2), into the classical equation (which determines the exact solution) for $T(\mathbf{r}, t)$:

$$\int_V \left[k \nabla^2 T(\mathbf{r}, t) + q'''(\mathbf{r}, t) - \rho C_p \frac{\partial T(\mathbf{r}, t)}{\partial t} \right] dv = 0 \quad (2.8)$$

Separation of the two ∇ operators and further substitution of equation (2.1) into equation (2.8) reveals:

$$\int_V \left[-\nabla \cdot \mathbf{q}_{\text{cond}}(\mathbf{r}, t) + q'''(\mathbf{r}, t) - \rho C_p \frac{\partial T(\mathbf{r}, t)}{\partial t} \right] dv = 0 \quad (2.9)$$

Thus, each term inside the integrand can be evaluated separately as such:

$$\int_V (-\nabla \cdot \mathbf{q}_{\text{cond}}(\mathbf{r}, t)) dv + \int_V (q'''(\mathbf{r}, t)) dv - \int_V \left(\rho C_p \frac{\partial T(\mathbf{r}, t)}{\partial t} \right) dv = 0 \quad (2.10)$$

We can now examine each term for its physical meaning:

1.

$$\int_V (-\nabla \cdot \mathbf{q}_{\text{cond}}(\mathbf{r}, t)) dv \quad (2.11)$$

Equation (2.11) represents the rate of heat entering the volume, V , by means of conductive heat transfer from the volume's surfaces, of area A . The divergence theorem can be applied here to further clarify the concept, by defining a surface normal vector $\mathbf{n}$:

$$-\int_V (\nabla \cdot \mathbf{q}_{\text{cond}}(\mathbf{r}, t)) dv = -\int_A (\mathbf{q}_{\text{cond}}(\mathbf{r}, t) \cdot \mathbf{n}) dA \quad (2.12)$$

2.

$$\int_V (q'''(\mathbf{r}, t)) dv \quad (2.13)$$

The equation (2.13) represents the amount of energy generation within the volume, V .

3.

$$\int_V \left(\rho C_p \frac{\partial T(\mathbf{r}, t)}{\partial t} \right) dv \quad (2.14)$$

Finally, the equation (2.40) represents the amount of energy stored in volume, V .

2.2 Initial and Boundary Conditions

We can say that the space domain in question is designated by Ω and is in $\mathcal{R}^3$ (i.e., three-dimensional space), and the boundary of this domain as Γ . Suppose we split this boundary into two separate regions, Γ_1 and Γ_2 , ($\Gamma_1 \cup \Gamma_2 = \Gamma$). As well, let the outward normal be designated at a boundary Γ_1 as $\mathbf{n1}$, and similarly $\mathbf{n2}$ for Γ_2 .

There are two main types of boundary conditions, Dirichlet and Neumann. Dirichlet boundary conditions (DBC) are simply those where a specified dependant variable is set on a boundary. Neumann boundary conditions (NBC) occur when the normal derivative of a dependant variable is specified on a boundary. A good way to illustrate these two types of boundary conditions is by using a thermal and stress analysis analogy, with temperature (T) and displacement (u) as the dependent variables:

1. A Dirichlet boundary condition (DBC) is a prescribed temperature on a boundary Γ_1 :

$$T(\mathbf{r}, 0) = T_d(\mathbf{r}) \quad \text{for all } \mathbf{r} \text{ on } \Gamma_1$$

A DBC, with respect to stress analysis, can be a prescribed displacement on Γ_1 :

$$u(\mathbf{r}, 0) = 0 \quad \text{for all } \mathbf{r} \text{ on } \Gamma_1$$

This is a Dirichlet boundary condition (DBC).

2. A Neumann boundary condition (NBC) can be specified by a flux condition upon boundary surfaces on Γ_2 , thus specifying the gradient at this surface:

$$q_{neum} = q_{rad} + q_{conv} + q_{ext} \quad \text{for all } \mathbf{r} \text{ on } \Gamma_2 \quad (2.15)$$

q_{neum} is the conductive flux normal to the surface Γ_2 , specified by equation (2.1),

q_{rad} is the radiation flux boundary condition, specified by equation (2.4),

q_{conv} is a convection flux boundary condition, specified by equation (2.3),

q_{ext} is a flux from an external heat source outside the domain.

Substituting in the above conditions, equation (2.15) becomes:

$$-(k\nabla T) \cdot \mathbf{n} = h_{rad}(T_s - T_\infty) + h(T_s - T_\infty) + q_{ext} \quad (2.16)$$

$$\text{for all } \mathbf{r} \text{ on } \Gamma_2 \quad (2.17)$$

Thus, in the context of a stress analysis, a NBC is a pressure on a face or set of faces.

2.3 Discretization

A common finite element method (FEM) is that of Bubnov-Galerkin. We first start with a domain and boundary with boundary conditions, such as those outlined above.

To discretize the domain, we must introduce what is called basis functions. These form a basis for a space in each individual element. In a common eight node brick, where nodes are simply placed at the vertices, the basis functions with the origin set at the centroid of the brick, are set as:

$$N_i = \frac{1}{8}(1 + rr_i)(1 + ss_i)(1 + tt_i), i = 1, 2, \dots, 8 \quad (2.18)$$

Sometimes in complex problems, higher order elements are needed, such as a 20 node brick. Here, nodes are at vertices and midpoints on edges. For these kind of elements, the basis functions are:

For vertex nodes:

$$N_i = \frac{1}{8}(1 + rr_i)(1 + ss_i)(1 + tt_i)(rr_i + ss_i + tt_i - 2) \quad (2.19)$$

For mid side nodes:

$$r_i = 0 \tag{2.20}$$

$$s_i = 1, -1$$

$$t_i = 1, -1$$

$$N_i = \frac{1}{4}(1 + r^2)(1 + ss_i)(1 + rr_i)$$

The reason for the need of higher order elements in a mesh can be shown with complex curvilinear heat flow paths in a material which cannot be approximated sufficiently accurately with simple 8-node bricks with their linear basis functions. An alternate solution would be to use a finer mesh with the same 8-node bricks.

To transform from their local coordinates (r, s, t) to the global coordinates (x, y, z) , we must make use of a linear mapping matrix called the Jacobian. This operator works as follows, shown for three dimensions:

$$\begin{bmatrix} dr \\ ds \\ dt \end{bmatrix} = [J] \begin{bmatrix} dx \\ dy \\ dz \end{bmatrix} \tag{2.21}$$

$$\begin{bmatrix} dx \\ dy \\ dz \end{bmatrix} = [J]^{-1} \begin{bmatrix} dr \\ ds \\ dt \end{bmatrix} \tag{2.22}$$

$$[J]^{-1} = \begin{bmatrix} x_{,r} & x_{,s} & x_{,t} \\ y_{,r} & y_{,s} & y_{,t} \\ z_{,r} & z_{,s} & z_{,t} \end{bmatrix} \quad (2.23)$$

Thus, for integrals over a domain, Ω , the following transforms a volume element from Cartesian (x, y, z) coordinates to (r, s, t) local element coordinates:

$$dxdydz = \det[J]drdsdt \quad (2.24)$$

2.4 General Finite Element Formulation - Energy Solver

2.4.1 Bubnov-Galerkin Finite Element Method

Once the domain is sufficiently discretized, and the boundary conditions applied, we can apply the Bubnov-Galerkin FEM to the problem. We can define the temperature, its gradient, and the Laplacian, respectively, in any element as a function of it's position:

$$\hat{T}(\mathbf{r}, t) = \sum_{i=1}^n T_i(t) N_i(\mathbf{r}) \quad (2.25)$$

$$\nabla \hat{T}(\mathbf{r}, t) = \sum_{i=1}^n T_i(t) \nabla N_i(\mathbf{r}) \quad (2.26)$$

$$\nabla \cdot \nabla \hat{T}(\mathbf{r}, t) = \sum_{i=1}^n T_i(t) \nabla \cdot \nabla N_i(\mathbf{r}) \quad (2.27)$$

where:

$\hat{T}(\mathbf{r}, t)$ is the approximate solution for one particular element

n is the product of the number of nodes in an element and the number of degrees of freedom (DOFs) of one node. In the case of temperature in an 8 node brick, the DOF of one node is 1, and thus $n = 8$.

$T_i(t)$ is the temperature at node i , a function of time,

Note that for an 8-node brick:

$$\nabla \cdot \nabla N_i(\mathbf{r}) \approx 0 \quad (2.28)$$

Bubnov-Galerkin method essentially says that we take a set of weighting functions in the test space (i.e., the vector space of basis functions where we exclude those nodes which have Dirichlet boundary conditions enforced on them) and make the error orthogonal to them, which gives us the best possible approximation to the exact solution in that space. In the context of a thermal analysis, this says that:

$$\int_{\Omega} W_i(\mathbf{r}) E(\mathbf{r}) d\Omega = 0 \quad (2.29)$$

$$E = T - \hat{T} \quad (2.30)$$

where:

$W_i(\mathbf{r})$ represents the weighting functions for Bubnov-Galerkin method,

E represents the error,

T represents the exact solution to the governing equation for the problem, from equation (2.8),

$\hat{T}$ represents the approximate solution to the governing equation for the problem,

Here, the Bubnov-Galerkin method uses a subset of the trial function space basis functions, $N_i(\mathbf{r})$ to represent the weighting functions $W_i(\mathbf{r})$. We shall denote these weighting functions as $N_j(\mathbf{r})$, and must remember that they are a subset of the trial function space basis functions, that exclude the basis functions which have Dirichlet BCs. Thus, if we use this in the classical heat equation (2.8), we get:

$$\int_{\Omega} \left(-k \nabla \cdot \nabla T_i(\mathbf{r}, t) + q'''(\mathbf{r}, t) - \rho C_p \frac{\partial T(\mathbf{r}, t)}{\partial t} \right) N_j(\mathbf{r}) d\Omega = 0 \quad (2.31)$$

We can apply Greens first identity to the first term in the above equation to yield:

$$\int_{\Omega} N_j(\mathbf{r})(-k\nabla\cdot\nabla T_i(\mathbf{r},t))d\Omega = \int_{\Gamma} N_j(\mathbf{r})(-k\nabla T_i(\mathbf{r},t))\cdot\mathbf{n}d\Gamma - \int_{\Omega} \nabla\cdot N_j(\mathbf{r})(-k\nabla T_i(\mathbf{r},t))d\Omega \quad (2.32)$$

Thus by substitution, we get the continuum weak form of the heat equation:

$$\begin{aligned} & \int_{\Omega} \nabla\cdot N_j(\mathbf{r})(k\nabla T_i(\mathbf{r},t))d\Omega - \int_{\Gamma} N_j(\mathbf{r})(k\nabla T_i(\mathbf{r},t))\cdot\mathbf{n}d\Gamma \\ & + \int_{\Omega} q'''(\mathbf{r},t)N_j(\mathbf{r})d\Omega - \int_{\Omega} \rho C_p \frac{\partial T(\mathbf{r},t)}{\partial t} N_j(\mathbf{r})d\Omega = 0 \end{aligned} \quad (2.33)$$

We call equation (2.33) the weak form for one main reason. By performing the transformation, we have weakened the restriction seen in equation (2.31) imposed by the double ∇ term, which requires the second derivative of the function $T_i(\mathbf{r},t)$ to exist. Since this is a difficult condition to ensure, we apply the above Green's identity to weaken this condition so that $T_i(\mathbf{r},t)$ need only be differentiable once.

Now that we have the continuum weak form, we can apply equations (2.25) and (2.26) to (2.33) to yield the discretized equations:

$$\begin{aligned} & \int_{\Omega} \nabla\cdot N_j(\mathbf{r})(k\nabla T_i(t)N_i(\mathbf{r}))d\Omega - \int_{\Gamma} N_j(\mathbf{r})(k\nabla T_i(t)N_i(\mathbf{r}))\cdot\mathbf{n}d\Gamma \\ & + \int_{\Omega} q'''(\mathbf{r},t)N_j(\mathbf{r})d\Omega - \int_{\Omega} \rho C_p \frac{\partial T(t)}{\partial t} N_i(\mathbf{r})N_j(\mathbf{r})d\Omega = 0 \end{aligned} \quad (2.34)$$

We can express this further as a series of vector and matrix quantities:

$$\begin{aligned}
& \int_{\Omega} \nabla \cdot [N_j] k \nabla [N_i]^T [T_i] d\Omega - \int_{\Gamma} [N_j] k \nabla [N_i]^T [T_i] \cdot \mathbf{n} d\Gamma \\
& + \int_{\Omega} q''' [N_j] d\Omega - \int_{\Omega} \rho C_p [T_i]_{,t} [N_i]^T [N_j] d\Omega = 0
\end{aligned} \tag{2.35}$$

where:

$[N_i]$ represents a column vector of basis functions, with $[N_i]^T$ its transpose,

$[N_j]$ represents a column vector of weighting functions, which in this case is a subset of the trial functions,

$[T_i]$ represents a column vector of nodal temperature values,

Applying the boundary conditions discussed earlier in equation (2.16) to the boundary terms where we integrate on Γ :

$$\begin{aligned}
& \int_{\Omega} \nabla \cdot [N_j] k \nabla [N_i]^T [T_i] d\Omega \\
& + \int_{\Gamma_2} [N_j] \left\{ h_{rad}([N_i]^T ([T_i] - T_{\infty})) + h([N_i]^T ([T_i] - T_{\infty})) + [N_i]^T [q_{ext}] \right\} d\Gamma \\
& + \int_{\Omega} q''' [N_j] d\Omega - \int_{\Omega} \rho C_p [T_i]_{,t} [N_i]^T [N_j] d\Omega = 0
\end{aligned} \tag{2.36}$$

We can rearrange and break each term in (2.36) into matrices, as follows:

$$\begin{aligned}
& \int_{\Omega} [B]^T [K] [B] [T_i] d\Omega \\
& + \int_{\Gamma_2} [N_j] \left\{ h_{rad}([N_i]^T [T_i] + h([N_i]^T [T_i]) + [N_i]^T q_{ext} \right\} d\Gamma
\end{aligned} \tag{2.37}$$

$$\begin{aligned}
&= \int_{\Gamma_2} [N_j] \left\{ h_{rad}([N_i]^T T_\infty) + h([N_i]^T T_\infty) \right\} d\Gamma \\
&\quad + \int_{\Omega} [q'''] [N_j] d\Omega + \int_{\Omega} \rho C_p [T_i]_{,t} [N_i]^T [N_j] d\Omega
\end{aligned}$$

where:

$[B]$ represents an 8×3 matrix of the first derivative of the basis functions,

$[K]$ represents the thermal conductivity tensor at a 3×3 matrix.

Thus, we can define:

1.

$$\int_{\Omega} \rho C_p [N_i]^T [N_j] d\Omega = [M] \quad (2.38)$$

This is defined as the mass (or capacitance) matrix.

2.

$$\int_{\Omega} [B]^T [K] [B] d\Omega + \int_{\Gamma_2} [N_j] \left\{ h_{rad}[N_i]^T + h[N_i]^T \right\} d\Gamma = [K] \quad (2.39)$$

This is defined as the stiffness matrix.

3.

$$\int_{\Gamma_2} [N_j] \left\{ h_{rad}([N_i]^T T_\infty) + h([N_i]^T T_\infty) - [N_i]^T q_{ext} \right\} d\Gamma + \int_{\Omega} [q'''] [N_j] d\Omega = [F] \quad (2.40)$$

This is defined as the external load vector.

With the above substitutions, we obtain the final FEM form of:

$$[M][T_i]_{,t} + [K][T_i] = [F] \quad (2.41)$$

With the above form of the equation in mind, we can alternatively define the residual in (2.30) as:

$$R = [M][T_i]_{,t} + [K][T_i] - [F] \quad (2.42)$$

Chapter 3

Plasticity Theory

3.1 Stress Strain Theory - Elastic and Plastic Deformations

The underlying equations for stress with elastic deformations in three dimensions are, using tensorial notation:

$$G = \frac{E}{2(1 + \mu)} \quad (3.1)$$

$$\varepsilon_{ij} = \frac{\sigma_{ij}}{2G} - \frac{\mu}{E} \sigma_{kk} \delta_{ij} \quad (3.2)$$

$$\sigma_{ij} = 2G\varepsilon_{ij} - \lambda\varepsilon_{kk}\delta_{ij} \quad (3.3)$$

$$\varepsilon_{kk} = \varepsilon_{xx} + \varepsilon_{yy} + \varepsilon_{zz} = \text{tr}(\varepsilon) \quad (3.4)$$

$$\sigma_{kk} = \sigma_{xx} + \sigma_{yy} + \sigma_{zz} = \text{tr}(\sigma) \quad (3.5)$$

$$\lambda = \frac{E\mu}{(1+\mu)(1-2\mu)} \quad (3.6)$$

where:

μ is Poisson's ratio,

E is Young's modulus,

G is the shear modulus,

λ a Lamé material constant,

δ_{ij} is the Kronecker delta, where $\delta_{ij} = 0$ if $i \neq j$, and $\delta_{ij} = 1$ if $i = j$ (an alternative way of expressing the identity matrix, $\mathbf{I}$),

σ_{ij} is the Cauchy stress tensor, which in an Cartesian (x, y, z) coordinate system is

defined as:

$$[\sigma] = \begin{bmatrix} \sigma_{xx} & \sigma_{xy} & \sigma_{xz} \\ \sigma_{yx} & \sigma_{yy} & \sigma_{yz} \\ \sigma_{zx} & \sigma_{zy} & \sigma_{zz} \end{bmatrix} \quad (3.7)$$

ε_{ij} is the strain tensor, which in an Cartesian (x, y, z) coordinate system is defined

as:

$$[\varepsilon] = \begin{bmatrix} \varepsilon_{xx} & \varepsilon_{xy} & \varepsilon_{xz} \\ \varepsilon_{yx} & \varepsilon_{yy} & \varepsilon_{yz} \\ \varepsilon_{zx} & \varepsilon_{zy} & \varepsilon_{zz} \end{bmatrix} = \begin{bmatrix} \varepsilon_{xx} & \frac{1}{2}\gamma_{xy} & \frac{1}{2}\gamma_{xz} \\ \frac{1}{2}\gamma_{yx} & \varepsilon_{yy} & \frac{1}{2}\gamma_{yz} \\ \frac{1}{2}\gamma_{zx} & \frac{1}{2}\gamma_{zy} & \varepsilon_{zz} \end{bmatrix} \quad (3.8)$$

γ_{xy} is the engineering shear strain component in the xy-plane, is a half of the tensorial shear strain component,

ε_{xy} is the tensorial shear strain component in the xy-plane,

σ_{xx} is the normal stress in the x-direction,

ε_{xx} is the normal strain in the x-direction.

We can use the Cauchy stress tensor to calculate the principle stresses (i.e., stress states with no shear components) by determining the eigenvalues of the Cauchy stress tensor. We can write the characteristic equation for (3.7) to get the stress invariants:

$$\sigma^3 - J_1\sigma^2 + J_2\sigma + J_3 = 0 \quad (3.9)$$

$$J_1 = tr(\sigma_{ij}) = \sigma_{ii} = \sigma_{xx} + \sigma_{yy} + \sigma_{zz} = \sigma_1 + \sigma_2 + \sigma_3 \quad (3.10)$$

$$\begin{aligned} J_2 &= \frac{1}{2}\sigma_{ij}\sigma_{ij} \\ &= \sigma_{xx}\sigma_{yy} + \sigma_{yy}\sigma_{zz} + \sigma_{xx}\sigma_{zz} - \sigma_{xy}^2 - \sigma_{xz}^2 - \sigma_{yz}^2 \\ &= \sigma_1\sigma_2 + \sigma_2\sigma_3 + \sigma_3\sigma_1 \end{aligned} \quad (3.11)$$

$$J_3 = \frac{1}{3}\sigma_{ij}\sigma_{jk}\sigma_{ki} = det[\sigma] \quad (3.12)$$

The same decomposition is applied to plasticity theory, except the stress tensor is decomposed in a slightly different manner:

$$\sigma_{ij} = \sigma_H \delta_{ij} + S_{ij} \quad (3.13)$$

where:

σ_H is the hydrostatic stress, determined by:

$$\sigma_H = \frac{1}{3}\sigma_{ii} = \frac{1}{3}(\sigma_{xx} + \sigma_{yy} + \sigma_{zz}) = \frac{1}{3}(\sigma_1 + \sigma_2 + \sigma_3) \quad (3.14)$$

S_{ij} is the deviatoric stress tensor, determined by:

$$S_{ij} = \sigma_{ij} - \sigma_H \delta_{ij} \quad (3.15)$$

If we decompose (3.15) as we did before, we can determine the invariants of the deviatoric stress tensor:

$$J'_1 = \text{tr}(S_{ij}) = S_{ii} = S_1 + S_2 + S_3 = 0 \quad (3.16)$$

$$\begin{aligned} J'_2 &= \frac{1}{2} S_{ij} S_{ij} \\ &= \frac{1}{6} \left[(\sigma_{xx} - \sigma_{yy})^2 + (\sigma_{yy} - \sigma_{zz})^2 + (\sigma_{zz} - \sigma_{xx})^2 \right] - \sigma_{xy}^2 - \sigma_{xz}^2 - \sigma_{yz}^2 \\ &= \frac{1}{6} \left[(\sigma_1 - \sigma_2)^2 + (\sigma_2 - \sigma_3)^2 + (\sigma_3 - \sigma_1)^2 \right] \end{aligned} \quad (3.17)$$

$$= \frac{1}{2} \left[S_1^2 + S_2^2 + S_3^2 \right] \quad (3.18)$$

$$J'_3 = \frac{1}{3} S_{ij} S_{jk} S_{ki} = S_1 S_2 S_3 = \det[S_{ij}] \quad (3.19)$$

3.1.1 Stress Space and the Yield Function

The above deviatoric stress invariants are very important with regards to determining yielding conditions for materials. First we must introduce the concept of the stress space. A symmetric 3×3 matrix, such as the stress tensor, has six independent stress components, and hence a six-dimensional stress space, to describe a stress state geometrically. A three-dimensional stress space with the principle stresses as the axes can be used, and is referred to as the Haigh-Westergaard stress space. We

can further describe a plane, called the octahedral plane, in this space as:

$$\sigma_1 + \sigma_2 + \sigma_3 = \sqrt{3}d \quad (3.20)$$

Here, d is the distance from the origin of the Haigh-Westergaard stress space to the plane. When d is zero, the plane is called the π -plane. It is called the octahedral plane simply because its normal passing through the origin makes equal angles with the principle stress axes. It should be noted that the normal of this plane which passes through the origin (i.e., a stress diagonal), can have any stress state projected upon it. On this line:

$$\sigma_1 = \sigma_2 = \sigma_3 = \frac{1}{3}J_1 \quad (3.21)$$

As previously stated, this line makes equal angles with the principle axes, with the angle being determined by $\cos^{-1}1/\sqrt{3}$. For a given stress point in the stress space, the projection of this point upon this line represents the state of hydrostatic stress, whereas the projection of this point onto the π -plane represents the deviatoric stress state. From this π -plane we can determine the yield surface for many materials. The yield surface is such that:

$$F(\sigma_{ij}) = 0 \quad (3.22)$$

Thus, the conditions for loading and unloading are such:

$$\begin{aligned}
F(\sigma) &< 0 && \text{(Elastic deformation)} \\
F(\sigma) &= 0 \quad dF > 0 && \text{(Loading, plastic deformation)} \\
F(\sigma) &= 0 \quad dF = 0 && \text{(Neutral loading)} \\
F(\sigma) &= 0 \quad dF < 0 && \text{(Unloading)}
\end{aligned} \tag{3.23}$$

$F(\sigma_{ij}) > 0$ has no physical meaning (neglecting viscous effects). Experiments have shown that for plastic yielding of metals, the hydrostatic stress has no effect [22]. We can use Von Mises stress equation to determine the yield criterion for a material at a point by using all the above stress components. It states that yielding will occur when the second invariant of the deviatoric stress tensor, J'_2 , reaches a critical value, k^2 , a material property. It can be found that:

$$k = \sigma_y / \sqrt{3} = \tau_y \tag{3.24}$$

where:

σ_y is the yield stress in uniaxial tension,

τ_y is the yield shear stress.

With this, the yield surface takes the form of a cylinder in principle stress space with its axis parallel to the hydrostatic stress axis, and thus, perpendicular to the

π -plane (or any deviatoric plane for that matter). In deviatoric stress space, the yield surface is a sphere. The intersection of the yield surface with the π -plane is known as the yield curve.

With this in mind:

$$\begin{aligned}\sigma_y &= \sqrt{\frac{1}{2} ((\sigma_{xx} - \sigma_{yy})^2 + (\sigma_{zz} - \sigma_{xx})^2 + (\sigma_{zz} - \sigma_{yy})^2) + 3 (\tau_{xy}^2 + \tau_{xz}^2 + \tau_{yz}^2)} \\ &= \sqrt{J_2'}\end{aligned}\tag{3.25}$$

We can say from this discussion, neglecting viscous effects, that:

$$J_2' - k^2 = 0 \quad \text{for yielding or plastic deformation} \tag{3.26}$$

$$J_2' - k^2 < 0 \quad \text{for elastic deformation} \tag{3.27}$$

Although the above condition is necessary for plastic deformation, it is not sufficient for plastic deformation to occur. If we consider the stress space and the stress state in terms of vectors, we can consider some arbitrary loading which results in a trial stress state, σ_T . Plastic deformation can only occur if this trial stress state lies outside the yield surface, with the plastic strain increment projected from this trial stress state onto the normal of the yield surface and scaled by some scalar (see equation 3.36).

We must distinguish between the elastic and plastic ranges when deforming. Linear elastic deformation obeys Hooke's law, where the stress level is linearly dependant upon the strain. However once the yield stress level is reached, then typically a certain hardening within the material occurs, while the loading in the specimen continues to increase and the deformation becomes plastic. Upon unloading after this plastic deformation, the body does not return to its original shape, and thus at zero stress, we can have some strain within the material. Thus we can define infinitesimal strain in three-dimensions as:

$$\varepsilon_{ij} = \varepsilon_{ij}^e + \varepsilon_{ij}^p \quad (3.28)$$

where:

ε_{ij}^p is the plastic strain component,

ε_{ij}^e is the elastic strain component,

ε_{ij} is the total strain tensor component,

We can also note that the dissipation caused by the plastic strain causes the plastic deformation to be path dependent. This is clear since after plastic deformation, a stress value does not have a unique strain value associated with it. Since it is path dependent, it is necessary to integrate the incremental stress-strain relationship along this path to determine the final deformation state.

We can now introduce the following equivalent stress and equivalent strain:

$$\begin{aligned}
\sigma_{eq} &= \frac{1}{\sqrt{2}} \sqrt{((\sigma_{xx} - \sigma_{yy})^2 + (\sigma_{yy} - \sigma_{zz})^2 + (\sigma_{zz} - \sigma_{xx})^2) + 6(\tau_{xy}^2 + \tau_{xz}^2 + \tau_{yz}^2)} \\
&= \frac{1}{\sqrt{2}} \sqrt{((\sigma_1 - \sigma_2)^2 + (\sigma_2 - \sigma_3)^2 + (\sigma_3 - \sigma_1)^2)} \\
&= \sqrt{\frac{2}{3} S_{ij} S_{ij}}
\end{aligned} \tag{3.29}$$

$$\begin{aligned}
\varepsilon_{eq} &= \sqrt{\left(\frac{2}{9}(\varepsilon_{xx} - \varepsilon_{yy})^2 + (\varepsilon_{yy} - \varepsilon_{zz})^2 + (\varepsilon_{zz} - \varepsilon_{xx})^2\right) + \frac{4}{3}(\varepsilon_{xy}^2 + \varepsilon_{xz}^2 + \varepsilon_{yz}^2)} \\
&= \sqrt{\left(\frac{2}{9}(\varepsilon_1 - \varepsilon_2)^2 + (\varepsilon_2 - \varepsilon_3)^2 + (\varepsilon_3 - \varepsilon_1)^2\right)}
\end{aligned} \tag{3.30}$$

Note that (3.29) is essentially von Mises yield criterion.

3.1.2 Plastic Flow Theory

3.1.2.1 Flow Rule

Now we must try to formulate a stress strain relationship for plastic deformation. Once plastic deformation is initiated, the constitutive laws of elastic deformation are no longer valid (i.e., Hooke's law). Since the plastic stress strain relationship is non-linear and path/history dependant, it is also more desirable to work in increments in the plastic range. The Levy-Mises theory of plasticity makes the following assumptions:

1. Elastic strains, ε_{ij}^e , are relatively small compared with plastic strains, and can thus be neglected, which for most metals, is true.
2. The principle axes of strain increment (or rate), $d\varepsilon_{ij}$ are coincident with the principle axes of stress, σ_{ij} .

Thus, Levy-Mises proposed that the total incremental strain is related to the deviatoric stress components:

$$d\varepsilon_{ij} = S_{ij}d\lambda \quad (3.31)$$

where:

$d\lambda$ is a scalar multiplier, which is determined from the yield criterion.

The Prandtl-Reuss equation is similar to that of Levy-Mises, with one key exception. Prandtl-Reuss proposed that for materials where one could not neglect the elastic strains, $\varepsilon_{ij} = \varepsilon_{ij}^e + \varepsilon_{ij}^p$. Thus:

$$d\varepsilon_{ij}^p = S_{ij}d\lambda \quad (3.32)$$

Prandtl-Reuss equations are suitable for most plasticity cases, whereas Levy-Mises is suitable for cases such as rigid plasticity where metal is drawn to large strains. Thus, in such a case, elastic strains can be neglected if spring back is not deemed to be important.

Similar to the equivalent strain rate, we can introduce the equivalent plastic strain increment:

$$\begin{aligned}
d\varepsilon_{eq}^p &= \left[\frac{2}{9}((d\varepsilon_{xx}^p - d\varepsilon_{yy}^p)^2 + (d\varepsilon_{yy}^p - d\varepsilon_{zz}^p)^2 + (d\varepsilon_{zz}^p - d\varepsilon_{xx}^p)^2) \right. \\
&\quad \left. + \frac{4}{3}((d\varepsilon_{xy}^p)^2 + (d\varepsilon_{xz}^p)^2 + (d\varepsilon_{yz}^p)^2) \right]^{1/2} \\
&= \frac{2}{\sqrt{3}} \sqrt{((d\varepsilon_1^p - d\varepsilon_2^p)^2 + (d\varepsilon_2^p - d\varepsilon_3^p)^2 + (d\varepsilon_3^p - d\varepsilon_1^p)^2)} \\
&= \frac{2}{\sqrt{3}} \sqrt{\frac{1}{2} d\varepsilon_{ij}^p d\varepsilon_{ij}^p}
\end{aligned} \tag{3.33}$$

Similarly formed equations can be formed for the equivalent total strain rate using the same forms as in (3.33). It should be noted that (3.33) determines the magnitude of the vector $d\varepsilon_{eq}^p$, but tells us nothing about the previous path or history. With these notations, it can be shown that the parameter $d\lambda$ relating incremental strain (for Levy-Mises) or incremental plastic strain (for Prandtl-Reuss) to the deviatoric stress tensor as:

$$d\lambda = \frac{3d\varepsilon_{eq}}{2\sigma_{eq}} \quad \text{using Levy-Mises,} \tag{3.34}$$

$$d\lambda = \frac{3d\varepsilon_{eq}^p}{2\sigma_{eq}} \quad \text{using Prandtl-Reuss} \tag{3.35}$$

Considering the Prandtl-Reuss equation, we introduce the concept of the plastic potential function:

$$d\varepsilon_{ij}^p = \frac{\partial Q(\sigma_{ij})}{\partial \sigma_{ij}} d\lambda \quad (3.36)$$

where:

$Q(\sigma_{ij})$ is the plastic potential function.

However, we have to define $Q(\sigma_{ij})$. For most metals, $Q(\sigma_{ij})$ is equal to $F(\sigma_{ij})$, the yield surface. With this substitution, (3.36) is known as the associated flow rule, and the term $\frac{\partial Q(\sigma_{ij})}{\partial \sigma_{ij}} = \frac{\partial F(\sigma_{ij})}{\partial \sigma_{ij}}$ is then a normal to the yield surface in stress space.

3.2 Hardening

A typical metallic engineering material undergoing plastic deformation will undergo a certain hardening behaviour. This hardening will affect the yield surface's location in the stress space. Thus, we must take this into account upon plastic deformation to accurately model plasticity. Perfectly plastic materials are assumed to have no change in yield surface, but do not occur in actual engineering practice. Thus, we can modify our definition of the yield surface, so that subsequent yield surfaces can be predicted:

$$F(\sigma, \alpha_i) = 0 \quad (3.37)$$

where:

σ is the stress tensor, as before,

α_i is the i -th internal variable hardening parameter, $i = 1, 2, \dots, n$.

This hardening parameter changes the yield surface only when plastic deformation occurs. Commonly, there are two types of hardening models, isotropic and kinematic.

Isotropic hardening assumes that the yield surface expands uniformly from the initial yield surface. This is seen as concentric circles on the $\pi - plane$, if von Mises criterion for yielding is used. Thus for Von Mises material, we get for isotropic hardening and using deviatoric invariants (neglecting hydrostatic stresses, J'_1 , as before):

$$F(J'_2, k) = f(J'_2) - k = 0 \quad (3.38)$$

where:

k is a material constant modeling the isotropic hardening effect, or in effect the change in radius of the yield surface circle on the $\pi - plane$.

Kinematic hardening models state that the yield surface size remains constant, but it translates in the stress space. This can be expressed as:

$$F(S, \alpha) = f(S, \alpha) - k_o = 0 \quad (3.39)$$

where:

k_o is a material constant modeling the isotropic hardening effect, or the (constant) radius of the yield surface circle on the $\pi - plane$ before plastic deformation, α is a second order tensor, representing the new center of the yield surface in the stress space, also known as the back stress.

Neither kinematic nor isotropic hardening are exact representations of what happens in real-life. Experiments have shown that the Baushinger effect is pronounced, representing kinematic hardening effects. This effect states that the yield stress of a component in the reversed direction of the original loading is much less than that of the original yield stress. This is due to the anisotropy of the dislocation distribution caused by the initial plastic yielding [22]. Although isotropic hardening is easier to model, kinematic hardening does a somewhat better job of modeling this effect.

3.2.1 Isotropic Hardening

With these definitions, we can return to the issue of hardening. For isotropic hardening there are two measures of hardening, recalling (3.38):

1.

$$\alpha = \varepsilon_{eq}^p = \int d\varepsilon_{eq}^p \quad (3.40)$$

This is commonly referred to as the strain-hardening model, and assumes the hardening is a function of the amount of plastic strain induced.

2.

$$\alpha = W^p = \int \sigma_{ij} d\varepsilon_{ij}^p \quad (3.41)$$

This is commonly referred to as the work-hardening model, and assumes that the yield surface depends solely on the total plastic work done.

For von Mises yield criterion, both the above hardening rules are equivalent. Recalling that $k = \sigma_y/\sqrt{3}$, where σ_y is the yield stress in a tensile test, and k is the yield stress in a shear test. If k is a function of α (i.e., $k(\alpha)$), von Mises says that only the second deviatoric stress invariant is critical in yielding:

$$F(J'_2, k) = J'_2 - \frac{1}{3}\sigma_y^2(\alpha) = 0 \quad (3.42)$$

For the strain hardening model, the yield surface can be determined by:

$$J'_2 = \frac{1}{3}\sigma_y^2(\varepsilon_{eq}^p) \quad (3.43)$$

For the work hardening model, yield surface is determined by:

$$J'_2 = \frac{1}{3}\sigma_y^2(W^p) = \frac{1}{3}\sigma_y^2\left(\int \sigma_{eq} d\varepsilon_{eq}^p\right) = \frac{1}{3}\sigma_y^2(\varepsilon_{eq}^p) \quad (3.44)$$

3.2.2 Kinematic Hardening

As stated, kinematic hardening is a translation of the yield surface in stress space. Thus, we can describe this translation by introducing a kinematic hardening form of the yield function:

$$F(\sigma) = f(\sigma - \alpha) - k_o = 0 \quad (3.45)$$

where:

α is a second order tensor, representing the new center of the yield surface in the stress space, also known as the back stress.

Thus for the kinematic hardening model, there exists a need to derive evolution equations for the backstress, α , in terms of ε^p or σ . There are two main hypotheses for this evolution. The first is by Prager [22], who proposed the following linear equation for the evolution of the backstress:

$$d\alpha = cd\varepsilon^p \quad (3.46)$$

where:

c is a material constant, and can be linear or non-linear, depending on the modeling complexity.

(3.46) says that the yield surface moves in the direction of the plastic strain increment, or in the direction of the normal to the yield surface at the loading point. However, it has been shown, by Zeigler [22] that Prager's proposal is inconsistent for two and three-dimensional problems. Thus, he proposed the following:

$$d\alpha = (\sigma - \alpha)d\nu \quad (3.47)$$

where:

$d\nu$ is scalar constant determined by the yield condition.

(3.47) gives a radial movement of the yield surface, with the direction dictated by the tensor difference $(\sigma - \alpha)$. However, for a von Mises yield criterion, in three-dimensions, both (3.46) and (3.47) give the same directions.

3.2.3 Viscous Effects

In the above discussion, we assumed that the stress state could not go beyond the yield surface, and that the plasticity was rate independent. However, with viscous effects, and thus rate dependency, included in this discussion this now becomes possible. The theory underlying this phenomena takes into account stress relaxation and creep, and rather than considering each of these phenomena separately, couples them into viscoelasticity, or viscoplasticity, as the case may be. Two basic one-dimensional models for viscoelasticity is the Maxwell and Voight models.

Maxwell considered viscous effects by using a spring and damper in series to model the viscoelastic effects. Voight used the spring and damper in parallel, but the underlying concepts are the same. If we apply the stress and strain analogy to these systems, we can apply a stress (i.e., a load) to the system, and the system undergoes a strain (i.e., a deformation). We will assume the spring to have a constant of elasticity G , and the damper to have a damping constant η . We recall that a spring produces a force in response to a displacement, and a damper produces a force in response to a velocity. Thus, we can get the following equations for the forces in the spring and damper, respectively:

$$\sigma = G\varepsilon \tag{3.48}$$

$$\sigma = \eta\dot{\varepsilon} \tag{3.49}$$

Thus, for an applied stress σ , the Maxwell model gives:

$$\dot{\varepsilon} = \frac{\dot{\sigma}}{G} + \frac{\sigma}{\eta}, \quad \varepsilon(0) = \frac{\sigma(0)}{G} \tag{3.50}$$

As well, the Voight model gives:

$$\sigma = G\varepsilon + \eta\dot{\varepsilon}, \quad \varepsilon(0) = 0 \tag{3.51}$$

With either of these models, we can consider a unit input of loading, $\sigma = \sigma_o[U(t)]$, where:

$$\begin{aligned} [U(t - t_o)] &= 1 \quad (t > t_o) \\ &= 0 \quad (t < t_o) \end{aligned} \quad (3.52)$$

For this loading input, we can show that the creep response, $\varepsilon(t)$ for the Maxwell and Voight models, respectively, are:

$$\varepsilon(t) = \left(\frac{1}{G} + \frac{1}{\eta} t \right) \sigma_o[U(t)] \quad (3.53)$$

$$\varepsilon(t) = \frac{1}{G} \left(1 - e^{-\frac{t}{\tau}} \right) \sigma_o[U(t)] \quad (3.54)$$

where:

τ is the relaxation time, η/G .

Similarly, to examine the stress relaxation response, we can consider a unit input of deformation, $\varepsilon = \varepsilon_o[U(t)]$. For this deformation input, we can show that the stress relaxation response, $\sigma(t)$ for the Maxwell and Voight models, respectively, are:

$$\sigma(t) = G e^{-\frac{t}{\tau}} \varepsilon_o[U(t)] \quad (3.55)$$

$$\sigma(t) = \eta \varepsilon_o [\delta(t)] + G \varepsilon_o [U(t)] \quad (3.56)$$

where:

$\delta(t)$ is the impulse function, where:

$$\begin{aligned} [\delta(t - t_o)] &= 0 \quad (t \neq t_o) \\ &= 1 \quad (t = t_o) \end{aligned} \quad (3.57)$$

The Maxwell model shows that a sudden loading creates an immediate deformation, due to the spring in series with the damper. By contrast, the Voight model has no initial deformation for the same loading, since the damper prevents instantaneous deformation. In time, the Voight model will gradually deform, due to the presence of the damper in parallel with the spring.

The sudden deformation causes the Maxwell model to create an immediate force upon deformation, due to the spring in series, followed by relaxation, according to (3.55). The Voight model shows that it would require an infinite stress at time t_o to produce a finite instant strain.

Both of the above models have limitations, and thus, the more general standard linear solid model, was proposed by Kelvin. It uses a combination of the Maxwell and Voight models, where a damper and spring in series are put in parallel with another

spring. Various viscoelasticity models can be made with combinations of the Maxwell and Voight models.

Carrying on the one-dimensional theory, we now take the above concepts to viscoplasticity, a special case of viscoelasticity, where the stress state can go beyond that of the yield surface. The model we wish to use has a damper and friction block in parallel, with a spring in series with this parallel combination. The parallel combination is only active once plastic deformation is present, and in the elastic region, the lone spring is only active (loading the system by σ , the load in the spring is $\sigma^e = \sigma$).

As we load this system up with a stress σ , then approach and then equal the yield stress, σ_y , then the friction block becomes active. If the stress level is less than yield, it isn't used.

Using this simple one-dimensional model as a reference, the theory can be applied to multiple dimensions, noting that

$$\dot{\varepsilon}_{ij} = \dot{\varepsilon}_{ij}^e + \dot{\varepsilon}_{ij}^{vp} \quad (3.58)$$

Recalling the yield function, with hardening, is of the form:

$$F(\sigma_{ij}, \alpha_i) = f(\sigma_{ij}) - k(\alpha_i) = 0 \quad (3.59)$$

We can recall that $k(\alpha_i)$ defines the radius of the yield surface, which is a

function of the hardening parameters, α_i . Thus, we can alternatively say that the onset of viscoplastic flow occurs when the function F reaches the yield stress, σ_{yo} :

$$F(\sigma_{ij}, \alpha_i) = \sigma_{yo} \quad (3.60)$$

Both of these above equations represent the same thing. That is, they say that viscoplastic flow will only occur when the stress state is beyond the yield surface defined by F . However now, with viscous effects, the surface acts more like a rubber band, rather than a hard surface, as was with rate-independent plasticity [50]. In the above model, the friction block is a switch that acts as a way to model the dissipative effects of plastic flow.

We can now modify the associated flow rule, (3.36), so that we can incorporate viscoplastic flow:

$$\varepsilon_{ij}^{vp} = \gamma \langle \Phi(F) \rangle \frac{\partial F(\sigma_{ij})}{\partial \sigma_{ij}} \quad (3.61)$$

$$\begin{aligned} \langle \Phi(F) \rangle &= \Phi(F) \quad \text{if } F > 0 \\ &= 0 \quad \text{if } F \leq 0 \end{aligned} \quad (3.62)$$

The role of the $\Phi(F)$ factor in (3.61) is to enforce the fact that we only want viscoplastic flow when we go beyond the yield surface.

Chapter 4

FEM Stress Analysis Theory

4.1 Finite Element Method for Stress

4.1.1 Continuum Mechanics Review

We can now introduce continuum mechanics notation for quasi-static stress analysis, which relates thermal and mechanical phenomena:

$$\nabla \cdot \sigma + b = 0 \tag{4.1}$$

$$\frac{D\rho}{Dt} + \rho \nabla \cdot v = \rho_{,t} + \nabla \cdot (\rho v) = 0 \tag{4.2}$$

$$\sigma \cdot n = T \tag{4.3}$$

where:

v is the velocity of some arbitrary body,

b is the body force acting at a point $p(x, y, z)$ in a domain Ω (examples are gravity, or inertia forces),

T is the traction forces acting on the surface of a domain.

Equation (4.1) represents the conservation of linear momentum, and equation (4.2) represents the conservation of mass, where the body is in static equilibrium with no inertia forces acting on it. Equation (4.3) is the definition of traction or the definition of stress.

We can define some basic deformation and strain continuum approaches. At $t = 0$, the initial state, a point in some arbitrary body is called a material point, P . After some deformation at a time $t = t_1$, the point moves to a new position in space, P_1 , and the position is referred to as a spatial point. The material point has with it coordinates X , where X is a general coordinate system in three dimensions (i.e., $X(X, Y, Z)$). Similarly, the spatial point has coordinates $x(x, y, z)$. The movement from P to P_1 can be expressed in terms of a displacement vector $\mathbf{U}$ where, assuming both material and spatial point use the same coordinate system origin:

$$\mathbf{U} = x(x, y, z) - X(X, Y, Z) \quad (4.4)$$

We alternatively describe the motion from point X to x as:

$$x = x(X, t) \quad (4.5)$$

Defining the deformation gradient:

$$\mathbf{F} = \frac{\partial x_i}{\partial X_i} \quad (4.6)$$

It can also be shown that the deformation gradient can be expressed as the product of a rotation tensor $\mathbf{R}$ and a stretch tensor $\mathbf{U}_S$, such that:

$$\mathbf{F} = \mathbf{R}\mathbf{U}_S \quad (4.7)$$

Let Q be a point in the close to a point P . We can define a vector between them, $Q - P = \mathbf{dX}$. Similarly, if these two points are deformed to the new spatial position, the vector changes into a new vector $\mathbf{dx}$. The deformation gradient (4.6) maps this deformation of the neighbourhood of a point:

$$\mathbf{dx} = \mathbf{F}\mathbf{dX} = \mathbf{R}(\mathbf{U}_S\mathbf{dX}) \quad (4.8)$$

If we have two sets of vectors in the undeformed ($\mathbf{dX}_1, \mathbf{dX}_2$) and deformed bodies

$(\mathbf{dx}_1, \mathbf{dx}_2)$, we can define the change in length of these two vectors and the angle between them by saying:

$$\mathbf{dx}_1 \cdot \mathbf{dx}_2 = \mathbf{dX}_1 \cdot \mathbf{C} \mathbf{dX}_2 \quad (4.9)$$

where:

$\mathbf{C}$ the right Cauchy-Green deformation tensor, a material tensor quantity, defined as:

$$\mathbf{C} = \mathbf{F}^T \mathbf{F} \quad (4.10)$$

Similarly, we can define the left Cauchy-Green deformation tensor, a spatial tensor quantity, as:

$$\mathbf{b} = \mathbf{F} \mathbf{F}^T \quad (4.11)$$

$$\mathbf{dX}_1 \cdot \mathbf{dX}_2 = \mathbf{dx}_1 \cdot \mathbf{b}^{-1} \mathbf{dx}_2 \quad (4.12)$$

We can now define Green's (or Lagrangian since it is using a material tensor) strain tensor as:

$$\varepsilon = \frac{1}{2}(\mathbf{C} - \mathbf{I}) = \frac{1}{2} \left(\frac{\partial U_i}{\partial X_j} + \frac{\partial U_j}{\partial X_i} - \frac{\partial U_k}{\partial X_i} \frac{\partial U_k}{\partial X_j} \right) \quad (4.13)$$

where $\mathbf{I}$ is the second order identity tensor.

It can be shown that the above quantities can be expressed using the deformed or Eulerian formulation. Lagrangian formulations are evaluated in the initial configuration, whereas Eulerian formulations are evaluated in the deformed configuration.

(4.13) is valid for finite strains as well as infinitesimal strains. However, if the $\partial U_i / \partial X_j$ terms are much less than unity, then the product terms can be neglected, and we can say for small or infinitesimal strains:

$$\varepsilon = \frac{1}{2} \left(\frac{\partial U_i}{\partial X_j} + \frac{\partial U_j}{\partial X_i} \right) = \frac{1}{2} (\nabla \mathbf{U} + \nabla \mathbf{U}^T) \quad (4.14)$$

We can define stress in different ways, depending on which configuration we wish to describe it in. Typically, we describe stress in terms of Cauchy stress, σ , that is, stress in terms of the deformed body. We can also define the Kirchhoff stress tensor, τ , such that it describes the stress with respect to the initial volume:

$$\tau = J\sigma \quad (4.15)$$

J is the Jacobian, and represents the volume change during the deformation:

$$J = \det(\mathbf{F}) \quad (4.16)$$

However, both Cauchy and Kirchhoff stresses are in terms of spatial quantities. To determine stresses in terms of material quantities (i.e., the undeformed body), one must use the First and Second Piola-Kirchhoff stresses. The first Piola-Kirchhoff stress

tensor expresses the current force with respect to the undeformed area and is defined as:

$$P = \tau \mathbf{F}^{-T} = J \sigma \mathbf{F}^{-T} \quad (4.17)$$

However, the first Piola-Kirchoff stress is not symmetric. The second Piola-Kirchoff stress tensor is symmetric, and maps the current force back to the undeformed body in the similar manner that the deformation gradient maps material points to spatial points. The second Piola-Kirchoff stress tensor is defined as:

$$S = J \mathbf{F}^{-1} \sigma \mathbf{F}^{-T} \quad (4.18)$$

4.1.2 Virtual Work Equation

To evaluate the stress state in a domain, we can use a similar approach as before for thermal analysis, and use a set of any admissible weighting functions, W_i , over the domain, applied to (4.1):

$$\int_{\Omega} W_i (\nabla \cdot \sigma + b) d\Omega = 0 \quad (4.19)$$

We can apply Green's identity to the first term in (4.19) to get the weak form of the equation:

$$\int_{\Omega} W_i(\nabla \cdot \sigma) d\Omega = - \int_{\Omega} \nabla W_i \sigma d\Omega + \int_{\Gamma} W_i(\sigma \cdot n) d\Gamma \quad (4.20)$$

Applying the above identity to (4.19), we obtain:

$$\int_{\Gamma} W_i(\sigma \cdot n) d\Gamma - \int_{\Omega} \nabla W_i^T \sigma d\Omega + \int_{\Omega} W_i^T b d\Omega = 0 \quad (4.21)$$

(4.21) is commonly referred to as the virtual work equation, and is valid for any deformation for a continuum. We can add further that for any body in equilibrium undergoing small virtual displacements, the internal virtual work is equal to the external virtual work:

$$\int_{\Omega} (\epsilon^{*T} \sigma) d\Omega = \int_{\Omega} \delta^{*T} b d\Omega + \int_{\Gamma} \delta^{*T} T d\Gamma \quad (4.22)$$

where:

ϵ^* is the virtual strain,

δ^* is the virtual displacement.

The LHS of (4.22) represents the internal virtual work; the actual stress, σ undergoing the virtual strain, ϵ^* , corresponding to the virtual displacement, δ^* . The RHS of (4.22) represents the external virtual work; the actual body forces and traction forces undergoing virtual displacements, δ^* . To obtain the approximate FEM solution, we can choose trial functions, such that:

$$\delta^* = N_i(\mathbf{r})\delta_i^* \quad (4.23)$$

$$\delta^{*T} = \delta_i^{*T} N_i(\mathbf{r})^T$$

where:

δ_i^* is the nodal virtual displacement.

4.1.3 Elastic Stress Analysis

For an elastic stress state, where each stress state has a unique strain state associated with it, we can relate stress so that is in a form where it relates to the nodal displacement:

$$[\sigma] = [D][\varepsilon_e] = [D][B][u_i] \quad (4.24)$$

where:

D is the elasticity tensor mapping stress to strain,

B is the strain to displacement map,

u_i is the displacement at node i.

Substituting (4.23) and (4.24) into (4.22), and now using matrix notation, gives:

$$\delta_i^{*T} \left[\int_{\Omega} [B]^T [D] [B] [u_i] d\Omega - \int_{\Omega} [N]^T [b] d\Omega - \int_{\Gamma} [N]^T [\tau] d\Gamma \right] = 0 \quad (4.25)$$

Since the weighting functions chosen are completely arbitrary, the above equation is true for any admissible weighting function, and thus:

$$\int_{\Omega} [B]^T [D] [B] [u_i] d\Omega - \int_{\Omega} [N]^T [b] d\Omega - \int_{\Gamma} [N]^T [\tau] d\Gamma = 0 \quad (4.26)$$

We should note that the above equation should be summed over each element in the domain. Thus, we can again simplify the notation by defining the following:

1. Defining the stiffness matrix:

$$\int_{\Omega} [B]^T [D] [B] d\Omega = [K] \quad (4.27)$$

2. Defining the nodal displacement vector.

$$[u_i] \quad (4.28)$$

3. Defining the external loading vector.

$$\int_{\Omega} [N]^T [b] d\Omega + \int_{\Gamma} [N]^T [\tau] d\Gamma = [F] \quad (4.29)$$

For a three-dimensional system with N nodes, $[K]$ and $[F]$ are of dimension $3N \times 3N$, and $[u_i]$ is of length $3N$. As in previous analyses, Dirichlet boundary conditions (DBC) can be applied to nodal displacements (i.e., fixing a node's displacement to a prescribed values), and the subsequent rows in columns in $[K]$ and $[F]$ can be eliminated, saving computational time. As with the other solvers, we are left with the reduced FEM matrix equation:

$$[K][u_i] = [F] \quad (4.30)$$

4.1.4 Elastic-Plastic Stress Analysis

The above analysis is valid for an elastic stress state only, where the stress state at any loading step, n , stays within the yield surface F . The main difference between an elastic and plastic FEM analysis is that as stated before, plastic deformations are history dependent. As such, the stress and strain are not unique to one another, and an incremental theory must be used. Thus, we can use the same derivation as above, but we must change (4.24) such that:

$$[\Delta\sigma] = [D_{ep}][\Delta\varepsilon] = [D_{ep}][B][\Delta u_i] \quad (4.31)$$

or:

$$[\Delta\sigma] = [D]([\Delta\varepsilon] - [\Delta\varepsilon_p]) = [D](\Delta\varepsilon_e) \quad (4.32)$$

where:

D_{ep} is the elastic-plastic D matrix,

D is the elastic D matrix,

$\Delta\varepsilon$ is the total strain increment,

Δu_i is the nodal displacement increment.

4.1.5 General Plasticity Theory - FEM Formulation

The theory presented here is from that of [41], [42], [50]. We first start by decomposing the deformation gradient $\mathbf{F}$ into elastic and plastic parts as such:

$$\mathbf{F} = \mathbf{F}^e \mathbf{F}^p \quad (4.33)$$

Next, we assume that there is a free energy function ψ associated with the stored energy from the elastic deformation, and if we assume the material is isotropic, then it is implied that ψ depends upon the elastic portion of the deformation gradient. This is done through the elastic portion of the left Cauchy-Green strain tensor:

$$\mathbf{b}^e = \mathbf{F}^e \mathbf{F}^{eT} \quad (4.34)$$

As well, we can define a vector, $\mathbf{q}$, which is a vector of internal variables, that can include hardening parameters, such that:

$$\mathbf{q} = -\frac{\partial\psi(\mathbf{b}^e, \xi)}{\partial\xi} \quad (4.35)$$

$$\psi = \psi(\mathbf{b}^e, \xi) \quad (4.36)$$

where:

ξ is a vector of hardening variables conjugate with $\mathbf{q}$.

To describe the dissipation associated with plastic flow, we can define a dissipation function φ such that:

$$\varphi = \tau \cdot d - \frac{d\psi}{dt} \quad (4.37)$$

where:

d is the rate of deformation tensor, defined as:

$$d = \frac{1}{2}(\nabla\mathbf{V} + \nabla\mathbf{V}^T) \quad (4.38)$$

$\mathbf{V}$ is the spatial description of velocity:

$$\mathbf{V} = \dot{\mathbf{U}} \quad (4.39)$$

From reference [41], we get the relation:

$$\dot{\mathbf{b}}^e = \nabla \mathbf{V} \cdot \mathbf{b}^e + \mathbf{b}^e \cdot \nabla \mathbf{V} + L[\mathbf{b}^e] \quad (4.40)$$

where:

$L(\mathbf{b}^e)$ is the Lie derivative of $\mathbf{b}^e$, defined as:

$$L[\mathbf{b}^e] = \mathbf{F} \frac{\partial \mathbf{C}^p^{-1}}{\partial t} \mathbf{F}^T \quad (4.41)$$

$\mathbf{C}^p$ is the plastic right Cauchy-Green strain tensor:

$$\mathbf{C}^p = \mathbf{F}^p{}^T \mathbf{F}^p \quad (4.42)$$

For Von mises materials, the plastic flow is incompressible, and thus:

$$\det(\mathbf{F}^p) = \det(\mathbf{C}^p) = 1 \quad (4.43)$$

When we apply the second law of thermodynamics to the dissipation function, we get the following relations:

$$\tau = 2\nabla_{\xi}(\mathbf{b}^e, \xi) \mathbf{b}^e \quad (4.44)$$

$$\varphi = \tau \cdot \left(-\frac{1}{2} L\mathbf{b}^e^{-1}\right) + \mathbf{q} \dot{\xi} \geq 0 \quad (4.45)$$

Letting the yield surface be denoted as $\phi(\tau, \mathbf{q})$, we can say that for a given $L[\mathbf{b}^e]$ and $\dot{\xi}$, the actual state $(\tau, \mathbf{q})$ will be the one that maximizes the dissipation function φ , so long as $\phi(\tau, \mathbf{q}) \leq 0$. The evolution equations for such a system are:

$$-\frac{1}{2}L[\mathbf{b}^e] = \gamma \frac{\partial \phi(\tau, \mathbf{q})}{\partial \tau} \mathbf{b}^e \quad (4.46)$$

$$\dot{\xi} = \gamma \frac{\partial \phi(\tau, \mathbf{q})}{\partial \mathbf{q}} \quad (4.47)$$

$$\phi(\tau, \mathbf{q}) \leq 0 \quad (4.48)$$

$$\gamma \geq 0 \quad (4.49)$$

$$\gamma \phi(\tau, \mathbf{q}) = 0 \quad (4.50)$$

where:

γ is a variable indicating whether yielding has occurred, if $(\tau, \mathbf{q})$ is inside the yield surface, $\gamma = 0$.

The above evolution equations assume no viscous effects. To account for viscous effects, one must introduce a viscosity term, η . We also introduce a function w , which acts like a pseudo-yield surface, such that:

$$w(z) > 0 \quad \text{if } z > 0 \quad (4.51)$$

$$w(z) = 0 \quad \text{if } z \leq 0 \quad (4.52)$$

Thus, like before we aim to maximize

$$\varphi - \frac{w(\phi(\tau, \mathbf{q}))}{\eta} \quad (4.53)$$

over $(\tau, \mathbf{q})$, but without the constraint of $\phi(\tau, \mathbf{q}) \leq 0$ as before. In other words, the stress state can exist outside the yield surface. Thus, the evolution equations become:

$$-\frac{1}{2}L[\mathbf{b}^e] = \gamma \frac{\partial \phi(\tau, \mathbf{q})}{\partial \tau} \mathbf{b}^e \quad (4.54)$$

$$\dot{\xi} = \gamma \frac{\partial \phi(\tau, \mathbf{q})}{\partial \mathbf{q}} \quad (4.55)$$

$$\gamma = \frac{g(\tau, \mathbf{q})}{\eta} \quad (4.56)$$

$$g = w' \quad (4.57)$$

It is assumed that g is similar to w in the sense that:

$$g(z) \geq 0 \quad \text{if } z \geq 0 \quad (4.58)$$

$$g(z) = 0 \quad \text{if } z \leq 0 \quad (4.59)$$

The evolution equations for viscoplasticity above reduce to that of rate independent plasticity as $\eta \rightarrow 0$. This implies that the relaxation time goes to zero, and the yield function becomes a hard constraint.

4.1.6 Infinitesimal Strain J_2 Flow Theory - FEM Formulation

Here we take the general framework of the previous section and take the case of infinitesimal strain with kinematic and isotropic hardening. Since for J_2 deformation theory plastic flow is dependent upon the deviatoric stress quantities only, we can define a yield function as:

$$\phi(\mathbf{s}, \beta, \alpha) = ||(\mathbf{s} - \beta)|| - \sqrt{\frac{2}{3}}K(\alpha) \quad (4.60)$$

where:

$\mathbf{s}$ is the deviatoric part of the Kirchoff stress tensor, τ :

$$\mathbf{s} = \mathbf{s} - \frac{1}{3}tr(\mathbf{s})I \quad (4.61)$$

β is the back stress (center of the yield surface), characterizing kinematic hardening,

α is the isotropic hardening parameter, defining the expansion of the yield surface in stress space,

$K(\alpha)$ is the isotropic hardening function, assumed to be of the form:

$$K(\alpha) = \sigma_y + K'\alpha \quad (4.62)$$

K' is a constant,

σ_y is the yield stress.

The deformation and internal variables follow the following evolution equations:

$$\dot{\varepsilon}^p = \gamma \mathbf{n} \quad (4.63)$$

$$\dot{\beta} = \frac{2}{3} \gamma H(\alpha) \mathbf{n} \quad (4.64)$$

$$\dot{\alpha} = \sqrt{\frac{2}{3}} \gamma \quad (4.65)$$

$$\dot{\mathbf{s}} = 2\mu(\dot{\mathbf{e}} - \dot{\varepsilon}^p) \quad (4.66)$$

$$= 2\mu(\dot{\mathbf{e}} - \gamma \mathbf{n}) \quad (4.67)$$

$$= 2\mu(\mathbf{I} - \gamma(\mathbf{n} \otimes \mathbf{n}))\dot{\mathbf{e}} \quad (4.68)$$

where:

H is the kinematic hardening function,

$\mathbf{e}$ is the deviatoric strain:

$$\mathbf{e} = \varepsilon - \frac{1}{3} \text{tr}(\varepsilon) \mathbf{I} \quad (4.69)$$

$\mathbf{n}$ is a normal to the yield surface:

$$\mathbf{n} = \frac{(\mathbf{s} - \beta)}{\|(\mathbf{s} - \beta)\|} \quad (4.70)$$

With that in mind, we define the constitutive equation:

$$\dot{\sigma} = D_{ep} \dot{\varepsilon} \quad (4.71)$$

$$D_{ep} = \lambda(I \otimes I) + 2\mu(\mathbf{I} - \gamma(\mathbf{n} \otimes \mathbf{n})) \quad (4.72)$$

where:

$\otimes$ is the tensor product operator,

μ is the shear modulus,

λ is Lamé's constant,

$\mathbf{I}$ is the fourth order identity tensor.

I is the second order identity tensor transformed into a 6×1 vector:

$$I = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \end{bmatrix}^T \quad (4.73)$$

$$I \otimes I = \begin{bmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \quad (4.74)$$

where:

$\mathbf{1}$ is a 3x3 matrix of all ones,

$\mathbf{0}$ is a 3x3 matrix of all zeroes.

If we introduce viscoplasticity, we can replace the evolution equations for rate independent plasticity with:

$$\dot{\varepsilon}^p = \frac{g(\phi(\sigma, \beta, \alpha))}{\eta} \mathbf{n} \quad (4.75)$$

$$\dot{\beta} = \frac{2}{3} \frac{g(\phi(\sigma, \beta, \alpha))}{\eta} H(\alpha) \mathbf{n} \quad (4.76)$$

$$\dot{\alpha} = \sqrt{\frac{2}{3}} \frac{g(\phi(\sigma, \beta, \alpha))}{\eta} \quad (4.77)$$

$$\dot{\mathbf{s}} = 2\mu(\dot{\mathbf{e}} - \dot{\varepsilon}^p) \quad (4.78)$$

$$= 2\mu(\dot{\mathbf{e}} - \frac{g(\phi(\sigma, \beta, \alpha))}{\eta} \mathbf{n}) \quad (4.79)$$

$$= 2\mu(\mathbf{I} - \frac{g(\phi(\sigma, \beta, \alpha))}{\eta} (\mathbf{n} \otimes \mathbf{n})) \dot{\mathbf{e}} \quad (4.80)$$

4.1.6.1 Incremental Strain Algorithm

This algorithm is outlined in [42], and is used in the code used in the analysis. If at a time t_n we are given the following data:

- $u_n \rightarrow$ the nodal displacements at t_n ,
- $\sigma_n \rightarrow$ the stress at all Gauss points at time t_n ,
- $\alpha_n \rightarrow$ the hardening parameter at all Gauss points at time t_n ,
- $\beta_n \rightarrow$ the back stress at all Gauss points at time t_n .

If we are given a displacement increment $dU : \Omega \rightarrow R^3$, in some body Ω , which is the displacement increment during some time interval $\Delta t = t_{n+1} - t_n$ such that:

$$u_{n+1} = u_n + dU \quad (4.81)$$

Once this data is known, we proceed to evaluate the following at each Gaussian integration point:

Step 1 Compute the strain increment, and its deviatoric counterpart, and the deviatoric stress at time t_n :

$$d\varepsilon = \frac{1}{2}(\nabla dU + \nabla dU^T) \quad (4.82)$$

$$d\mathbf{e} = d\varepsilon - \frac{1}{3}tr(d\varepsilon)I \quad (4.83)$$

$$\mathbf{s}_n = \sigma - \frac{1}{3}tr(\sigma)I \quad (4.84)$$

$$(4.85)$$

Step 2 Assuming all elastic strains, calculate a trial deviatoric stress such that:

$$\mathbf{s}^{\text{tr}}_{n+1} = 2\mu d\mathbf{e} + \mathbf{s}_n \quad (4.86)$$

Test this trial stress for plastic loading:

$$\hat{\phi}_{n+1}^{\text{tr}} = ||\mathbf{s}^{\text{tr}}_{n+1} - \beta_n|| - \sqrt{\frac{2}{3}}K(\alpha_n) \quad (4.87)$$

If:

$$\hat{\phi}_{n+1}^{\text{tr}} \leq 0 \quad (4.88)$$

then the trial stress is one where the strains are totally elastic, and thus we can update the stress and proceed to the next timestep:

$$\mathbf{s}_{n+1} = \mathbf{s}^{\text{tr}}_{n+1} \quad (4.89)$$

$$\alpha_{n+1} = \alpha_n \quad (4.90)$$

$$\beta_{n+1} = \beta_n \quad (4.91)$$

$$\Delta\gamma = 0 \quad (4.92)$$

If the stress step is plastically loaded, then we must perform one more step before updating.

Step 3 This is the “radial return” step, where we solve for $\Delta\gamma$ from the following equation:

$$\hat{\phi}_{n+1}^{\text{tr}} - g^{-1} \frac{\eta \Delta\gamma}{\Delta t} - 2\mu \Delta\gamma \quad (4.93)$$

$$-\sqrt{\frac{2}{3}}[K(\alpha_n + \sqrt{\frac{2}{3}}\Delta\gamma) - K(\alpha_n)] \quad (4.94)$$

$$-\sqrt{\frac{2}{3}}[H(\alpha_n + \sqrt{\frac{2}{3}}\Delta\gamma) - H(\alpha_n)] \quad (4.95)$$

$$= 0 \quad (4.96)$$

Note that in this step the stress increment returns back towards the origin of the yield surface, hence the term “radial return”.

Step 4 Update the plastically loaded step:

$$\mathbf{n} = \frac{(\mathbf{s}^{\text{tr}}_{n+1} - \beta_n)}{||\mathbf{s}^{\text{tr}}_{n+1} - \beta_n)||} \quad (4.97)$$

$$\mathbf{s}_{n+1} = \mathbf{s}^{\text{tr}}_{n+1} - 2\mu\Delta\gamma\mathbf{n} \quad (4.98)$$

$$\alpha_{n+1} = \alpha_n + \sqrt{\frac{2}{3}}\Delta\gamma \quad (4.99)$$

$$\beta_{n+1} = \beta_n + \sqrt{\frac{2}{3}}[H(\alpha_{n+1}) - H(\alpha_n)]\mathbf{n} \quad (4.100)$$

$$\sigma_{n+1} = [tr(\sigma_n) + \kappa tr(d\varepsilon)]I + \mathbf{s}_{n+1} \quad (4.101)$$

where:

κ is the bulk modulus.

Chapter 5

Theory of Fracture

5.1 Deformation Mechanisms

Crystalline metals deform in numerous ways. Plastic deformation via plastic flow is a kinetic process, requiring movement of some sort, usually of dislocations or grain boundaries. Reference [12] breaks deformation mechanisms down into several categories, and will be discussed individually.

5.1.1 Brief Dislocation Definition

Before going further into deformation mechanisms, it is important to have some idea of how dislocations influence deformations in the material in question. A dislocation is a defect in a material, and is characterized by a Burgers vector $\mathbf{b}$, which is of an order

of magnitude of the lattice parameter, a_o . When the Burgers vector is perpendicular to the dislocation line, the dislocation is called an edge dislocation, and when the Burgers vector is parallel to the dislocation line, the dislocation is called an screw dislocation. Dislocations create stress and strain fields surrounding it that are not thermodynamically stable. As shown below, these dislocations have a large influence in plastic flow at lower temperatures in a crystalline material.

5.1.2 Elastic Collapse

When a material fails catastrophically at its ideal strength, it is said to have undergone elastic collapse. The “ideal strength” is that which a perfect crystal suddenly destroys itself by tearing, or when the strain rate becomes infinite.

5.1.3 Low Temperature Plasticity

Generally at low temperatures, plastic deformation is the result of dislocation glide, provided there are adequate independent slip systems available. The rate of glide, or flow, is dictated by obstacles, such as other dislocations, precipitates, solutes, grain boundaries, or friction with the lattice itself. Obstacles can generally be classed into two areas, discrete obstacles (i.e., precipitates) bypassed or cut through individually by the moving dislocation, or extended and diffuse obstacles (i.e., lattice friction), barriers that are collectively overcome by dislocation motion. Each of these classes

tends to have a different flow rate equation associated with it, with both equations dependent upon the energy to overcome the obstacle.

During deformation, the number of dislocations increases dramatically, often by orders of magnitude. After a certain activation shear stress is reached, sufficient energy is present to create more dislocations. However, not all dislocations are mobile; mobility is dependent upon the dislocation being on a specific gliding plane.

Upon encountering grain boundaries, dislocations become blocked; thus grain boundaries are another type of discrete obstacle. These dislocations pile-up at the grain boundary. This pile-up is one cause of hardening in crystalline metals, and is characterized by back stress. When these dislocations pile up under a particular stress state, when this stress state is reversed, the dislocations “unpile” and thus, the yield strength to initiate plastic flow in the reversed stress state after initial plastic flow is not equal to that in the initial stress state, it is less. This is known as the Baushinger effect, and is common in crystalline metals. Generally at lower temperatures, the yield strength of most metals is not dependent upon the flow rate, so this type of plasticity is commonly known as rate-independent plasticity.

5.1.4 High Temperature Plasticity

At temperatures approaching and over half the melting temperature of a metal, rate dependent plasticity, or creep, becomes prominent. Generally there are three portions

in a typical creep strain vs. time graph, a primary zone, a secondary zone, and a tertiary zone. The primary zone occurs at low creep strains where the creep strain rate eventually slows and reaches a steady state. At this point, it is said to have reached the secondary zone. As the metal continues to creep, there reaches a point where the creep strain becomes such that the creep strain rate again begins to increase. Thus, it has said to enter the final or tertiary zone, and the rate increases until final rupture. In the tertiary zone, microcracks have already formed, thus initiating tertiary creep. These cracks may continue to grow in tertiary creep zones, but final rupture occurs upon coalescence of these growing cracks.

Power law creep is typically of the form:

$$\frac{d\varepsilon_{ij}^c}{dt} = A_o \exp\left(\frac{-Q}{RT}\right) \sigma^n \quad (5.1)$$

where:

A_o, n are creep material constants,

Q is the activation energy for creep,

R_g is the universal gas constant,

$d\varepsilon_{ij}^c/dt$ is the creep strain rate,

T is the absolute temperature of the material.

The exponent n ranges in values from 3 to 10, depending on the material and the temperature at which creep is occurring. As mentioned above, the creep strain rate can depend upon whether or not the creep is at steady state (secondary) or transient (primary or tertiary).

Below half the melting temperature, glide controlled creep can happen in most metals, if the activation energy for glide is small. At higher temperatures, the glide becomes a predominantly thermally activated process, and creep behaviour starts.

At higher temperatures, glide can be accompanied by dislocation climb. Essentially, obstacles that once prevented dislocation movement can now be climbed over by the dislocation. The dislocation then glides to the next obstacle, and the process repeats. The main difference from climb versus glide dislocation movements is that climb involves the diffusion of vacancies to or from the climbing dislocation, whereas at higher temperatures, glide is a thermally activated process.

However, at high stresses the power law can break down, and strain rates are higher than predicted. This process is a transition from only a climb controlled process to a glide and climb controlled process. Dynamic recrystallization can also occur at high temperatures as well as the above mechanisms. When this occurs, the strain rate is higher than predicted.

Creep can also be diffusion controlled. Diffusion through and around the grain boundaries creates a strain, assuming that there is grain boundary sliding as well as

the diffusion. When lattice diffusion dominates creep, it is known as Nabarro-Herring creep, and if grain boundary diffusion dominates, it is Coble creep. Coble creep is associated with lower temperatures than Nabarro-Herring creep.

5.2 Microscopic Aspects of Fracture

Considering the above methods of damage and deformation, one can apply these ideas and concepts to a fracture mechanics aspect, when considering the microscopic scale. A stress acting normal to a crystallographic plane, or a shear stress acting parallel to this plane can both break atomic bonds and cause motion along this plane. The critical stress at which these two phenomena occur are called the theoretical cohesive or cleavage stress and the theoretical gliding or shear stress, respectively [33]. Estimates for cohesive stress and glide stresses are given as:

$$\sigma_s = \frac{E}{10} \quad (5.2)$$

$$\gamma_s = \frac{Ea_o}{100} \quad (5.3)$$

where:

E is the Young's modulus,

a_o is the lattice parameter of the crystalline in question (in the order of nm),

σ_s is the estimated theoretical cohesive or cleavage stress,

γ_s is the estimated theoretical gliding or shear stress.

Experiments have shown that plasticity, in some form, is necessary to initiate cracking. As mentioned above, when the amount of plastic deformation is low before rupture, the rupture is said to be brittle in nature, whereas large amounts of plastic deformation before rupture indicate a ductile rupture.

Fractures can be classified into different methods, here in order of increasing ductility:

Brittle intergranular fracture, which, as the name implies, the crack propagates along grain boundaries.

Transgranular cleavage fracture, where the crack propagates through grains under cleavage stress.

Shear fracture, where the crack propagates through grains by shear through the grains.

Ductile dimple fracture, where the crack propagates through grains via the nucleation, growth, and coalescence of pores within the material.

5.2.1 Brittle Cleavage Fracture

In a polycrystalline solid under loading, a cleavage crack can occur and propagate in three main stages. First, the dislocations pile up at a grain boundary or particle, and nucleate a microcrack. This microcrack then grows through the grain at which it was nucleated. Finally, the crack encounters the next grain along the cleavage plane, called grain boundary “crossing”, and the process starts again. Grains on two sides of a boundary have different orientations, and thus there are different cleavage plane orientations. The continuation of the cleavage crack relies upon meeting of the two planes at a mean point by means of a “river” or “step” between the two [33]. At each stage here, an energy barrier must be overcome to continue the crack growth.

5.2.2 Brittle Intergranular Fracture

The methods of intergranular fracture are similar to cleavage fracture, in that the initiation of a microcrack occurs ahead of dislocation pile-ups at grain boundaries. However, the extension of the crack does not go through the grain, it follows along the grain boundaries, under the action of a normal applied stress. Grain boundary crossing is not applicable to intergranular fracture [33]. For intergranular fracture, the cohesive energy of the grain boundary must be lower than that of the surface energy of the grain, or else a transgranular fracture will occur.

5.2.3 Ductile Fracture

As mentioned above, ductile fracture occurs in three main stages. First, pores or voids are nucleated within a material. Next, these voids grow to a critical size, and then finally the voids grow to a point of coalescence, where they become linked to one another and create the final rupture path.

Voids are nucleated at second phase particles. This occurs in two main ways. The first way is by the fracture of non-metallic inclusions by dislocation pile-up, and the growth occurs assuming the stress state is low enough to prevent a cleavage type fracture. The second way is by separation of particles with the matrix by a process known as interfacial decohesion [33].

At early stages of void growth, analysis of the growth can be considered by considering each individual void by itself. Their interaction is neglected. Typically in ductile materials, elastic deformations are neglected, and plastic deformations are considered. For microscopic modeling, either cylindrical voids or spherical voids are considered. Derivations of growth rates can be found in [33, pgs. 136-140]. It has been shown that as stress triaxiality ratio increases (with stress triaxiality ratio defined as σ_H/σ_{eq} , where σ_H is the hydrostatic stress and σ_{eq} is the equivalent stress, defined by a von Mises type equation.) ductility decreases. These voids continue to grow until they reach the point of coalescence.

An approach has been suggested by Gurson [33, pg. 142], where it is assumed

that a body contains voids of equal size, with a void volume fraction, f . Gurson first derived a potential function for flow, Φ as:

$$\Phi = \left(\frac{\sigma_{eff}}{\sigma_y} \right)^2 + 2f \cosh \left(\frac{3\sigma_H}{2\sigma_y} \right) - (1 + f^2) = 0 \quad (5.4)$$

where:

σ_{eff} is the effective stress, which can be found using von Mises equation,

σ_y is the yield stress,

σ_H is the hydrostatic stress,

f is the volume fraction of voids in the material.

Essentially, in (5.4), when $f = 0$, the yield surface is that of a von Mises material, and as f approaches 1, the yield surface shrinks to a point where the load carrying capacity of the material vanishes. Actual fracture of the material occurs at an f value much less than 1. Since (5.4) does not account for void interaction and coalescence, it overestimates ductility. To account for this, a modification of (5.4) is proposed [33, pg. 144]:

$$\Phi = \left(\frac{\sigma_{eff}}{\sigma_y} \right)^2 + 2fq_1 \cosh \left(\frac{3q_2\sigma_H}{2\sigma_y} \right) - (1 + q_3f^2) = 0 \quad (5.5)$$

where:

q_1, q_2, q_3 are factors for void growth ($q_1 = 1.5, q_2 = 1.0, q_3 = q_1^2$ [33, pg. 144]).

(5.5) tends to model overall behaviour well, but does not take into account a rapid drop in stress carrying capacity as fracture conditions are approached. To take this into account:

$$f = \begin{cases} f & \text{for } f \leq f_c \\ f_c - k(f - f_c) & \text{for } f > f_c \end{cases} \quad (5.6)$$

k is a factor for void growth effects, (usually 3.4 or 5.16 [50]),

f_c is a critical volume fraction of voids where, beyond which, coalescence effects cannot be neglected, and varies between 0.04 and 0.15 ([33, pg. 145], [50]).

As well, it has been determined that depending on the material, the initial volume fraction of voids, f_o , is generally in the range of 0 to 0.06. The void volume fraction growth rate can be decomposed into 2 parts, one for nucleation rate of the voids, and one for growth rate of the existing voids:

$$\dot{f} = \dot{f}_n + \dot{f}_g \quad (5.7)$$

where:

$\dot{f}_n$ is the void nucleation growth rate,

$\dot{f}_g$ is the void growth rate,

Models for the both growth rates have been proposed [50]:

$$\dot{f}_g = (1 - f)\dot{\varepsilon}_{kk}^p \quad (5.8)$$

$$\dot{f}_n = B\dot{\varepsilon}_{eff}^p \quad (5.9)$$

$$B = \frac{f_N}{s_N\sqrt{2\pi}} \exp\left(-0.5\left(\frac{\dot{\varepsilon}_{eff}^p - \dot{\varepsilon}_N}{s_N}\right)^2\right) \quad (5.10)$$

where:

$\dot{\varepsilon}_{kk}^p$ is the trace of the plastic strain rate tensor,

$\dot{\varepsilon}_{eff}^p$ is the effective plastic strain rate,

$\dot{\varepsilon}_N$ is the maximum plastic strain rate at which void nucleation rate is a maximum,

s_N is the standard deviation of the nucleation function, taken as 0.1,

f_N is the volume fraction of voids that can nucleate, taken as 0.02.

5.2.4 Shear Fracture

Shear fracture essentially occurs when shear localizes in the location of failure, along shear bands at specific angles with respect to the load axis. The shear is transgranular. Small slip or shear band occur at lower strains, and grow in number and size as the strain is increased. The final fracture surface tends to nucleate at a free surface and follows along these formed shear bands.

5.2.5 Ductile-Brittle Transition

In typical ductile fracture applications, there comes a point when the fracture actually becomes brittle in nature. At this point, the fracture has said to undergone a ductile-brittle transition. This can occur after a critical stress (i.e., the cleavage stress) is reached in the part after a certain amount of deformation. Temperature can cause this transition to occur prematurely, as a sufficiently low temperature will cause even the most ductile metals to behave in a brittle manner, since the energy to move dislocations increases as temperature decreases.

Grain size plays a role in this transition. All other things being equal, coarser grains tend towards a cleavage, or brittle fracture, whereas ductile fracture occurs more frequently in smaller grained specimens. With finer grains, the cleavage stress for brittle fracture occurs only when a certain amount strain hardening (i.e., dislocation pile-up), whereas coarser grains can fracture at stresses near the yield stress with no deformation.

5.2.6 Fatigue Cracking

Commonly, fatigue crack growth is determined by a variation in the stress intensity factor K along the crack tip in question during a fatigue loading cycle. This is determined by taking the difference between the maximum value of K and the minimum value of K and taking the difference. Once this is determined, a typical Paris-Erdogan

crack growth equation for fatigue crack growth is given as:

$$\frac{da}{dN} = c(\Delta K)^m \quad (5.11)$$

where:

$\frac{da}{dN}$ is the crack extension da per cycle, N ,

c, m are material constants,

ΔK is the change in stress intensity factor during a typical loading cycle.

Generally, it is best to take this difference ΔK after cyclic steady state is achieved, but if it is taken before this is achieved, the difference will be larger than during steady state. However, if values are taken before steady state is achieved, then the number of cycles to a certain crack length a by (5.11) will end up being less than that if ΔK values from steady state were to be used [50].

5.3 Fracture Mechanics Review

A crack is generally viewed as a creation of two new surfaces inside a material, with a leading edge known as the crack tip. For a typical crack, there are three modes of relative motion or tearing of the two planes. The first, imaginatively named Mode I, is the opening mode, which can be visualized by a pure tension test on a notched

specimen. The second and third deal with shears, with Mode II being in-plane shear, and Mode III being out-of-plane shear. Plane strain is more likely to occur at the center of thick plates, where there is too much material to deform in a third direction, and plane stress is likely to occur near free surfaces. Recall, that there can be no stress component normal to a free surface (assuming there is no external pressure or traction on it).

5.3.1 Fully Elastic Treatment

At the crack tip, there can be stress concentrations due to the “notch” effect of the tip. Thus, there is a stress increase in this region, compared to the nominal region surrounding the crack. To characterize this, a stress intensity factor, K is introduced in elastic fracture mechanics. It depends on the geometry of the cracking body, the mode of fracture, and the magnitude of the loading. K is the product of the stress and the square root of length, and thus has S.I. units of $MPa\sqrt{m}$. As well, each mode of cracking can have a stress intensity factor associated with it (i.e., K_I , K_{II} , K_{III}). These can be added using a superposition rule, depending on the loading conditions. K_I represents the stress intensity factor for an opening cracking mode, K_{II} for in-plane sliding, and K_{III} for the anti-plane sliding mode.

For elastic treatment of cracks, there is a singularity around the tip of the stress (i.e., it tends towards infinity) as the radius around the tip goes to zero. It has been

shown that this singularity is of the form $1/\sqrt{r}$, where r is the radial distance from the crack tip. For two-dimensional symmetry, the equations for stresses around the crack tip can be expressed in terms of radial distance r from the tip, and an angle θ from the reference x-axis, in the form of:

$$\sigma_{ij} = \frac{K_n}{\sqrt{2\pi r}} \tilde{\sigma}_{ij}(\theta) \quad (5.12)$$

where:

σ_{ij} is the stress component,

K_n is the stress intensity factor mode, ($n = \text{I, II, or III}$),

$\tilde{\sigma}_{ij}(\theta)$ is a function of θ , usually expressed as a series of trigonometric terms based on the angle θ at which one is describing the stress field around the crack tip (see [33, pgs. 12-14]).

As for the extension of a crack, one must consider the energy needed to extend the crack. This was first proposed by Griffith for very brittle materials. Consider a crack of length a , and we wish to determine the energy required to extend the crack an infinitesimal distance δa . For this, we need:

$$G\delta a = \delta W - \delta U \quad (5.13)$$

where:

δW is the variation of the work done by external forces,

δU is the variation of the stored elastic strain energy,

$G\delta a$ is the variation in available energy for crack extension.

In (5.13), G can have dimensions of energy per unit area, or force per unit length, and is referred to as the strain energy release rate with respect to crack length, or crack extension force. It can be shown that the relationship between K and G for an isotropic material is:

$$G = \frac{K_I^2}{E'} + \frac{K_{II}^2}{E'} + \frac{K_{III}^2}{2\mu} \quad (5.14)$$

where:

μ is the shear modulus,

E' is defined as:

$$E' = \begin{cases} E & \text{for plane stress} \\ E/(1 - \nu^2) & \text{for plane strain} \end{cases} \quad (5.15)$$

E is Young's modulus,

ν is Poisson's ratio,

Unstable crack extension is said to occur when the K reaches a critical value, K_C . That said, a crack can grow at values lower than K_C , but its growth is said to

be stable. Generally, the first mode of tearing is considered the most dangerous, and K_C is usually given in literature as K_{IC} . Conversely, due to (7.22), it can be said that the crack extends upon reaching a critical energy requirement, G_C . The energy criteria is that to create two new surfaces. For a transgranular fracture (i.e, through grains):

$$G_C = 2\gamma_S \quad (5.16)$$

For a intergranular fracture (i.e., along grain boundaries):

$$G_C = \gamma_A + \gamma_B + \gamma_{AB} \quad (5.17)$$

where:

γ_S is the energy required to form a new surface in one phase,

γ_A is the surface energy of phase A,

γ_B is the surface energy of phase B,

γ_{AB} is the cohesive energy of the grain boundary between phases A and B.

The preceeding treatment of fully elastic fracture mechanics is suitable for brittle materials such as glasses or ceramics.

5.3.1.1 Elastic Treatment With Plastic Zone

In the above section, any plastic deformation was neglected, although it is known that there will always be some degree of plasticity at the crack tip region in even slightly ductile materials. At low applied loads, there is a small zone around the crack tip that is entirely plastic. Even at higher loads, where the surrounding material may be plastic, there is still a region of higher plastic flow due to the stress concentration effect of the crack tip. This zone varies in size whether the material is in a plane stress or plane strain state. The size of this plastic zone is given by Irwin as a radius around the crack tip in the xy-plane for plane strain and plane stress, respectively, as:

$$r_p = \frac{1}{6\pi} \left(\frac{K_I}{\sigma_y} \right)^2 \quad (5.18)$$

$$r_p = \frac{1}{2\pi} \left(\frac{K_I}{\sigma_y} \right)^2 \quad (5.19)$$

There is an increase in the yield tensile stress under plane strain conditions, and thus, the factor of 1/6 in (5.18) versus 1/2 in (5.19).

If this zone of plastic yielding is sufficiently small, then we can still use linear elastic theory outlined above to characterize the crack tip region. In a thick plate, the plane stress plastic zone exists at the edges of the plate, whereas the plane strain

plastic zone occurs in the interior of a sufficiently thick plate. The dominance of one zone over another depends on the thickness of the plate being used.

The same energy arguments can be made as before (see (5.13)), although now a plastic strain energy must be added:

$$G_C = 2\gamma_S + \gamma_P \quad (5.20)$$

where:

γ_P is the plastic strain energy, in units of energy per unit crack length.

5.3.2 Elastic-Plastic Treatment

The preceding sections dealt with linear elastic fracture mechanics, with and without a plasticity correction. Here, we wish to extend the thoughts to one where plasticity is considered in a strain hardening material to determine fracture characteristics. For elastic-plastic, or non-linear treatment of cracks, there is again a singularity around the tip of the stress as the radius around the tip goes to zero. Similar to elastic fracture mechanics, it has been shown that this singularity is of the form $1/r$, where r is the radius from the crack tip. The general assumption is that of a power law hardening material, where in uniaxial tension the expression is:

$$\frac{\varepsilon}{\varepsilon_y} = \left(\frac{\sigma}{\sigma_y} \right)^n \quad (5.21)$$

where:

ε_y, σ_y are the reference yield strains and stresses, respectively,

n strain hardening exponent, depending on the material and the stress/strain states.

Thus, as with the elastic cases in (5.12), we can characterize the stress and strain states around the crack tip with (r, θ) components in the form of:

$$\frac{\sigma_{ij}}{\sigma_y} = \left(\frac{J}{I\sigma_y\varepsilon_y r} \right)^{1/(1+n)} \tilde{\sigma}_{ij}(\theta, n) \quad (5.22)$$

$$\frac{\varepsilon_{ij}}{\varepsilon_y} = \left(\frac{J}{I\sigma_y\varepsilon_y r} \right)^{n/(1+n)} \tilde{\varepsilon}_{ij}(\theta, n) \quad (5.23)$$

where:

I is a numerical non-dimensional factor usually given in tables

$\tilde{\varepsilon}_{ij}(\theta, n), \tilde{\sigma}_{ij}(\theta, n)$ are non-dimensional functions of θ and n , such that their maximum effective value is 1.

J is the “J-integral”.

5.3.3 J-integral derivation

If we consider a two-dimensional body where the crack faces are free of loading, we can define a contour integral around the crack tip, called the J-integral. This integral

can be used in plane stress or plane strain states. We can derive this J-integral as follows, as in [20]. If we consider a domain B_o with a boundary ∂B_o . We can say that for equilibrium:

$$\int_{\partial B_o} [N_\alpha S_{\alpha i} v_i - N_\alpha q_\alpha] dS_o = \int_{B_o} \frac{d}{dt} \left[U + \frac{1}{2} \rho v_i v_i \right] d\mathbf{X} \quad (5.24)$$

where:

α is an index across three dimensions, $\alpha = x, y, z$,

dS_o is a reference area of size dS_o ,

N_α is the unit normal,

q_α is a “nominal” heat flux vector,

$S_{\alpha i}$ is a component of the stress tensor in the undeformed configuration such that the traction, t_i is:

$$t_i = N_\alpha S_{\alpha i} \quad (5.25)$$

v_i is the velocity vector,

U is the internal energy state per unit mass,

ρ is the density,

$d\mathbf{X}$ is a vector referred to as the undeformed current configuration (material points).

(5.24) essentially says that for a crack growing at steady state, the rate of energy input into B_o (i.e. the LHS of (5.24)) must be equal to the rate of energy increase in B_o (i.e. the RHS of (5.24)). Note that this derivation is evaluated in the reference configuration with material points $\mathbf{X}$. For simplicity, body forces are neglected, and all heat sources other than those from the boundary of B_o are neglected as well.

With these underlying assumptions in place, we can say that the rate of change in internal energy, or the local energy equation, is:

$$\dot{U} = S_{\alpha i} v_{i,\alpha} - q_{\alpha,\alpha} \quad (5.26)$$

In the derivation from [20], he states that the domain B_o is fixed such that it always consists of the same spatial points. In other words, the area enclosed by B_o is fixed in space, but the material “slides” through it [50]. With reference to a crack tip, the domain B_o is fixed to the tip, but this tip can move through the material it is cracking. With this idea in hand, we change the notation of this domain from B_o to $B_o(t)$, since at any time t it can be defined. The rate of change of internal energy within this domain $B_o(t)$ is then:

$$\begin{aligned} \frac{d}{dt} \int_{B_o(t)} \left[U + \frac{1}{2} \rho v_i v_i \right] d\mathbf{X} &= \\ \int_{B_o(t)} \frac{d}{dt} \left[U + \frac{1}{2} \rho v_i v_i \right] d\mathbf{X} &+ \int_{\partial B_o(t)} \left[U + \frac{1}{2} \rho v_i v_i \right] (N_\alpha C_\alpha) dS_o \end{aligned} \quad (5.27)$$

where:

C_α are the components of the velocity of ∂B_o .

By substitution of (5.26) into the above equation, and by use of the divergence theorem, we can say the volume integral on the RHS of (5.27) satisfies (5.26):

$$\frac{d}{dt} \int_{B_o(t)} \left[U + \frac{1}{2} \rho v_i v_i \right] d\mathbf{X} = \int_{\partial B_o(t)} N_\alpha [S_{\alpha i} v_i - q_\alpha + C_\alpha (U + \frac{1}{2} \rho v_i v_i)] dS_o \quad (5.28)$$

We can apply (5.28) to find the energy flux into the tip of a moving crack. To apply the above equation to a moving crack in a plane, we define the following coordinate system such that:

$$X_1 < L(t) \quad X_2 = 0 \quad -\infty < X_3 < \infty \quad (5.29)$$

where:

$L(t)$ is the length of the crack from the beginning of the crack to the tip.

One can think of these coordinates as X_1 being parallel to the crack path, X_2 is normal to the cracked surface near the tip, with X_3 being tangent to the crack tip, or the cross product of X_1 and X_2 . For steady state crack growth in a homogeneous domain, the state of stress and deformation is considered to be independent of X_3 . If

we take a contour Γ around the crack tip in the (X_1, X_2) plane we can say the energy per unit length of entering this contour is:

$$CJ = \int_{\Gamma} [N_{\alpha}(S_{\alpha i}v_i - q_{\alpha})ds + (U + \frac{1}{2}\rho v_i v_i)C dX_2] \quad (5.30)$$

where:

C is the velocity of crack growth, $\dot{L}(t)$,

ds is an infinitesimal section of the contour Γ .

If we consider two concentric contours around the crack tip, and assume that all loading moves with the crack tip, we can say that the region bounded by these contours contains a constant amount of energy. Thus, the energy flux into either of these contours is constant, therefore the contour chosen is arbitrary. When this is the case:

$$v_i = -C \frac{\partial x_i}{\partial X_1} = -C x_{i,1} = -C F_{i,1} \quad (5.31)$$

where:

F is the deformation gradient.

$$J = \int_{\Gamma} [(U + \frac{1}{2}\rho C^2 F_{i,1} F_{i,1})dX_2 - N_{\alpha}(S_{\alpha i} F_{i,1} + C^{-1} q_{\alpha})ds] \quad (5.32)$$

We can make a few simplifying assumptions to make (5.32) appear as it does in most fracture mechanics texts. We can take the case of a slow moving steady state crack, where $C \rightarrow 0$, the heat flux is also approaching zero ($q \rightarrow 0$), and the material such that we can define the internal energy U as the strain energy density W (i.e., the material is elastic), we find that (5.32) reduces to:

$$J = \int_{\Gamma} (W dX_2 - N_{\alpha} S_{\alpha i} F_{i,1} ds) \quad (5.33)$$

Thus, the J-integral is an contour integral, taking a path denoted as Γ , and starts from the lower crack surface and ends up on the upper crack surface. In 2-D, it is a line integral, in 3-D, it is a surface integral, requiring faces along the surface to be evaluated. In terms of a local Cartesian coordinate system set up along the crack tip as before, (5.33) becomes:

$$J = \int_{\Gamma} \left(W dy - \bar{T} \frac{\partial \bar{u}}{\partial x} ds \right) \quad (5.34)$$

where:

W is the strain energy density,

$$W = \int_0^{\varepsilon} \sigma_{ij} d\varepsilon_{ij}^{el} \quad (5.35)$$

$\bar{T}$ is the traction vector on Γ , with outward normal n_j , defined as before:

$$\bar{T}_i = \sigma_{ij}n_j \quad (5.36)$$

$\bar{u}$ is the displacement vector,

ds is the arc length for a line integral, $d\Gamma$ if it is a surface integral,

dy is the incremental length perpendicular to the crack extension path.

The J-integral is path-independent of the contour being used, assuming the same plane is being examined. In plasticity theory, W is not a strain energy density, since most of this energy is dissipated in the material due to plastic flow. As a result, not all of this energy is available for crack extension. However, for J_2 deformation theory (i.e., plasticity theory based on the second deviatoric stress invariant), the J-integral derives from a potential [33, pg. 165]. This can be shown several ways, one being to consider a contour moving with a crack tip. Consider (5.34) where the integral moves with the tip a distance da , the extension distance of the crack along the plane considered:

$$Jda = \int_{\Gamma} W dy da - \bar{T} \frac{\partial \bar{u}}{\partial x} ds da \quad (5.37)$$

where:

$\int_{\Gamma} W dy da$ is the stress work done or strain energy stored during translation of the contour,

$\bar{T} \frac{\partial \bar{u}}{\partial x} ds da$ is the work done by tractions on the contour during motion,

$J da$ is thus the total energy going through the contour for crack extension da .

With this above argument, the J-integral can also be called the energy release rate of a crack. It should be noted that the above derivation is only valid for cracks extending in its own plane, that is, in the x-direction.

Thus, we can say that, for a 2-D case, that per unit thickness of material:

$$J \delta a = -\delta \Pi \quad (5.38)$$

where:

$\delta \Pi$ is the variation in energy of the cracked body for a virtual crack increment δa in its own plane.

For 3-D cases, (5.38) is slightly changed by taking a surface integral instead of a 2-D contour integral. Thus per unit thickness of material:

$$\oint J(s) \delta \mathbf{a}(s) \mathbf{n}(s) ds = -\delta \Pi \quad (5.39)$$

where:

$\delta \mathbf{a}(s) \mathbf{n}(s)$ is the virtual infinitesimal crack extension of length δa , normal to the crack front along this width s , along unit normal $\mathbf{n}(s)$,

$\delta\Pi$ is the variation of total potential energy of the crack body for a virtual infinitesimal crack extension $\delta\mathbf{a}(s)\mathbf{n}(s)$,

s is a discrete width along the crack tip,

$J(s)$ is the surface J-integral evaluated around the crack tip of width s .

For a finite width s , it is said that the J-integral evaluated over this width is said to be an “average” value of J. If this s value is infinitesimal, then the J value is said to be the “local” value of J:

$$\lim_{s \rightarrow 0} \frac{J}{s} = J_{local} \quad (5.40)$$

For purely elastic materials undergoing elastic cracking, J is equivalent to the strain energy release rate, (Equation 5.14), and thus a J value can be related to a stress intensity factor, K.

The preceeding discussion dealt with the J-integral as a vector with only an x-component. That is, we only considered the growth of the crack in the x-direction. In three dimensions, the J-integral is also a vector, but with components in (x, y, z) . [35] claims that for three dimensions, the x, y, and z components of the J-integral, respectively, are:

$$J_x = \int_{\Gamma} \left(W dy - \bar{T} \frac{\partial \bar{u}}{\partial x} ds \right) - \int_S \left(\sigma_{iz} \frac{\partial \bar{u}}{\partial x} \right)_{,z} dS \quad (5.41)$$

$$J_y = \int_{\Gamma} \left(-W dx - \bar{T} \frac{\partial \bar{u}}{\partial y} ds \right) - \int_S \left(\sigma_{iz} \frac{\partial \bar{u}}{\partial x} \right)_{,z} dS \quad (5.42)$$

$$J_z = - \int_S \left(\sigma_{iz} \frac{\partial \bar{u}}{\partial x} \right)_{,z} dS \quad (5.43)$$

where:

S is the surface bounding the volume in question.

Note that for a two-dimensional problem, σ_{iz} is zero. As well, for a plane strain problem, the differential of the $\left(\sigma_{iz} \frac{\partial \bar{u}}{\partial x} \right)_{,z}$ term with respect to z also becomes zero, so it is also zero. Note also that the above equations are for the value along a width of crack, so they represent an “average” value of J .

For simplicity, in our studies, we will assume plane strain conditions, and only consider the x -component of the J -integral, so that equation (5.41) reduces to equation (5.34).

5.3.4 Evaluation of J -integral terms

In a FEM formulation, we can evaluate the strain energy term in the J -integral as thus:

$$\int_{\partial\Omega} W dy = \int_{\partial\Omega} \int_0^\epsilon \sigma_{ij} \cdot d\varepsilon_{ij} dy \quad (5.44)$$

$$= \sum_{k=1}^{GP} \left(\int_0^\varepsilon \sigma_{ij} \cdot \varepsilon_{ij} n_y \right) \det J w_k \quad (5.45)$$

$$= \sum_{k=1}^{GP} \int_0^\varepsilon \begin{pmatrix} \sigma_{xx} \\ \sigma_{yy} \\ \sigma_{zz} \\ \sigma_{xy} \\ \sigma_{yz} \\ \sigma_{xz} \end{pmatrix}^T \cdot \begin{pmatrix} \varepsilon_{xx} \\ \varepsilon_{yy} \\ \varepsilon_{zz} \\ \varepsilon_{xy} \\ \varepsilon_{yz} \\ \varepsilon_{xz} \end{pmatrix} (n_y \det J w_k) \quad (5.46)$$

$$= \sum_{k=1}^{GP} \int_0^\varepsilon \begin{pmatrix} \sigma_{xx} & \sigma_{xy} & \sigma_{xz} \\ \sigma_{yx} & \sigma_{yy} & \sigma_{yz} \\ \sigma_{zx} & \sigma_{zy} & \sigma_{zz} \end{pmatrix} \begin{pmatrix} \varepsilon_{xx} & \varepsilon_{xy} & \varepsilon_{xz} \\ \varepsilon_{yx} & \varepsilon_{yy} & \varepsilon_{yz} \\ \varepsilon_{zx} & \varepsilon_{zy} & \varepsilon_{zz} \end{pmatrix} (n_y \det J w_k) \quad (5.47)$$

Similar to the above FEM evaluation of the strain energy term, we can evaluate the traction and displacement x-gradient term in the J-integral as:

$$\int_{\partial\Omega} \bar{T} \frac{\partial \bar{u}}{\partial x} ds = \int_{\partial\Omega} \sigma_{ij} n_j u_{i,x} ds \quad (5.48)$$

$$= \sum_{k=1}^{GP} \left[\begin{pmatrix} \sigma_{xx} & \sigma_{xy} & \sigma_{xz} \\ \sigma_{yx} & \sigma_{yy} & \sigma_{yz} \\ \sigma_{zx} & \sigma_{zy} & \sigma_{zz} \end{pmatrix} \cdot \begin{pmatrix} n_x \\ n_y \\ n_z \end{pmatrix} \right] \cdot \begin{pmatrix} u_{x,x} \\ u_{y,x} \\ u_{z,x} \end{pmatrix} (\det J w_k) \quad (5.49)$$

where:

w_k are the weights used at each Gauss point for Gaussian integration.

GP are the number of Gauss points used for Gaussian integration.

5.3.5 T^* Integral

A parameter called the T^* integral is defined in [6], which is an extension of the J-integral, is given as:

$$T^* = \int_{\Gamma_{12345}} (W n_1 - t_i u_{i,1}) ds + \int_{V-V_\epsilon} (W_{,1} - (\sigma_{ij} u_{i,1})_{,j}) dv \quad (5.50)$$

where:

W is the total accumulated increments of stress work density.

Figure 5.1: Notation for the T^* integral, from [6, pg. 538].

Here it is shown for the x-direction crack growth, but similar arguments could be made for y-direction crack growth, and z-direction crack growth.

Note the first term in the T^* integral is the J integral itself, with the contour taken around the crack tip, without the contour (1, 0, 5). The main difference between the T^* integral and the J integral is the volume integral term, which takes into account the plasticity effects inside the volume $V - V_\epsilon$. The J-integral, however, neglects this effect, and this is crucial for fatigue, since it models the unloading and reloading effects after reverse plastic deformation.

As a crack grows and reaches steady state growth, one would expect a J value to steadily increase, but the T^* value tends to reach a steady state value. Recall that J is defined for steady state growth, so nothing with the contour changes with time. Another advantage to the T^* integral is that parameters that are valid for the J-integral (such as Equation 5.14) can also be valid for T^* [6].

However, with that said, the J-integral is a very solid starting place for examination of elastic-plastic fracture, as it can be used for elastic problems and plastic problems, with reasonable results. Once a reliable J-integral code is created, it is a matter of modification and testing to create a similarly based T^* code.

Chapter 6

Software Development

6.1 Introduction

To perform a J-integral analysis on the critical solder ball, software had to be developed and tested that would transform a mesh and an accompanying set of results from one arbitrary coordinate system to one where the J-integral could be applied. Recall, that the J-integral requires a particular local coordinate system associated with the crack. Both this rigid body rotation software and the J-integral software had to be developed and tested.

6.2 J-integral Software Testing

6.2.1 Testing Requirements

As with any new software, the need to test and verify the results is crucial. The J-integral software is no different. The J-integral needed to meet several requirements:

1. The J-integral values must be zero (within precision) when a contour encloses no stress singularities.
2. The J-integral must be path independent when the contour is free of gross plastic deformation. That is, the values given by the software should be the same for each different contour enclosing the same singularity, within a precision.
3. The values returned by the J-integral software must be verified by analytic solutions to determine if they are valid or not.

6.2.2 Testing Methods

Two distinct types of tests were conducted. The first was to test for zero values of the J-integral in the absence of singularities. This could use a mesh which contained a constant stress and strain state. To do this, a mesh of a unit cube, $1m \times 1m \times 1m$ was created, consisting of eight 20-node bricks. The loading was such that the boundary nodes were set to be Dirichlet, and a function of their nodal coordinates. This way,

there would be no singularities in the mesh upon loading. The boundary Dirichlet were prescribed with a polynomial such that:

$$\bar{u} = a_0 + a_1x + a_2y + a_3z \quad (6.1)$$

$$\bar{v} = b_0 + b_1x + b_2y + b_3z \quad (6.2)$$

$$\bar{w} = c_0 + c_1x + c_2y + c_3z \quad (6.3)$$

where:

$(\bar{u}, \bar{v}, \bar{w})$ are the x, y, and z nodal displacement components, respectively,

$a_0, a_1, \dots, c_2, c_3$ are coefficients prescribed by the user in a design file.

The cube was made of aluminum alloy 380 (Al380), with the properties:

Table 6.1: Aluminum 380 Material Properties

E	$6.9 \times 10^{10} Pa$
ν	0.33

where:

E is Young's modulus,

ν is Poisson's ratio.

Figure 6.1: Infinite plate with collinear cracks (from [5, pg. 82]).

The second test combined the test for path independence with verification of the J-integral values. If the test problem was kept as purely elastic, then the following relation could be employed to determine if the J-integral values were valid (assuming negligible K_{II} and K_{III} modes of loading, and plane strain conditions):

$$J = \frac{K_I^2(1 - \nu^2)}{E} \quad (6.4)$$

It follows that:

$$K_I = \sqrt{\frac{JE}{(1 - \nu^2)}} \quad (6.5)$$

Like the cube, the plate was also made from aluminum alloy 380 (see Table 6.1).

A formula was found in [5] for an infinite plate with collinear cracks undergoing a constant tensile stress as:

$$K_I = \sigma \sqrt{\pi a} \left(\frac{W}{\pi a} \tan \frac{\pi a}{W} \right)^{\frac{1}{2}} \quad (6.6)$$

One can simulate the conditions in Figure 6.1 by creating a mesh with zero normal Dirichlet boundary conditions (DBC) on faces containing edges EF and CD, simulating material surrounding in an infinite plate. To enforce a plane strain condition in the plate, a zero normal DBC on the front and back faces of the plate was also employed. As well, with these conditions, it is only necessary to mesh half the width, W , so that one gets a mesh of an edge crack. The DBCs enforce the mesh to act as if it is in a larger plate by restricting bending.

One restriction of the code was that it was a displacement based loading, meaning that DBCs were applied to the ends of the plates to induce a stress. In other words, the displacement could be specified, but not the loading. To resolve this, a very long plate (with respect to the width) was created. Tests were run with shorter plates, but the stress as a result of the displacement was not constant along the loading ends. With the elongated plate, it was found that the stress normal to the loaded end was constant across the entire surface. As well, the relatively long plate only had a marginally longer computational time than the shorter plates.

The width of the plate was 1.0m (x-direction), the thickness being 1.5m (z-direction), with a total plate height of 20m (y-direction, see Figure 6.2). The thickness was meshed using three rows of elements, so that each element had a thickness of 0.5

Figure 6.2: The “infinite” plate mesh on the left, with embedded refined cracked mesh on the right.

meters. Essentially a two-dimensional quad was made and extruded along the z-axis at five points specified in space, making three elements in the z-direction. The mesh was made recursively in the respect that first the large coarse mesh was made, and was recursively refined twice around the predetermined crack tip to get the embedded refined mesh seen in Figure 6.2.

If a mesh is incompatible, meaning that between elements there are “unmatched” nodes, a master-slave relationship was to be made between the meshes in question. This is the case between the recursively refined areas in the mesh in Figure 6.2 (for example, the mesh containing the purple and blue parts). That means that nodes on the boundary of the slave mesh (typically the finer mesh) were set to have the same displacements as a master element set, so that there would be no separation between these meshes.

The crack was modeled for three lengths of 8.33cm, 25cm, and 41.67cm, so for each of these lengths a new mesh was made. This was achieved by leaving the area that was to contain a crack with duplicate nodes on adjacent elements, and merging those nodes in which there was to be no crack.

To enforce plane strain conditions, DBCs were enforced on all boundaries, except the loaded ends, so that their normal displacement was zero. On the loaded ends (i.e., the faces at the extreme y-coordinates in Figure 6.2), the DBCs were set at $1 \times 10^{-4}m$ per time step in the y-direction, and zero displacement in the x-direction and z-direction. Thus, the loading was a pure y-displacement. With the test having three time steps, the total displacement at the final time step was $3 \times 10^{-4}m$. As well, 5 different contours were taken around the crack tip, at the plate's center and encompassing the crack tip.

Figure 6.3: Coordinate system for mode I cracking [40, pg. 31].

Another way to verify the results was to examine the purely asymptotic solution for stress fields near a crack in an infinite plate:

$$\sigma_{yy}(r, \theta) = \frac{K_I}{\sqrt{2\pi r}} \cos \frac{\theta}{2} \left(1 - \sin \frac{\theta}{2} \sin \frac{3\theta}{2} \right) \quad (6.7)$$

where:

σ_{yy} is the stress normal to the cracked face,

r is the radius from the crack tip,

θ angle from the x-axis.

If we solve for K_I in (6.7) and take $\theta = 0$, we get:

$$K_I = \sigma_{yy} \sqrt{2\pi r} \quad (6.8)$$

We can plot (6.7) and extrapolate as $r \rightarrow 0$ to get an approximation of K_I .

Literature has been written for the use of special elements near crack tips to accurately capture crack tip effects, such as bricks with mid-nodes moved towards the crack tip in question. However, in these tests, no special elements were used, only regular 20-node bricks were employed.

6.2.3 Testing Results

6.2.3.1 Test Cube

Tests were run for various loading conditions with the test cube described above. Each time, a constant stress state was encountered due to the boundary DBCs prescribed. As well, the J-integral values ranged in order from 1×10^{-9} to 1×10^{-12} , proving that the J-integral did return reasonable values in the absence of singularities under a constant stress state.

6.2.3.2 “Infinite” Plate

Figure 6.4: Embedded refined mesh showing the elements taken in the first (red and green), third (dark blue and yellow), and fifth (smaller purple and light blue mesh) contour for J-integral evaluation, along with half of the finest embedded refined cracked mesh (larger purple and light blue mesh).

Three of the five contours taken are shown in Figure 6.4. All contours were taken concentrically around the crack tip, and evaluated at the center of the plate. They are shown cut away in Figure 6.4 for comparison of sizes. All contours were taken at a constant width of 0.5m. Since each crack length had a geometrically similar

mesh, these values were constant for each crack length evaluated.

There was little difference between J-integral values for each contour. To try to quantify the difference, the average value of the J-integral for the time step was taken and then the percentage difference between the particular contour's value and the average was taken. The majority of values only differed from their average value by between 1% to 3%, with the only "large" difference coming from the J-integral values from the first and smallest contour containing 4 elements. Strangely this only occurred with crack lengths of 8.33cm and 41.67cm, and both times the difference from the average was about 10%. Despite this, the majority of the values were essentially identical, typically within a decimal point, so with this, the path independence was proven.

It was noted that the J-integral value returned by the code was for the thickness of material evaluated (referring to Figure 6.2, along the z-direction). Thus, since each contour was 0.5m in thickness, the value had to be divided by 0.5m to get the true J-integral value per unit width.

Thus, for comparison, the average of the five contours was taken and an equivalent K value computed using equation 6.5. The average values are shown in the following tables as "K (Jint)".

As well, the nodal stresses in the y-direction (σ_{yy}) across the top and bottom y-coordinate edges of the plates were evaluated and shown to be constant, and thus

were used in equation 6.6. These values are shown in the following tables as “K (Periodic BC)”.

Finally, as a final means of comparison, the nodal values of σ_{yy} were evaluated along the crack plane with $\theta = 0$, and substituted into equation 6.8. These values were then plotted (see Appendix) and linearly extrapolated to $r = 0$. Thus, the x-intercept on these plots can be taken as an approximation of K_I . These values are shown as “K (Inf. Plate)” in the following tables:

Table 6.2: J-integral values - Time step 1 (per unit length)

Contour	Crack Length = 8.33cm	Crack Length = 25.0cm	Crack Length = 41.67cm
1	3.02136	9.55586	18.75784
2	3.4109	10.78956	19.01122
3	3.43008	10.85092	19.04126
4	3.45292	10.92848	19.05828
5	3.45796	10.94438	19.06182

Table 6.3: J-integral values - Time step 2 (per unit length)

Contour	Crack Length = 8.33cm	Crack Length = 25.0cm	Crack Length = 41.67cm
1	12.08548	38.2234	75.0314
2	13.64358	43.1582	76.0448
3	13.72034	43.4036	76.165
4	13.81166	43.714	76.2332
5	13.83182	43.7774	76.2472

Table 6.4: J-integral values - Time step 3 (per unit length)

Contour	Crack Length = 8.33cm	Crack Length = 25.0cm	Crack Length = 41.67cm
1	27.1924	86.0028	168.8204
2	30.698	97.106	171.101
3	30.8708	97.6582	171.3712
4	31.0762	98.3562	171.5246
5	31.1216	98.4994	171.5564

Table 6.5: K values - Time step 1 (All K values in $Pa\sqrt{m}$)

Crack Length (cm)	K (Jint)	K (Periodic BC)	K (Inf. Plate)
8.33	5.09×10^5	5.24×10^5	5.48×10^5
25.0	9.06×10^5	9.19×10^5	8.57×10^5
41.67	1.21×10^6	1.22×10^6	1.21×10^6

Table 6.6: K values - Time step 2 (All K values in $Pa\sqrt{m}$)

Crack Length (cm)	K (Jint)	K (Periodic BC)	K (Inf. Plate)
8.33	1.02×10^6	1.05×10^6	1.10×10^6
25.0	1.81×10^6	1.84×10^6	1.71×10^6
41.67	2.42×10^6	2.43×10^6	2.42×10^6

Table 6.7: K values - Time step 3 (All K values in $Pa\sqrt{m}$)

Crack Length (cm)	K (Jint)	K (Periodic BC)	K (Inf. Plate)
8.33	1.53×10^6	1.57×10^6	1.64×10^6
25.0	2.72×10^6	2.76×10^6	2.57×10^6
41.67	3.64×10^6	3.66×10^6	3.63×10^6

6.2.4 Figures for J-integral Testing

Figure 6.5: Close-up view of the through 41.67cm crack with displacement magnified $500\times$ for emphasis.

Figure 6.6: Z-displacement of the plate mesh, showing that the entire width is in plane strain.

Figure 6.7: Asymptotic solution for an infinite plate, (6.8), extrapolated to $r \rightarrow 0$ for 8.33cm crack.

Figure 6.8: Asymptotic solution for an infinite plate, (6.8), extrapolated to $r \rightarrow 0$ for 25.0cm crack.

Figure 6.9: Asymptotic solution for an infinite plate, (6.8), extrapolated to $r \rightarrow 0$ for 41.67cm crack.

Figure 6.10: Comparison of three solutions for 8.33cm crack.

Figure 6.11: Comparison of three solutions for 25.0cm crack.

Figure 6.12: Comparison of three solutions for 41.67cm crack.

6.3 Rigid Body Rotation Software Testing

6.3.1 Motivation

The motivation behind this project is to transform a problem analyzed in one coordinate system into another coordinate system where the transformation is a rigid body motion, i.e., a rotation plus a displacement. In particular to transform the mesh and associated results from a stress analysis in one coordinate system into a set of results and a mesh in another coordinate system which is compatible with a J-integral analysis. Thus, to achieve this, we can take any mesh in a configuration (“A”) and rotate it to a new configuration (“B”), such that we rotate the mesh by:

$$B = \mathbf{R}A + \mathbf{v} \quad (6.9)$$

We can achieve this by taking some rotation matrix $\mathbf{R}$ and some displacement vector $\mathbf{v}$. Then apply the rotation to the vector and second order tensor data in the results, such that:

$$\begin{aligned} \mathbf{u}_B &= \mathbf{R}\mathbf{u}_A \\ \sigma_B &= \mathbf{R}\sigma_A\mathbf{R}^T \end{aligned} \quad (6.10)$$

Conversely, we can rotate results back to configuration A from B by applying:

$$\begin{aligned}
\mathbf{u}_A &= \mathbf{R}^T \mathbf{u}_B \\
\sigma_A &= \mathbf{R}^T \sigma_B \mathbf{R}
\end{aligned} \tag{6.11}$$

6.3.2 Proof of Stress Rotation

Suppose we define a rotation matrix, $\mathbf{R}$ such that:

$$\mathbf{R} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix} \tag{6.12}$$

Thus, the transpose of $\mathbf{R}$ is $\mathbf{R}^T$:

$$\mathbf{R}^T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{bmatrix} \tag{6.13}$$

Choosing $\mathbf{R}$ is such that if $\mathbf{R}$ is composed of three column vectors,

$$\mathbf{R} = \begin{bmatrix} |c_1| & |c_2| & |c_3| \end{bmatrix} \tag{6.14}$$

we can say that $|c_1|$ is a vector in B which is equal to:

$$\left[\begin{array}{c} |c_1| \end{array} \right]_B = \left[\begin{array}{c} 1 \\ 0 \\ 0 \end{array} \right]_A \quad (6.15)$$

We start with the definition of traction, τ in terms of stress σ and outward normal $\mathbf{n}$, in both configurations A and B:

$$\begin{aligned} \sigma_A \cdot \mathbf{n}_A &= \tau_A \\ \sigma_B \cdot \mathbf{n}_B &= \tau_B \end{aligned} \quad (6.16)$$

As above, if we wish to rotate from A to B, we can apply:

$$\begin{aligned} \mathbf{u}_B &= \mathbf{R}\mathbf{u}_A \\ \mathbf{n}_B &= \mathbf{R}\mathbf{n}_A \\ \mathbf{n}_A &= \mathbf{R}^T\mathbf{n}_B \end{aligned} \quad (6.17)$$

Substituting equation (6.17) into equation (6.16), we get:

$$\begin{aligned} \mathbf{R}\sigma_A \cdot \mathbf{n}_A &= \mathbf{R}\tau_A \\ \mathbf{R}\sigma_A \cdot \mathbf{R}^T\mathbf{n}_B &= \mathbf{R}\tau_A \\ &= \tau_B \end{aligned} \quad (6.18)$$

Thus, the stress state in A is related to B by:

$$\sigma_B = \mathbf{R} \sigma_A \cdot \mathbf{R}^T \quad (6.19)$$

$$\sigma_B \cdot \mathbf{n}_B = \tau_B \quad (6.20)$$

6.3.3 Expected Results - A to B rotation

Suppose we define quantities associated with configuration A as σ^A , and quantities associated with configuration B as σ^B . By applying the above rotation matrix and equation 6.10, we get:

$$[\sigma] = \begin{bmatrix} \sigma_{xx}^A & \sigma_{xy}^A & \sigma_{xz}^A \\ \sigma_{yx}^A & \sigma_{yy}^A & \sigma_{yz}^A \\ \sigma_{zx}^A & \sigma_{zy}^A & \sigma_{zz}^A \end{bmatrix}_A = \begin{bmatrix} \sigma_{xx}^A & -\sigma_{xz}^A & \sigma_{xy}^A \\ -\sigma_{xz}^A & \sigma_{zz}^A & -\sigma_{yz}^A \\ -\sigma_{xy}^A & -\sigma_{zy}^A & \sigma_{yy}^A \end{bmatrix}_B \quad (6.21)$$

MrcShell notation has the following order for the vector representation of a symmetric second order tensor:

$$[\sigma] = \begin{bmatrix} \sigma_{xx} & \sigma_{yy} & \sigma_{zz} & \sigma_{xy} & \sigma_{yz} & \sigma_{xz} \end{bmatrix} \quad (6.22)$$

Thus, rotation from A to B would show the following change in order of the stress vector:

$$[\sigma] = \begin{bmatrix} \sigma_{xx}^A & \sigma_{yy}^A & \sigma_{zz}^A & \sigma_{xy}^A & \sigma_{yz}^A & \sigma_{xz}^A \end{bmatrix}_A = \begin{bmatrix} \sigma_{xx}^A & \sigma_{zz}^A & \sigma_{yy}^A & -\sigma_{xz}^A & -\sigma_{yz}^A & \sigma_{xy}^A \end{bmatrix}_B \quad (6.23)$$

For the displacement, we can consider the following transformations, applying equation 6.10:

$$[\mathbf{u}]_B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}_A = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}_B \quad (6.24)$$

$$[\mathbf{u}]_B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}_A = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}_B \quad (6.25)$$

$$[\mathbf{u}]_B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}_A = \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix}_B \quad (6.26)$$

6.3.4 Expected Results - B to A rotation

Following the above discussion, we expect the rotation from B to A to use equation (6.11) and transform the results as such:

$$[\sigma] = \begin{bmatrix} \sigma_{xx}^B & \sigma_{yy}^B & \sigma_{zz}^B & \sigma_{xy}^B & \sigma_{yz}^B & \sigma_{xz}^B \end{bmatrix}_B = \begin{bmatrix} \sigma_{xx}^B & \sigma_{zz}^B & \sigma_{yy}^B & \sigma_{xz}^B & -\sigma_{yz}^B & -\sigma_{xy}^B \end{bmatrix}_A \quad (6.27)$$

For the displacement, we can consider the following transformations, applying equation 6.11:

$$[\mathbf{u}]_A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}_B = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}_A \quad (6.28)$$

$$[\mathbf{u}]_A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}_B = \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix}_A \quad (6.29)$$

$$[\mathbf{u}]_A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}_B = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}_A \quad (6.30)$$

Note that the transpose of the rotation matrix inverts the rotation back to configuration A.

6.3.5 Test Procedure

6.3.5.1 Problems A and B

For simplicity, mesh A is a cube. The centroid of mesh A has coordinates (0.5,0.0,0.5) and the vertices of its bounding box are (0,0.5,0) and (1,-0.5,1).

Mesh B is generated by rotating Mesh A 90 degrees about the +x axis, using equation (6.12). The rotation is counterclockwise (CCW) when viewed from the +x-axis. The result is that a point (x_A, y_A, z_A) is mapped to a point $(x_A, z_A, -y_A) = (x_B, y_B, z_B)$. The bounding box of mesh B is (0,0,0.5) and (1,-1,0.5).

This is done by using the *RBT* method, which is a method of *MrcoUMesher*. The usage is as such:

```
MrcoUMesher <instance> <myDoc> [eType (4,8)]
```

Methods:

```
RBT <o1> <e11> <e12> <e13> <o2> <e21> <e22> <e23>
```

where:

< o1 > is the origin in configuration 1, or A in our case,

< e11 >, < e12 >, < e13 > are the basis vectors for configuration 1, or A in our case,

< o2 > is the origin in configuration 2, or B in our case,

< e21 >, < e22 >, < e23 > are the basis vectors for configuration 2, or B in our case,

Here, the rotation matrix $\mathbf{R}$ is computed by the following equations:

$$[\mathbf{R}] = \begin{bmatrix} \langle e11 \rangle \cdot \langle e21 \rangle & \langle e11 \rangle \cdot \langle e22 \rangle & \langle e11 \rangle \cdot \langle e23 \rangle \\ \langle e12 \rangle \cdot \langle e21 \rangle & \langle e12 \rangle \cdot \langle e22 \rangle & \langle e12 \rangle \cdot \langle e23 \rangle \\ \langle e13 \rangle \cdot \langle e21 \rangle & \langle e13 \rangle \cdot \langle e22 \rangle & \langle e13 \rangle \cdot \langle e23 \rangle \end{bmatrix} \quad (6.31)$$

It should be noted that if $(\langle e11 \rangle, \langle e12 \rangle, \langle e13 \rangle)$ are set to be the identity matrix, then the $(\langle e21 \rangle, \langle e22 \rangle, \langle e23 \rangle)$ forms the *column* vectors in the rotation matrix R :

$$\mathbf{R} = \begin{bmatrix} \langle e21 \rangle & \langle e22 \rangle & \langle e23 \rangle \end{bmatrix} \quad (6.32)$$

In problem A we apply Dirichlet x-displacements of -1×10^{-4} to the -y face. We also apply zero Dirichlet BC to nodes on +y face (with $(r, s, t) = (0, 1, 0)$) with bitfield (1,1,1).

In problem B we apply Dirichlet x-displacements of -1×10^{-4} to the -z face. This is the exact same set of nodes that have non-zero Dirichlet applied to them in problem A. We also apply zero Dirichlet BC to nodes on +z face (with $(r, s, t) = (0, 0, 1)$) with bitfield (1,1,1).

To test the method, we perform an analysis in configuration A. Then we rotate the results to configuration B. We can call this analysis “1”. Conversely, we perform an analysis with identical boundary conditions, as described above, to the previous

analysis in configuration B, and rotate back to A. We can call this analysis “2”.

We rotate these results using the *TransformResultsByRBT* method, which is a method of *MrcoDataUtility*:

```
MrcoDataUtility <instance>  
  
TransformResultsByRBT <array of 1st coordinate sys>  
  
<array of 2nd coordinate sys> <name of resultsIn>  
  
<name of resultsOut>
```

where:

array of 1st coordinate sys takes the form:

$$[\langle o1 \rangle \langle e11 \rangle \langle e12 \rangle \langle e13 \rangle]$$

array of 2nd coordinate sys takes the form:

$$[\langle o2 \rangle \langle e21 \rangle \langle e22 \rangle \langle e23 \rangle]$$

name of resultsIn is the results file which is read,

name of resultsOut is the results file which is written out after rotation.

Again, it should be noted that if ($\langle e11 \rangle$, $\langle e12 \rangle$, $\langle e13 \rangle$) are set to be the identity matrix, then the ($\langle e21 \rangle$, $\langle e22 \rangle$, $\langle e23 \rangle$) forms the *column* vectors in the rotation matrix *R*:

$$\mathbf{R} = \begin{bmatrix} < e21 > & < e22 > & < e23 > \end{bmatrix} \quad (6.33)$$

Thus, we expect that the results of the analysis in “1” in configuration A to have the same results as analysis “2” in configuration A. Likewise, we expect that the results of the analysis in “2” in configuration B to have the same results as analysis “1” in configuration B.

It is key to the test that both mesh and results are rotated identically, and that boundary conditions are identical.

The analysis on Mesh B should have the same results as the analysis on Mesh A rotated into the configuration of Mesh B. The analysis on Mesh A should have the same results as the analysis on Mesh B rotated into the configuration of Mesh A.

6.3.6 Test Results

6.3.6.1 Test Cube

To check that displacements and stresses were mapped properly using the rotation matrix from equation (6.12), the results were both visually verified using Ensign, and data files checked. These data files provided Gauss point stress and displacement gradient values. As well, the results files which contained data such as nodal displacements, Gauss point stresses and nodal stresses, could also be checked.

The first set of tests to be checked were the test cube described above, with displacements described above.

Figure 6.13: Rotation from configuration A (cube on left) to configuration B (cube on right) showing the magnitude of displacement.

Figure 6.15: Rotation from configuration A (cube in front) to configuration B (cube behind) showing Y-displacement.

Figure 6.14: Rotation from configuration A (cube on left) to configuration B (cube on right) showing X-displacement.

Figure 6.16: Rotation from configuration A (cube on left) to configuration B (cube on right) showing Z-displacement.

As figures 6.13 to 6.16 show, the displacements were mapped properly. Checking the data files and results files showed that the expected results from configuration A to configuration B were changed as expected. Recalling the MrcoShell notation for representation of a symmetric second order tensor:

$$[\sigma] = \begin{bmatrix} \sigma_{xx} & \sigma_{yy} & \sigma_{zz} & \sigma_{xy} & \sigma_{yz} & \sigma_{xz} \end{bmatrix} \quad (6.34)$$

Below is the stress state from the data file before rotation for the first Gauss point in element 1:

```
ElementId: 0
GPId: 0
stress: -111560.777457 -3636137.637033 -272571.428811
409700.457349 -209005.467992 48245.121785
```

After rotation, the data file showed the stress had transformed as such:

```
ElementId: 0
GPId: 0
stress: -111560.777457 -272571.428811 -3636137.637033
-48245.121785 209005.467992 409700.457349
```

Comparison of these unrotated and rotated values with equation (6.23) shows

that the stresses were transformed as expected. This was verified visually in Ensign as well.

Next, the transpose of the rotation matrix was applied to configuration B, thus rotating it back to configuration A. These data files were again examined and the unrotated configuration B data file for the first element and first Gauss point are shown below:

```
ElementId: 0
GPIId: 0
stress: -111560.770432 -272571.427867 -3636137.642123
-48245.125067 209005.476027 409700.465574
```

Again, the rotated results from B to A for the first element and first Gauss point are shown below:

```
ElementId: 0
GPIId: 0
stress: -111560.770432 -3636137.642123 -272571.427867
409700.465574 -209005.476027 48245.125067
```

Note that not only does this rotation agree with the expected results in equation (6.27), but the rotated results from B to A agree with the results from the finite element analysis performed in configuration A. The same is true for the results from

configuration B. Again, this was verified visually in Ensight. Thus we can say that the rigid body rotation rotates results as expected given the proper rotation matrix $\mathbf{R}$.

6.3.6.2 Test Plate

The next step was to verify the J-Integral code with rotated results. This was performed using the plate mesh created to originally test the J-Integral, with a crack length of 41.667cm. First, the test was run on the plate in unrotated "natural crack coordinates", so that no rotation was needed to yield a proper J-Integral result. The boundary conditions and timestep length was the same as the original test. This was done so that J-Integral values could be compared to the already verified values. We shall refer to this analysis as "A".

Then the test was performed on a mesh that was rotated using the same rotation matrix $\mathbf{R}$ as above, with identical boundary conditions in the rotated system. It was then rotated back into the "natural crack coordinates" so that a J-Integral analysis could be performed upon it. We shall refer to this analysis as "B". Thus, if the software is correct, both analyses A and B would have identical results.

Below are the numerical results from the analysis, along with the same results from the original testing performed earlier:

Thus, the new J-Integral software works as expected, producing accurate results

Figure 6.17: Original plate mesh in natural crack coordinates.

Figure 6.18: Plate mesh rotated using rotation matrix $\mathbf{R}$

Table 6.8: J-integral values (per unit length)

Contour	Original Values	Test A	Test B
1	18.75784	16.88802	16.88802
2	19.01122	19.0625	19.0625
3	19.04126	19.17108	19.17108
4	19.05828	19.31202	19.31202
5	19.06182	19.34021	19.34021

that are essentially path independent. One may notice a slight difference in values from the original J-Integral testing performed earlier. This may be attributed to precision issues from test to test, but the differences are so slight, they are negligible. As well, the first contour seems to have smaller values than the other values. This may be because since the first contour only encompassed four elements, the elements may not have given sufficiently accurate stresses to capture the singularity at the crack tip accurately.

Chapter 7

Thermal, Stress, and Cracking

Analysis of Dos 100 ASIC

7.1 Special FEM Methods

7.1.1 A General Multigrid Method

A simple approach to getting a more accurate solution in finite element methods is to simply use a finer mesh. However, as the number of elements increases, so does the computational time. Thus, as with many engineering practices, it is a compromise between accuracy and time. One way to reduce computational time is to use what is known as a multigrid method. As in [4], it is best to demonstrate the method in one-dimension for clarity.

If we consider a domain $\Omega = (x : 0 \leq x \leq 1)$ divided into N subintervals, with nodes $x_j = jh$. Here, $h = \frac{1}{N}$ is a constant width subinterval. With these subdivisions we denote the domain as Ω^h . The system of linear equations derived from this grid and put into matrix form is:

$$\mathbf{A}\mathbf{u} = \mathbf{f} \quad (7.1)$$

$$\mathbf{e} = \mathbf{u} - \mathbf{v} \quad (7.2)$$

$$\mathbf{r} = \mathbf{A}\mathbf{u} - \mathbf{f} \quad (7.3)$$

where:

$\mathbf{A}$ the stiffness matrix of the discrete system,

$\mathbf{u}$ the exact solution vector to the continuous system,

$\mathbf{v}$ the approximate solution vector to the discrete system,

$\mathbf{e}$ the error vector,

$\mathbf{r}$ the residual vector,

$\mathbf{f}$ the load vector,

We can rewrite the above as:

$$\mathbf{A}\mathbf{v} = \mathbf{f} - \mathbf{r} \quad (7.4)$$

$$\mathbf{A}\mathbf{e} = \mathbf{r} = \mathbf{f} - \mathbf{A}\mathbf{v} \quad (7.5)$$

(7.5) is known as the residual equation, and says that the error satisfies the same set of equations as the unknown $\mathbf{u}$ when $\mathbf{f}$ is replaced by $\mathbf{r}$ [4]. For a complex function, a finer grid with smaller h value will give a more accurate solution.

A way to improve iteration speed is to use an initial guess on the problem. A way to get this initial guess is to perform an analysis on a coarse mesh and use these results on a finer mesh. Noting from [4] that iterations on the equation (7.1) with an arbitrary initial guess $\mathbf{v}$ is exactly the same as iterating on (7.5) setting $\mathbf{e} = 0$.

To “transfer” data from one grid to another is through prolongation. Prolongation is essentially an interpolation of data between grids, where data from a coarse grid is used as an initial guess on finer grids via interpolation of that data. Denoting fine grid properties by superscript Ω^h and coarse grid properties by Ω^{2h} , a linear prolongation operator I_{2h}^h is used to interpolate the approximate solution such that:

$$I_{2h}^h \mathbf{v}^{2h} = \mathbf{v}^h \quad (7.6)$$

With this, we can define a simple “coarse grid correction scheme” that illustrates the multigrid concept:

Step 1 Iterate on Ω^h using initial guess $\mathbf{v}^h$:

$$\mathbf{A}^h \mathbf{v}^h = \mathbf{f}^h \quad (7.7)$$

Step 2 Compute:

$$\mathbf{r}^{2h} = I_{2h}^h (\mathbf{f}^h - \mathbf{A}^h \mathbf{v}^h) \quad (7.8)$$

Step 3 Solve on Ω^{2h} :

$$\mathbf{A}^{2h} \mathbf{e}^{2h} = \mathbf{r}^{2h} \quad (7.9)$$

Step 4 Correct solution on Ω^h :

$$\mathbf{v}^h \leftarrow \mathbf{v}^h + I_{2h}^h \mathbf{e}^{2h} \quad (7.10)$$

Note the $\leftarrow$ denotes an overwriting of values.

Step 5 Repeat **Step 1** using new $\mathbf{v}^h$ value.

7.2 Description of Domain and Problem

The domain chosen for the analysis here is that of an application specific integrated circuit (ASIC), to be analyzed.

In the FEM analysis, each part is distinguished by its own part type, with a corresponding material associated with it. Each of these types may have several

different sections to it, and each section given its own number. This can be used to analyze different sections of one particular part in detail.

For example, using a shoe as an example, one can set the upper material part type as “upper”, with material “leather”. Suppose there are 4 pieces of leather that compose this upper part of the shoe. Then one would assign 4 different part numbers to these leather “upper” pieces, say 1, 2, 3, and 4. The sole may have a part type “sole”, and is made of “rubber”. If this is a one piece sole, then it may only require one part number, say “5”.

An important feature of this system is that each part number must be unique, but as described above, a set of part numbers may belong to one particular part type.

Figure 7.1: The ASIC mesh used in this analysis.

Figure 7.3: Detail of the underlying layers underneath the heat sink and pedestal.

Figure 7.2: Detail of the pedestal underneath the heat sink.

Figure 7.4: Detail of the solder ball layer in the coarse ASIC mesh.

Figure 7.5: ASIC with part numbers labeled.

Referring to Figure 7.5, one can see the several layers of the ASIC. Note from Figure 7.1 that since the ASIC and the load is symmetrical (true since the flux boundary condition is constant across the silicon die and all material properties are isotropic), only a quarter of it was meshed, to save in computational expense.

Note the solder balls (part number 5) just above the PCB. The solder balls perform an electrical and mechanical connection from the ASIC to the PCB, and are arranged in an array of 23 by 23 solder balls in the total ASIC. The solder balls are composed of eutectic Sn-Pb eutectic solder (63Wt%Sn-37Wt%Pb), of approximately 1.27 mm pitch, 500 μm height and 700 μm diameter at its mid-height. The main

Table 7.1: ASIC part numbers and part types

Part Number	Part Type	Material Type
1	PCB	FR4
2	PCB	FR4
3	Substrate	PTFE
4	Substrate	PTFE
5	Solderballs	Solder
6	Heat Spreader	HeatSpreader
7	SiDie Adhesive	Adhesive
8	Stiffener Adhesive	Adhesive
9	Stiffener	Stiffener
10	Pyralux	Pyralux
11	SiDie	SiDie
12	FakeSB (Underfill)	Underfill
13	Copper-Invar-Copper (CIC)	CIC
14	Overfill	Overfill
15	Pedestal	Pedestal
16	Heatsink	Aluminum 380
18	Heatsink	Aluminum 380

focus of the analysis was to try to estimate the low cycle fatigue life of the solder balls in the ASIC, so particular attention was focused upon the solder balls, and the mating components, the PCB and the substrate. As well, note the location of the silicon die (part number 11), shown as the purple square a few layers above the solder balls. This was the heat source in the ASIC.

7.3 Thermal Analysis

7.3.1 Procedure and Results

There was a flux boundary condition between the silicon die and the layer underneath it (the underfill) to represent the power generation from the silicon die. As well, a convection boundary condition was placed on the heat sink, using data from previous analyses to give an appropriate level of convection from it.

The silicon die acts as the heat source in the ASIC, generating heat during operation. This heat flows throughout the ASIC, with a temperature distribution shown in Figure 7.6. In this analysis, the maximum temperature achieved is approximately 360K, or about 87 degrees Celsius. The relaxation time, or time reach a steady state temperature distribution, is of the order of a few minutes for a package such as the DOS 100.

The temperature distribution shown in Figure 7.7, was used in the subsequent

Figure 7.6: The temperature distribution in the ASIC after 90 seconds of heating.

stress analysis. This temperature as a function time was selected because it matches the temperature profile used in accelerated tests that were planned but regrettably were cancelled. These temperature profiles were computed by running an implicit transient energy solver for one time step of 900 seconds for each the four stages of heating, hot soak, cooling and cold soak. At any time inside a time step, the temperatures were linearly interpolated with respect to time, to produce the profile shown in Figure 7.7. Note that in an accelerated test within a furnace, there are no thermal gradients within the package, whereas here we have spatial thermal gradients due to the heating of the silicon die. Also note that typical thermal relaxation times in actual ASICs are of the order of a few minutes, rather than the time used here in this test.

7.4 ASIC Stress Analysis

7.4.1 Procedure

Figure 7.7: The temperature cycle used for the stress analysis (actual temperatures shown are for the maximum temperatures in the solder balls). This is similar to a prescribed furnace temperature pattern for an accelerated test. This distribution was obtained at the Gauss point within the solder ball with the maximum effective plastic strain amplitude.

The stress analysis of the ASIC was performed using the multigrid technique described earlier. The first step was to perform an analysis on the entire ASIC (see Figure 7.1) using progressively finer meshes to get more accurate solutions. A total of four meshes were used for the entire ASIC.

The boundary conditions for the ASIC was that of a symmetry condition along the edges of the PCB, such that there was no normal displacement along these edges

of symmetry. This was done by setting Dirichlet boundary conditions (DBC) normal to these edges of zero. As well, the heat sink was not included in this analysis, since it was not the focus of the analysis. Instead another DBC of zero displacement was placed on the layer directly underneath the heat sink. This was valid since this layer was quite soft in reality, and thus, transfers very little force to the heat sink.

Figure 7.8: Side view of the package after one temperature cycle. Note warping in the package (magnified 500 times for emphasis).

7.4.2 Material Properties

Of particular interest here are the mechanical properties of the solder used, and the layers they are attached to, the PCB and substrate. Note that the solder modeled here is that of a Sn-Pb eutectic solder (63Wt%Sn-37Wt%Pb).

In our analysis, with the range of temperatures achieved (between 293K and about 346K in the critical solder ball), the properties below were kept constant:

Table 7.2: Solder Material Properties

Property	Value
Coeff. of thermal expansion (CTE)	$24.5 \times 10^{-6} \text{ 1/K}$
Poisson Ratio	0.4

Those properties which were not constant are shown in graphical form. Note that the CTE is an order of magnitude different between the solder and both the PCB material and substrate material.

Figure 7.9: Viscosity ($Pa \cdot s$) values versus temperature (K) for the solder used in the analysis.

Figure 7.10: Yield stress (Pa) values versus temperature (K) for the solder used in the analysis.

Figure 7.11: Young's modulus (Pa) values versus temperature (K) for the solder used in the analysis.

Figure 7.12: CTE ($1/K$) values versus temperature (K) for the PCB material FR4 used in the analysis.

Figure 7.13: CTE ($1/K$) values versus temperature (K) for the substrate material PTFE used in the analysis.

7.4.3 Results

The stresses in the ASIC arise from the temperature distribution induced by the heat from the silicon die. The distribution of heat in the ASIC causes the components to heat up and since they have differing coefficients of thermal expansion (CTE), stresses build up between layers and warping occurs in the package.

This stress analysis for one load cycle was taken over 40 timesteps, with each timestep having a length of 90 seconds, for a total analysis time of 3600 seconds. The temperature distribution was modeled after a prescribed furnace temperature cycle, and was taken from the thermal analysis at each timestep. Recall that the thermal analysis was run for one time step of 900 seconds, so this temperature was prolonged so that it was valid for the timestep value used in the stress analysis (see Figure 7.7). For example, at the first timestep in the stress analysis, the data from the thermal analysis was prolonged to 90 seconds, and at the second timestep it was prolonged to 180 seconds, and so on. This was to simulate testing temperature cycling.

From Figure 7.7 we can see that the temperature is linearly increased from the initial temperature of 293K to steady state at 900 seconds, held there for 900 seconds, then linearly decreased to room temperature of 293K. Note that the plot shown (Figure 7.7) is taken from solder ball data containing the solder ball with the highest temperatures.

Once the stress analysis is run, it is desired to determine which solder ball is predicted to fail first. This is achieved using a modified Coffin-Manson equation relating the amplitude of the plastic strain in each load cycle to the number of cycles to crack initiation. This assumes that the solder balls will fail by low cycle fatigue, which is typically the case.

[10] suggests for “stiff leadless”, a Weibull exponent of $1/4$. Probabilities used were 0.01%, 0.001%, and 0.0001%, translating to probabilities of one in 10000 components failing, one in 100000 components failing, and one in 1 million components failing, respectively.

Figure 7.14: Failure distribution predicted by the Coffin-Manson equation. Note red shows lowest CTF (most likely to fail), blue shows highest CTF (least likely to fail).

Using equation (1.6), the Gauss point containing the largest variation of effective plastic strain was sampled in each solder ball in the array, and then applied to the equation. The results are shown in Figure 7.14. The legend shows the solder balls

Table 7.3: Coffin Manson CTF Results (Coarse Mesh)

Probability	CTF
0.01%	114616
0.001%	65545
0.0001%	36244

with the highest probability for failure in red, and the lowest probability for failure in blue. Results are shown for varying probabilities in Table 7.4. Each time the failure distribution for the solder balls was the same, but the CTF for each individual solder ball would vary depending on the probability chosen.

From this, it was determined that the solder ball directly underneath the silicon die and lying along the package's diagonal line of symmetry (in Figure 7.14 this is the solder ball fifth from the left and fifth up from the package center) would be the first to initiate a crack. Once this was determined it was desired to find a way to characterize cracking in this solder ball, since it was deemed to be most likely to crack first. Since there is substantial plasticity in the solder ball at various points in the thermal cycle, it invalidated the use of linear elastic fracture mechanics. However, it was proposed to attempt to use the J-integral in such a case where there is plastic deformation.

Figure 7.15: Domain chosen for analysis of critical solder ball. Shown is the PCB, solder ball array, and substrate from the entire ASIC (shown in purple, light blue and yellow, respectively), with the critical solder ball mesh location.

Figure 7.16: Refined solder ball mesh.

7.5 Solder Ball Stress Analysis

7.5.1 Procedure

Figure 7.17: Critical solder ball showing location of maximum effective plastic strain (location is furthest from package center, shown in red, at solder ball and PCB interface). Displacements magnified 500 times for emphasis.

The domain of the subsequent analysis is changed from the entire ASIC to just the solder ball in question (see Figure 7.15). This solder ball was taken out of the larger domain, along sections of the attached PCB and substrate. Dirichlet boundary conditions (DBC) were applied to the edges of the PCB (shown in red in Figure 7.15) and substrate (shown in green in Figure 7.15). The DBCs applied were from that of the previous analysis of the entire ASIC, so that these DBCs essentially deformed the solder ball as if it was still in the ASIC package.

The multigrid method was again used to refine the solder ball (see Figure 7.16)

and determine the location of the maximum effective plastic strain (see Figure 7.17), which is the location of the initial crack predicted by the Coffin-Manson equation. It was found that the location of the maximum effective plastic strain was at a point close to the PCB and solder ball interface, furthest from the package center. One can see that the CTE mismatch is greater between the PCB and solder versus the substrate and solder (see Figures 7.12 7.13). Thus, one would expect a greater shearing action at the PCB/solder interface.

Table 7.4: Coffin Manson CTF Results (Coarse Mesh)

Probability	CTF
0.01%	3475
0.001%	6661
0.0001%	11845

Since the mesh was much finer in this analysis than the proceeding coarse ASIC analysis, the stresses were more accurately modeled around the PCB and solder ball interface, and the resulting peak stresses, and thus amplitude of plastic strain, was higher. The Coffin-Manson equation was again applied to this refined solder ball, with the Weibull exponent of $1/4$ again applied. Results are shown in Table .

7.5.2 Results

Figure 7.18: Effective plastic strain plot at each timestep for the finest refined solderball. This plot was obtained at the Gauss point within the solder ball with the maximum effective plastic strain amplitude.

Referring to Figure 7.20 and Figure 7.7, we can examine what is happening within the solder ball during the loading cycle. One can see the stress increase slowly until time step 7 (630 seconds). Up until this point, the rate of stress increase is because of the thermal strain rate is greater than the rate of stress decrease because of relaxation. At the maximum, the two are equal and balance. After the maximum, the rate of stress relaxation is greater than the rate of stress generation by the thermal strain rate [50].

At time step 10 (900 seconds), the maximum temperature is achieved and the

Figure 7.19: Effective strain plot at each timestep for the finest refined solderball. This plot was obtained at the Gauss point within the solder ball with the maximum effective plastic strain amplitude.

Figure 7.20: Effective stress plot at each timestep for the finest refined solderball. This plot was obtained at the Gauss point within the solder ball with the maximum effective plastic strain amplitude.

solder ball is held at this temperature for the next 900 seconds. One would expect the stress to decrease, since viscous effects are included, however, we see the stress increase with time. This may be due to several reasons, the most probable being load transfer among the solder balls in the total package, or continued shear interaction between the solder and substrate and PCB. As the temperature is kept constant there may be continued warping in the package, thus causing an increase in stress during this step.

Once the high temperature dwell is over (time step 20, 1800 seconds), the ASIC is cooled, and the solder viscosity increases. As this occurs, the package begins to warp more severely, and thus there is a large increase in stresses as the solder ball is loaded as it hardens while cooling. At time step 30 (2700 seconds), the cold soak begins, and the stress in the solder ball relaxes slowly. We can see this creep effect as the effective strain plot (Figure 7.19) and effective plastic strain plot (Figure 7.18) both show a slow increase in strain during this time.

7.6 Solder Ball Cracking Analysis

7.6.1 Procedure

To model a crack within the solder ball, contact elements were used and the solder ball was split into three part numbers, 5, 6, and 7. An advantage of this is that there

Figure 7.21: The solder ball mesh with contact elements between part numbers 6 (yellow) and 7 (purple).

Figure 7.22: Photo of an actual solder ball with cracking near the substrate and solder ball interface [51].

is no need to remesh each different crack length, one can use the same mesh with a layer of contact elements, and specify varying stiffnesses for cracked versus uncracked segments.

Referring to Figure 7.21 part number 5 is blue, part number 7 is purple, and part number 6 is yellow. The layer close to the PCB was set as part number 6, and the layer sandwiched between this layer in the bulk solder was set as 7 (see Figure 7.21). To model the crack, a special contact stiffness formula was used to distinguish between uncracked solder and cracked solder, where the stiffness is a function of the gap width between the contact faces:

$$K(gap) = K_{max} \left(1 - \frac{gap - offset}{sharpness + |gap - offset|} \right) \quad (7.11)$$

where:

K_{max} is the maximum value of contact stiffness, specified in the design file,

gap is the gap between faces of the contact element,

$sharpness$ a factor that makes $K(gap)$ a step function when set to zero,

$offset$ $K(gap)$ is symmetric about this value of offset.

Examination of the above equation shows that varying the parameters can greatly influence the traction between the two contact faces for a given gap or displacement. The contact elements with this type of stiffness formulation act as a spring,

where stiffness in three directions with respect to the contact face in question can be specified, one normal direction, and two tangential directions. The formulation was the same in all three directions, and the contact force between the faces:

$$Force(gap) = K_{gap} * gap \quad (7.12)$$

Table 7.5: Contact stiffness parameter values

	Cracked	Uncracked
$K(max)$	2×10^9	2×10^{14}
$offset$	1×10^{-9}	2×10^{-4}
$sharpness$	1×10^{-9}	5×10^{-5}

Figure 7.23: Specified stiffness function for uncracked portion of solderball with contact elements (see Table 7.5 for uncracked parameter values).

Geometry was used to specify the uncracked versus cracked contact faces. Not-

Figure 7.24: Resulting force between contact elements with uncracked stiffness formula specified (see Table 7.5 for uncracked parameter values).

Figure 7.25: Specified stiffness function for cracked portion of solderball with contact elements (see Table 7.5 for cracked parameter values).

Figure 7.26: Resulting force between contact elements with cracked stiffness formula specified (see Table 7.5 for cracked parameter values).

ing the solder ball arrangement in Figure 7.14, we can see that the critical solder ball in question is at equal x and y increments from the origin of the global coordinate system. As well, noting that the point of maximum effective plastic strain is at the point furthest from the package center in this solder ball (i.e., the origin), this is where we will take our crack to start. The crack in question was modeled to be a straight through crack, which was at a constant z -coordinate (i.e., in the xy -plane only) for simplicity. This would agree with experiments for the solder ball setup considered here (see Figure 7.22).

Consider the setup shown in Figure 7.27, where the solder ball is the circle in question. In the setup considered here the x and y intercepts are equal, so $\theta = 45$

Figure 7.27: Geometry for contact cracking.

degrees.

If we take the line with equation:

$$y = y_{int} - x \quad (7.13)$$

$$= x_{int} - x \quad (7.14)$$

to be the line at the crack tip, then we can calculate a nominal crack length with relation to the point of maximum effective plastic strain (EPS). With reference to Figure 7.27, we can call the point of maximum EPS, R_{max} . This is the radius from the origin of the global coordinate system to the point of maximum EPS. We then take the point where the crack tip line and the radius from the origin of the global coordinate system intersect, and call this R_{min} . Then, we can define a nominal crack length as:

$$a_o = R_{max} - R_{min} \quad (7.15)$$

Knowing R_{max} , we can specify a_o to get a desired R_{min} . Note also that since $\theta = 45$ degrees:

$$x_{int} = y_{int} \quad (7.16)$$

$$= \sqrt{R_{min}^2 + R_{min}^2} \quad (7.17)$$

$$= \sqrt{2R_{min}^2} \quad (7.18)$$

$$= \sqrt{2(R_{max} - a_o)^2} \quad (7.19)$$

Substituting back into our crack tip line equation we have:

$$y = \sqrt{2(R_{max} - a_o)^2} - x \quad (7.20)$$

We can use equation (7.20) as a test in a logical expression to see if a face is to have a cracked or uncracked stiffness formulation. Each face has a centroid specifying its location in space. We can sample each contact face's y-coordinate and x-coordinate. We can then substitute that x-coordinate into equation (7.20), and if the resulting y-coordinate is greater than the actual y-coordinate, then we specify the cracked stiffness formula for that face. Otherwise, all other faces remain uncracked.

This is illustrated in Figure 7.28. The red faces indicate a cracked condition, where the yellow ones remain uncracked, since they are underneath the "Crack Line".

The Dirichlet boundary conditions for these solder ball contact cracking meshes were taken from the results from the previous stress analysis on the regular refined solder ball mesh and applied to the outer periphery of the PCB and substrate. This simulated being part of the ASIC by having displacements determined by the surrounding package. Note that with this loading, there are no other cracks assumed in the other solder balls, and that the displacements from heating the ASIC essentially remains constant throughout the propagation of the crack.

Figure 7.28: Cracking surface on part number 6.

With this, analyses were run for crack lengths of $20\mu m$, $50\mu m$, $100\mu m$, $200\mu m$. Since the contact solder ball mesh itself does not lend itself to perform a J-integral analysis on it easily, we chose to create so-called crack objects within the solder ball to prolongate displacement results from these contact crack solder ball to the crack objects. Once the displacements are specified on the boundary of the crack object, the stresses and strains can be solved everywhere within it.

This was done by creating crack object meshes that lie completely within the solder ball, so Dirichlet boundary conditions can be applied to the boundary with values from the displacements of the contact cracking solder ball analysis. The crack object meshes were simple quads extruded along the above “crack line” (see 7.28) and refined to be 20 node bricks for the four different cracklengths in question. The meshing code removed any elements with nodes outside the solder ball periphery, so

that the mesh would be completely contained within the solder ball.

Figure 7.29: Closeup of the $20\mu m$ contact crack created in the contact solder ball mesh.

Figure 7.31: Closeup of the $50\mu m$ contact crack created in the contact solder ball mesh.

Figure 7.30: Closeup of the $100\mu m$ contact crack created in the contact solder ball mesh.

Figure 7.32: Closeup of the $200\mu m$ contact crack created in the contact solder ball mesh.

Figure 7.33: Half of contact solder ball with full $20\mu m$ crack object embedded within it, showing orientation.

Figure 7.35: Half of contact solder ball with full $50\mu m$ crack object embedded within it, showing orientation.

Figure 7.34: Half of contact solder ball with full $100\mu m$ crack object embedded within it, showing orientation.

Figure 7.36: Half of contact solder ball with full $200\mu m$ crack object embedded within it, showing orientation.

These meshes were used to analyze the crack tip phenomena at the four discrete lengths by prolonging the displacement data from the larger solder ball meshes to the boundary of the crack objects. Once this was done, the crack objects and the associated stress analysis results had to be rotated into a coordinate system that the J-integral software could use. To accomplish this, the following rotation matrix was applied (noting the orientation of θ in Figure 7.27):

$$\mathbf{R} = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ 0 & 0 & -1 \\ \sin\theta & \cos\theta & 0 \end{bmatrix} \quad (7.21)$$

This rotation matrix is orthogonal, and with a determinant of 1, both strict requirements for a proper rotation matrix. Thus, with $\theta = 135^\circ$ the crack object could be rotated so that it was aligned in the proper J-integral coordinate system, using the rigid body rotation code discussed earlier.

Figure 7.37: $20\mu m$ crack object mesh, in unrotated coordinates (note coordinate axes).

Figure 7.39: $20\mu m$ crack object mesh, in unrotated coordinates (note coordinate axes).

Figure 7.38: $20\mu m$ crack object mesh, in rotated coordinates (note coordinate axes).

Figure 7.40: $20\mu m$ crack object mesh, in rotated coordinates (note coordinate axes).

7.6.2 Results

Obtaining J-integral values was more challenging than anticipated. This was due to negative values being present in contours taken far from the crack tip, which would indicate that there would be no energy present for crack extension, but logically with high stresses and plasticity present, this is not the case. However, if we took contours very close (i.e., within a few elements) of the crack tip we could get reasonable values. With each crack length, at least 2 contours were taken, and the average between the values taken and deemed to be the J-integral value for that timestep and crack length. Any negative values were disregarded. One should note that path independence does tend to break down upon the presence of large amounts of plasticity. With that said, once the data was collected, Excel was used to determine an equivalent stress intensity factor, K , using:

$$J = \frac{K_I^2(1 - \nu^2)}{E} + \frac{K_{II}^2(1 - \nu^2)}{E} \quad (7.22)$$

where:

E is Young's modulus,

ν is Poisson's ratio,

We can combine the K_I and K_{II} factors into one equivalent stress intensity factor to get an approximation of the total stress intensity at the crack tip:

$$J = \frac{K_{eq}^2(1 - \nu^2)}{E} \quad (7.23)$$

$$K_{eq} = \sqrt{\frac{JE}{(1 - \nu^2)}} \quad (7.24)$$

$$K_{eq} = K_I + K_{II} \quad (7.25)$$

Here, we assume plane strain, and little antiplane (K_{III}) shear cracking, which is valid since we are taking samples along a line of symmetry in the ASIC. Although strictly true for linear elastic cases only, this will provide an approximation for fatigue crack growth using the Paris-Erdogan equation (equation 5.11):

[27] gives values for the material constants mentioned above for 63Sn-37Pb solder, and are shown below. For this study, since two values are given for identical loading and frequency conditions, the average of the two were taken:

Table 7.6: Paris-Erdogan constants for solder from [27]

Constant	Test 1	Test 2	Average
c	2.07×10^{-7}	3.46×10^{-7}	2.77×10^{-7}
m	3.30	3.21	3.26

Note that the constants in Table 7.6 are for a da/dN in units of $cm/cycle$, and ΔK in units of $MPa \cdot \sqrt{m}$. Although this is strictly true when the problem is in the elastic regime only, it should give an approximation of the fatigue life of the solder ball. As well, J itself is not a fatigue parameter, as it is valid for steady state cracking

only, but what we are striving for is an approximation for fatigue crack growth. With these considerations in mind, the data was taken from the mesh and an equivalent K value plotted for each crack length. The minimum and maximum values for each crack length were taken, and this provided an approximation of the range of stress intensity factors for a cycle, ΔK . We can re-arrange the Paris-Erdogan equation to get the number of cycles as:

$$N = \int \frac{da}{c(\Delta K)^m} \quad (7.26)$$

Thus, if we can express ΔK in terms of crack length, we can integrate the above equation. For each crack length, a dimensionless parameter η was calculated, where:

$$\eta = \frac{a}{R} \quad (7.27)$$

where:

a is the crack length,

R solder ball radius at the cracking plane in question ($R \approx 248\mu m$),

Once these ΔK values were plotted, a quadratic line of best fit was used to determine the equation relating η to ΔK . This equation was found to be:

$$\Delta K = 0.5744\eta^2 + 0.7502\eta + 0.5841 \quad (7.28)$$

Table 7.7: η and ΔK values for varying crack lengths

η	Crack Length (μm)	ΔK ($MPa\sqrt{m}$)
0.0807	20	0.765
0.2016	50	0.535
0.4032	100	1.11
0.8065	200	1.54

We can take the starting crack length as $20\mu m$ ($\eta = 0.0807$) and the end crack length at the center of the solder ball ($\eta = 1.0$). Substituting the above equation into our equation for cycles to failure, we get:

$$N = \int_{0.0807}^{1.0} \frac{248 \times 10^{-6} d\eta}{2.77 \times 10^{-7} (0.5744\eta^2 + 0.7502\eta + 0.5841)^{3.26}} \quad (7.29)$$

The commercial software package Maple was used to integrate the above equation. The value calculated was 87189 cycles. However, this is not the cycles to failure one can expect for this frequency and temperature range. The temperature used in finding the Paris Erdogan constants in [27] was 24 degrees Celsius, with a frequency of 10 Hz. To scale this number, [26] suggests:

$$N_{act} = \lambda N \quad (7.30)$$

N_{act} is the “actual” number of cycles to failure for the loading in question,

N number of cycles to failure predicted using constants from [27],

λ a scaling factor:

$$\lambda = \left(\frac{f_{act}}{f_{test}} \right)^{1/3} \left(\frac{MaxT_{test}}{MaxT_{act}} \right)^2 \quad (7.31)$$

f_{act} actual cycling frequency used in analysis, 1 cycle every 3600 seconds,

f_{test} test cycling frequency used for determining fatigue constants, 10 Hz,

$MaxT_{act}$ actual maximum temperature achieved in the analysis, approximately 73 degrees Celsius,

$MaxT_{test}$ testing maximum temperature achieved to determine fatigue constants, 24 degrees Celsius.

With this:

$$\lambda = \left(\frac{1}{36000} \right)^{1/3} \left(\frac{24}{73} \right)^2 = 0.003273 \quad (7.32)$$

Thus, the calculated cycles to failure are 286, using this scaling factor used above.

Figure 7.41: J-integral data plotted as equivalent stress intensity factor K , for each timestep for the $20\mu\text{m}$ crack length (blue dots are data points, black line is a line of best fit).

Figure 7.42: J-integral data plotted as equivalent stress intensity factor K , for each timestep for the $50\mu\text{m}$ crack length (blue dots are data points, black line is a line of best fit).

Figure 7.43: J-integral data plotted as equivalent stress intensity factor K , for each timestep for the $100\mu\text{m}$ crack length (blue dots are data points, black line is a line of best fit).

Figure 7.44: J-integral data plotted as equivalent stress intensity factor K , for each timestep for the $200\mu\text{m}$ crack length (blue dots are data points, black line is a line of best fit).

Figure 7.45: ΔK values for varying normalized crack lengths (blue dots are data points, black line is a line of best fit).

Chapter 8

Conclusions and Discussion

8.1 Discussion of Results

The cycles to failure found in this cracking analysis should be in addition to the cycles to failure predicted by the Coffin-Manson equation. Recall, the Coffin-Manson equation predicts cycles to crack initiation, whereas the Paris-Erdogan equation predicts cycles for crack growth from initiation to a pre-determined “failure” criterion. For example, if we take the CTF value found by the Coffin-Manson equation using a 0.0001% probability of failure (3475 cycles to failure), we can see that total CTF value would be 3475 cycles plus the 286 cycles of crack growth predicted by the fracture analysis. Thus, the total CTF would be approximately 3761 cycles to failure. Thus, the initiation of the crack comprises the majority of the life of the solder connection

(approximately 92% of the life). This implies that in this type of package, crack growth is not as important as crack initiation.

As mentioned earlier, a true fatigue life analysis should be performed only when cyclic steady state is achieved. To save computational time, the first cycle was used in the analysis. This would tend to be overly conservative in our CTF calculations, as the cyclic stress and plastic strains are expected to be greater here than during cyclic steady state.

Figure 8.1: Effective stress at the bottom of the solder ball, where cracking analysis was to take place.

Referring to Figure 8.1, we can see a ring of high stress along the edges of the solder ball at the PCB interface. The presence of these stresses is a violation of the assumptions of the derivation of the J-integral [50]. This may be the cause of the high ΔK values at lower crack lengths versus higher crack lengths (Figure 7.45, $20\mu m$ versus $50\mu m$). The high stresses around the edges here would suggest

Figure 8.2: Effective stress at the bottom of the solder ball, with a contact crack length of $200\mu m$.

that a crack would go around the edges before going through the bulk material. The relatively higher crack growth rate at low crack lengths suggests that at low crack lengths, the stresses on the lower perimeter of the solder ball influence the growth rate significantly. However, as the crack grows, these perimeter stresses in the cracked ligament become less (see Figure 8.2), and have less influence on the crack growth rate since the crack is moving away from it.

As well, examination of Figures 7.29 and 7.31 show that the contact solder ball crack at a crack length of $20\mu m$ is shaped more like an thumbnail-type edge crack rather than the more through crack at $50\mu m$. This is unavoidable using the contact stiffness formulation and the mesh used in the solder ball analysis. Since the crack object is assumed to be a straight through crack, and it takes these far field parameters from this contact solder ball mesh and its accompanying results, this

would cause inaccurate results.

Once the analysis goes past this initial stage, the J-integral seems to increase as one would expect with increasing crack length. We must note that the analysis done here is for the J-integral component in the x-direction, so it does not capture variations in crack growth in the y-direction or z-direction.

Also note from Figure 8.2 that a second area of high stress occurs in the solder ball, at the point closest to the package center. This would indicate the possibility that a secondary crack may initiate there, in addition to the crack already propagated.

Both the Coffin-Manson based CTF results and the fracture mechanics CTF results have pros and cons. The main disadvantage to the Coffin Manson CTF approach is that one must have a good idea of what the Weibull exponent is for that particular package, otherwise there can be great differences in the predicted CTF (see Figure 8.3). This can only be achieved through experimentation for that specific package, which for this package, has not been performed at the writing of this thesis. However based on results from similar ball grid array (BGA) packages, the results achieved are quite good [50]. As well, the Weibull statistical approach takes into account variations in quality due to various causes, such as assembly abnormalities. As was seen in this study, variation of the percentage of probable failed components, as well as the refinement of the mesh used causes a large variation in the predicted CTF (see Figure 8.3). The finer solder ball mesh has a larger plastic strain amplitude

versus the coarser solder ball mesh, and thus has a lower calculated cycles to failure than the coarse mesh (see Figure 8.4).

As mentioned earlier, the Coffin-Manson equation describes the cycles until crack initiation. However, it does not specify the geometry of this initial crack, so it is difficult to say what the crack length is implied by the Coffin-Manson approach and what is the crack length should be for the Paris-Erdogan crack growth. Thus, in our line of best fit used in Figure 7.45 was a quadratic line, so that the variation of ΔK values at low crack lengths is shown to vary and may be inaccurate.

Figure 8.3: Varying the Weibull exponent in the modified Coffin Manson equation for 0.01% probability and 0.0001% probability. Data shown is for the finest solder ball mesh.

The fracture mechanics approach shown here has disadvantages simply because we are mixing elements of linear elastic fracture mechanics and elastic-plastic fracture

Figure 8.4: Cycles to failure predicted by the Coffin Manson equation for coarse mesh and fine mesh.

mechanics. Strictly speaking, one should not apply one and the other at the same time. Although [26] used the J-integral for fracture mechanics of solder without commenting on its limitations, it is far from clear how accurate the J-integral is for fracture mechanics of solder balls under cyclic loading at high temperatures. As well, the J-integral is not a fatigue parameter; it may give conservative results as shown in [6]. The reason for being conservative is that the J-integral is strictly a steady state crack growth parameter, so it does not accurately handle unloading and reloading of the crack. As a crack grows and reaches steady state, not cyclic loading, with all other parameters being equal, [6] found that a J value would tend to steadily increase.

Limitations to the J-integral approach are shown in the case in the $20\mu m$ crack, where stresses from sources other than the crack tip appear to cause the J-integral

values to be higher than one would expect. Thus, in the presence of other stress singularities, the J-integral cannot be expected to accurately capture the stress singularity caused by cracking. The J-integral method is analogous to a speedboat running at steady speed on a large calm lake. Since the lake is calm, the boat progresses along, and the wake it causes is caused by the progression of the boat in the water. Likewise, the J-integral is the representation of the stress fields caused by the crack progressing in the material. However, if the boat encounters rough waters, then the wake becomes disturbed and does not represent the passing boat accurately. In our case, the “rough waters” is the edge effects caused by the short $20\mu m$ in proximity to the edge of the solder ball. The J-integral would thus capture the crack tip stress singularity as well as the surrounding stress singularities in the material. It is an imperfect analogy, but illustrates the purpose of the J-integral.

With that being said, the J-integral method here can be used in many applications where steady state crack growth is desired to be modeled, such as pipeline structural analysis, pressure vessel analysis, or nuclear applications. Such applications can have a small crack initiate on the inside of the pipe/pressure vessel, and assuming a constant pressure, evaluate the time taken to reach the outer surface and rupture. One must note however that the tractions on the cracking faces are not zero, and that the growth cannot be modeled as the crack approaches the outer periphery of the pipe/vessel (growth here would tend not to be steady state). As well, since

the J-integral can be reduced to an linear elastic fracture mechanics parameter, it is suitable for brittle fracture analysis, such as in ceramics or glasses.

The J-integral and T^* integral essentially exhaust the continuum approaches to crack characterization. One would have to resort to a Gurson type models, where one models pore initiation and growth. Also, with respect to the size of the mesh used in the solder ball analysis, it is difficult to use a continuum approach since FEM elements in the solder ball can be approximately the same size as the grain size. This is a shortcoming that cannot be avoided when using finite elements with the assumption of an isotropic continuum on such a small part such as this. The element sizes used here may not contain the several hundred grains needed to give a suitable approximation of the average stress within an element due to grain boundary interaction. One should use a model that utilized theory for single crystal plasticity to more accurately capture these effects.

8.2 Suggested Future Work

Although a considerable amount of effort and work has been put into this analysis, there is still more than can be done to give more accurate results:

1. Extend the J-integral code so that it computes the J-integral for all three J-integral vectors in the (x, y, z) directions.

2. Extend the J-integral code to cover theory covered by the T^* integral.
3. Develop a Gurson model for crack growth in solder alloys and other ductile metals.
4. Physical testing of the integrated circuit in question so that numerical results can be verified, and a better value found for the Weibull exponent.

Bibliography

- [1] Atluri, S. N., et. al., *Incremental Path-Independent Integrals in Inelastic and Dynamic Fracture Mechanics*, Engineering Fracture Mechanics, Vol. 20, No. 2, pgs. 209-244.
- [2] Barsom, J.M., and Rolfe, S.T., *Fracture and Fatigue Control in Structures - Applications of Fracture Mechanics*, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, United States of America, 1977.
- [3] Bonet, J., and Wood, R.D., *Nonlinear continuum mechanics for finite element analysis*, Cambridge University Press, Cambridge, United Kingdom, 1997.
- [4] Briggs, W. L., *A Multigrid Tutorial*, Society for Industrial and App. Mathematics, Philadelphia, United States of America, 1987.
- [5] Broek, David *Elementary Engineering Fracture Mechanics*, 4th Ed., Martinus Nijhoff Publishers, The Netherlands, 1986.

- [6] Brust, F. W., et. al., *A Combined Numerical/Experimental Study of Ductile Crack Growth After a Large Unloading, Using T^* , J , and CTOA criteria*, Engineering Fracture Mechanics, Vol. 23, No. 3, pgs. 537-550.
- [7] Chen, W., *Constitutive Equations for Engineering Materials; Vol. 2: Plasticity and Modelling*, Elsevier Science B.V., Amsterdam, The Netherlands, 1994.
- [8] Chung, T.J., *The Finite Element Analysis in Fluid Dynamics*, McGraw-Hill, Inc., United States of America, 1978.
- [9] Darveaux, R. “*Crack Initiation and Growth in Surface Mount Solder Joints*”, *Proc. ISHM International Symposium on Microelectronics*, 1993, pgs 86-97.
- [10] Engelmaier, W. *Design for Reliability for Surface Mount Solder Attachments: Physics of Failure and Statistical Failure Distributions*, Engelmaier Associates, Inc., Mendham, New Jersey, USA.
- [11] Eriksson, et. al. *Computational Differential Equations*, Chalmers U., Gotenburg, Sweden.
- [12] Frost, H.J., and Ashby, M.F., *Deformation-Mechanism Maps; The Plasticity and Creep of Metals and Ceramics*, Pergamon Press, Great Britian, 1982.
- [13] Fung, Y.C., *A First Course In Continuum Mechanics, Second Edition*, Prentice-Hall Inc., New Jersey, United States of America, 1977.

- [14] Garofalo, F., *Fundamentals of Creep and Creep-Rupture in Metals*, The Macmillan Company, New York, United States of America, 1965.
- [15] Gdoutos, E. E., *Fracture Mechanics - An Introduction*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1993.
- [16] Gebhart, Benjamin. *Heat Conduction and Mass Diffusion* McGraw-Hill, Inc., United States of America, 1993.
- [17] Gresho, P.M., and Sani, R.L., *Incompressible Flow and the Finite Element Method*, John Wiley and Sons, Ltd., United States of America, 2000.
- [18] Harper, Charles A., *Electronic Packaging and Interconnection Handbook*, McGraw & Hill, United States of America, 1991.
- [19] Heinrich, J.C., and Pepper, D.W., *Intermediate Finite Element Method: Fluid Flow and Heat Transfer Applications*, Taylor & Francis, United States of America, 1999.
- [20] Willis, J.R. *Modelling of Defects and Fracture Mechanics*, Edited by G. Herrmann, Springer-Verlag, New York, United States of America, 1993.
- [21] Incropera, F.P., and DeWitt D.P., *Introduction to Heat Transfer* John Wiley & Sons, United States of America, 1990.

- [22] Khan, S.A., and Huang, Sujian, *Continuum Theory of Plasticity*, John Wiley & Sons, Inc., Canada, 1995.
- [23] Lau, John A., and Pao, Yi-Hsin *Solder Joint Reliability of BGA, CSP, Flip Chip, and Fine Pitch SMT Assemblies*, McGraw & Hill, United States of America, 1997.
- [24] Lau, John A. *Solder Joint Reliability - Theory and Applications*, McGraw & Hill, United States of America, 1991.
- [25] Lau, John A. *Ball Grid Array Technology*, McGraw & Hill, United States of America, 1995.
- [26] Lau, John A. *Thermal Fatigue Life Prediction of Flip Chip Solder Joints by Fracture Mechanics Method*, Engineering Fracture Mechanics, Vol. 45, No. 5, pgs. 643-654.
- [27] Logsdon, W, et. al., *Fracture Behaviour of 63Sn-37Pb Solder*, Engineering Fracture Mechanics, Vol. 36, No. 2, pgs. 183-218.
- [28] Lemaitre, J., *A Course on Damage Mechanics*, Springer-Verlag, Germany, 1992.
- [29] Lewis, R.W., et. al., *The Finite Element Method in Heat Transfer Analysis*, John Wiley & Sons, West Sussex, England, 1996.
- [30] Mase, George, E., *Continuum Mechanics*, McGraw-Hill, Inc., United States of America, 1970.

- [31] Martin, P. L., "*Electronic Failure Analysis Handbook*", McGraw-Hill, United States of America, 1999.
- [32] McDowell, L. David, et al. *A Unified Creep-Plasticity Theory for Solder Alloys*, Fatigue of Electronic Materials, 1994, pgs. 42-59.
- [33] Miannay, D.P., *Fracture Mechanics*, Springer-Verlag, New York, United States of America, 1998.
- [34] Miller, A.K., *Unified Constitutive Equations For Creep and Plasticity*, Elsevier Applied Science, New York, United States of America, 1987.
- [35] Miyamoto, H., et. al., *Three-Dimensional J Integral*, Dept. of Mechanical Engineering, Faculty of Science and Technology, Science University of Tokyo, <http://kikuanni.me.noda.sut.ac.jp/paper/paper5.htm>
- [36] Mocanita, Mihaela, *Computational Mechanics of Welding Complex Structures*, Carleton University, Ottawa, Canada, September 2000.
- [37] Norrie, D.H., and deVries G., *Introduction to Finite Elements*, Academic Press Inc., New York, 1978.
- [38] Owen, J., and Hinton, E., *Finite Elements In Plasticity: Theory and Practice*, Pineridge Press Ltd., Swansea, United Kingdom, 1980

- [39] Roache, P.J., *Elliptic Marching Methods and Domain Decomposition*, CRC Press, United States of America, 1995.
- [40] Barsom, J.M., and Rolfe, S.T., *Fracture and Fatigue Control in Structures - Applications of Fracture Mechanics*, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, United States of America, 1977.
- [41] Simo, J.C., *Algorithms for static and dynamic multiplicative plasticity that preserve the classical return mapping schemes of infinitesimal theory*, Comp. Meth. in Appl. Mech. and Eng., v. 99, 1992.
- [42] Simo, J.C., Hughes, T.J.R. *Computational Inelasticity*, Springer-Verlag, 1998
- [43] Strang, Gilbert, *Introduction to Applied Mathematics*, Wellesley-Cambridge Press, Wellesley, Mass., United States of America, 1986.
- [44] Suresh, S., *Fatigue of Materials*, Cambridge University Press, United Kingdom, 1998.
- [45] Tu, P.L., et al., “*Effect of Intermetallic Compounds on Thermal Fatigue of Surface Mount Solder Joints*”, *IEEE Transactions on Components, Packaging, and Manufacturing Technology - Part B*, Vol. 20, No. 1, Feb. 1997. pgs 87-93.
- [46] Viswanadham, P., et. al., *Failure Modes and Mechanisms in Electronic Packages*, Chapman & Hall, United States of America, 1998.

- [47] Wilshire, B., and Owen, D.R.J, *Engineering Approaches to High Temperature Design: Vol. 2 in the Series "Recent Advances in Creep and Fracture of Engineering Materials and Structures"*, Pineridge Press, Swansea, United Kingdom, 1983.
- [48] Zienkiewicz, O.C., *The Finite Element Method*, McGraw-Hill, Inc., London, England, 1977.
- [49] Zienkiewicz, O.C., and Taylor, R.L. *The Finite Element Method; Vol. 2: Solid Mechanics*, 5th Ed., Springer-Verlag, 1998
- [50] Goldak, J., Private Communication, 2002.
- [51] Aldea, V., Private Communication, 2002.