1

National Library
of Canada

Bibliothèque nationale
du Canada

Canadian Theses Service

Service des thèses canadiennes

Ottawa, Canada
K1A 0N4

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, tests publiés, etc.) ne sont pas microfilmés.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30

**Three-Dimensional, Finite Deformation,
Thermal-Elasto-Plastic Finite Element Analysis**

by

Alan Sinclair Oddy
B. Eng. (Mechanical), M. Eng. (Materials)

A thesis submitted to
the Faculty of Graduate Studies And Research
in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy in Engineering

Department of
Mechanical and Aeronautical Engineering
Carleton University
Ottawa, Ontario
Canada

October 1987

©1987
Alan Sinclair Oddy

Permission has been granted to the National Library of Canada to microfilm this thesis and to lend or sell copies of the film.

The author (copyright owner) has reserved other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without his/her written permission.

L'autorisation a été accordée à la Bibliothèque nationale du Canada de microfilmer cette thèse et de prêter ou de vendre des exemplaires du film.

L'auteur (titulaire du droit d'auteur) se réserve les autres droits de publication; ni la thèse ni de longs extraits de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation écrite.

ISBN 0-315-40608-9

P
The undersigned hereby recommend to
The Faculty of Graduate Studies and Research
acceptance of the thesis,

**Three-Dimensional, Finite Deformation,
Thermal-Elasto-Plastic Finite Element Analysis**

submitted by

A. S. Oddy, B.Eng., M.Eng.
in partial fulfilment of the requirements
for the degree of Doctor of Philosophy

Chairman, Department of
Mechanical and Aeronautical Engineering

Thesis Supervisor

.....
External Examiner

Carleton University
November 1987

Abstract

Various processes such as welding, casting and forging have been used to manufacture metal objects for centuries. The complex thermal and deformation histories which occur in these processes usually leave the article with very large residual stresses. These residual stresses can seriously reduce the service life or in extreme cases render the object unfit for service. These objects nearly always have complex shapes, experience very large deformations and temperature changes and in some cases undergo phase transformations. A three-dimensional (3D), finite deformation, thermal-elasto-plastic finite element algorithm using centred strains and the Green-Naghdi stress rate is presented. A comparison of numerical results to several analytical solutions shows that the algorithm correctly solves the mathematics underlying large deformation thermal stress problems.

One example of a particularly difficult problem of this type is the thermal deformation which occurs in welding. Thermal expansion causes plastic deformation which can result in very large residual deformations and stresses. A comparison between numerical predictions and experimental measurements of the residual stresses in a short edge weld shows acceptable agreement. Traditionally, the prediction of residual stresses in welds has usually only been possible by assuming that the conditions can be adequately represented by a two-dimensional (2D) analysis. Experimental evidence available in many places suggests that the deformation pattern is never 2D. A full 3D analysis of a butt weld also shows deformation patterns, stress fields and strain fields which prove that 2D plane strain conditions do not exist in welds.

Acknowledgements

Nothing is ever done by one person alone. This thesis is no exception. The assistance of various individuals and organizations has been invaluable. I would like to thank my advisors, Dr. M.J. Babby and Dr. J.A. Goldak for the many hours of assistance which they gave me. The technical assistance of Dan Downey, Stephanie Hunn, Warren Hik, Balvant Patel and many others has resulted in a better thesis in a shorter time than I would have been able to do by myself. Without the generous financial assistance of Carleton University, the Natural Sciences and Engineering Research Council (NSERC) and the Independent Order of the Daughters of the Empire (IODE) it would not have been possible for me to return to university. Most of all I would like to thank my wife Moyra, who is responsible for more of this than she realises and more than is wise for me to admit.

Table of Contents

1.	Introduction	1
2.	Method	10
2.1	Finite Deformation	10
2.2	Stress Rate	16
2.3	Incremental Theory	38
3.	Validation	44
3.1	Rigid Body Rotation Test	46
3.2	Hyperelastic Tensile Test	49
3.3	Thermal-Elastic Deformation Test	52
3.4	Transformation Plasticity Test	56
4.0	Results	60
4.1	Comparison with Experiment	60
4.2	Butt Weld	75
5.0	Conclusions	90
6.0	Future Work	91
7.0	References	93
8.0	Appendix 1	100

List of Figures

	Page
Figure 1. Multistep Finite Deformation.	15
Figure 2. 10° Rigid Body Rotation.	48
Figure 3. Hyperelastic Tensile Test, Force vs Elongation.	51
Figure 4. Thermal-Elastic Bending Test.	54
Figure 5. Thermal-Elastic Bending Test. Displacement vs Temperature Difference.	55
Figure 6. Transformation Plasticity Test. Axial Stress vs Temperature. The Transformation occurred between -600°C and 300°C.	57
Figure 7. Finite Element Mesh used in Edge Weld Analysis.	64
Figure 8. Temperature Distribution Around Heat Source, Time = 7.6 s; Edge Weld, Outer Surface. Contours in °C.	65
Figure 9. Temperature Distribution Around Heat Source, Time = 7.6 s; Edge Weld, Symmetric Plane. Contours in °C.	66
Figure 10. Deformation Around Heat Source, Time = 7.6 s; Edge Weld, Outer Surface.	67
Figure 11. Deformation Around Heat Source, Time = 7.6 s; Edge Weld, Symmetric Plane.	68
Figure 12. Permanent Deformation Pattern after Complete Cooling; Edge Weld	69
Figure 13. Longitudinal Residual Stress (σ_{xx}); Edge Weld. Contours in Pa.	70
Figure 14. Residual Shear Stress σ_{xy} ; Edge Weld. Contours in Pa.	72
Figure 15. Experimental and Numerical Longitudinal Residual Stress σ_{xx} ; Edge Weld.	71
Figure 16. Finite Element Mesh used in Bead-on-plate Analysis Rotated to Expose Top.	77
Figure 17. Deformation Around Heat Source, Time = 7.6 s; But Weld.	79

List of Figures, cont'd

	Page
Figure 18. Permanent Deformation Pattern after Complete Cooling; Butt Weld.	80
Figure 19. Longitudinal Residual Stress (σ_{xx}); Butt Weld. Contours in Pa .	81
Figure 20. Residual Shear Stress σ_{xy} ; Butt Weld. Contours in Pa .	82
Figure 21. Longitudinal Residual Stress (σ_{xx}) on Cross-section at Midlength; Butt Weld. Contours in Pa .	83
Figure 22. Transverse Residual Stress (σ_{yy}) on Cross-section at Midlength; Butt Weld. Contours in Pa .	84
Figure 23. Through-thickness Residual Stress (σ_{xx}) on Cross-section at Midlength; Butt Weld. Contours in Pa .	85
Figure 24. Residual Shear Stress (σ_{xy}) on Cross-section at Midlength; Butt Weld. Contours in Pa .	86
Figure 25. Residual Shear Stress (σ_{yz}) on Cross-section at Midlength; Butt Weld. Contours in Pa .	87
Figure 26. Residual Shear Stress (σ_{zx}) on Cross-section at Midlength; Butt Weld. Contours in Pa .	88

Nomenclature

A_0	Initial cross-sectional area, hyperelastic tensile test.
B	Derivative operator.
D_{avg}	Average material property tensor.
D_{ijmn}	Elastic material property tensor.
$\dot{D}_{ijmn}$	Rate of change in elastic material property tensor.
d	Beam depth, Thermal-Elastic test.
E	Young's Modulus, hyperelastic tensile test.
E_{ij}	Green-Lagrange strain tensor.
e_{ij}	Almansi strain tensor.
F	Axial Force, hyperelastic tensile test.
$f(s)$	Scaling factor relating fraction transformed and transformation plasticity.
$g(s)$	Scaling factor relating fraction transformed and transformation plasticity rate.
ℓ	Final length of uniaxial stretch specimen.
ℓ_0	Initial length of uniaxial stretch specimen.
F	Deformation gradient tensor.
J_{ij}	Jacobian of transformation.
K	Empirical constant relating transformation plasticity to deviatoric stress.
K'	Empirical constant for transformation plasticity rate.
K	Stiffness matrix.
n	Time at beginning of an increment.
$n + 1$	Time at end of an increment.
$n + \gamma$	Time at at some intermediate point during an increment.
R_{n+1}	Rotation tensor of final orientation.
$R_{n+\frac{1}{2}}$	Rotation tensor of midpoint orientation.
R^T	Transposed rotation tensor.
R	Radius of yield surface in hyperstress space.
$\mathfrak{R}$	External nodal force vector.
$\mathfrak{R}_I$	Initial stress pseudoforce vector.
$\mathfrak{R}_{Th}$	Thermal and transformation pseudoforce vector.
r	Radius of curvature, Thermal-Elastic test.
S_{ij}	Deviatoric stress.
T	Temperature.

$\dot{T}$	Rate of change of temperature.
Δt	Duration of increment.
U	Symmetric, positive-definite tensor related to deformation.
u_i	Displacement.
u^{corr}	Correction to nodal displacements.
$\frac{\Delta V}{V}$	Fractional volume change in transformation.
X_j	Initial geometry.
x_j	Final geometry.
Δx	Horizontal displacement, Thermal-Elastic test.
Y	Flow stress, from radius of yield surface.
Δy	Vertical displacement, Thermal-Elastic test.
z	Fraction transformed.
$\dot{z}$	Transformation rate.
Δz	Increment in fraction transformed.
$\bar{\alpha}$	Instantaneous coefficient of thermal expansion.
α	Average coefficient of thermal expansion.
γ	Relative position in time, within an increment.
δ_{ij}	Kronecker delta.
ϵ_{ij}	Strain tensor.
ϵ^{corr}	Correction to incremental strain vector.
ϵ^{El}	Total elastic strain at beginning of increment.
ϵ^{net}	Net strain tensor; i.e., sum on incremental total strains.
$\epsilon^{spurious}$	Spurious strain caused by rotation.
ϵ^{true}	True strain in uniaxial stretch or compression.
$\epsilon^{virtual}$	Virtual strain in virtual work equation.
$\Delta \epsilon_{mn}^{Th}$	Incremental thermal strain tensor rotated to initial orientation.
$\dot{\epsilon}_{ij}^{Th}$	Incremental thermal strain rate tensor rotated to initial orientation.
$\Delta \epsilon_{ij}^{Tot}$	Incremental total strain tensor rotated to initial orientation.
$\dot{\epsilon}_{ij}^{Tot}$	Incremental total strain rate tensor rotated to initial orientation.
$\dot{\epsilon}_{ij}^{Tot'}$	Incremental total strain rate tensor in unrotated orientation.
$\Delta \epsilon_{ij}^{Tr}$	Incremental transformation strain tensor rotated to initial orientation.
$\dot{\epsilon}_{ij}^{Tr}$	Incremental transformation strain rate tensor rotated to initial orientation.
$\Delta \epsilon_{ij}^{TrP}$	Incremental transformation plasticity strain tensor rotated to initial orientation.

$\dot{\epsilon}_{ij}^{TrP}$	Incremental transformation plasticity strain rate tensor rotated to initial orientation.
$\Delta \epsilon_{ij}^{TrV}$	Reversible, hydrostatic, incremental volume change occurring during transformation, rotated to initial orientation.
$\dot{\epsilon}_{ij}^{Tr}$	Incremental transformation strain rate tensor rotated to initial orientation.
Θ	Angle of rotation.
ν	Poisson's ratio, hyperelastic tensile test.
Φ	Scaling factor on estimated incremental transformation plasticity.
Ψ	Residual force vector.
σ^b	Cauchy stress vector existing at end of increment.
σ'	Current guess of Cauchy stress existing at end of increment.
σ^{corr}	Correction to Cauchy stress existing at end of increment.
σ_{ij}	Cauchy stress tensor.
$\dot{\sigma}'$	Rotated stress rate.
$\dot{\sigma}$	Unrotated Cauchy stress rate.
$\Delta \sigma$	Stress increment.
$\dot{\tau}$	Green-Naghdi stress rate.

1. Introduction

There exists a class of problems in solid mechanics which involve transient stresses, transient temperatures, plastic deformation, thermal expansion/shrinkage, large strains, large displacements and phase changes. Their significance varies but all these phenomena are present in manufacturing processes such as casting, forging, rolling, and welding. One goal of computer aided engineering (CAE) is the rigorous analysis of the heat flow, metal flow, stresses and strains in these processes by correctly modelling the physics involved.

One phenomenon of considerable interest is that of residual stresses. Residual stresses are those stresses existing in a body when there are no external loads or body forces. One source of them is the plastic deformation which occurs in many manufacturing processes. If an initially stress free body is subjected to an arbitrary set of forces it is possible that non-uniform plastic deformations will occur. If these loads are removed the body will elastically unload. Plastic deformation has deformed some parts of the body more than others. It may be that a stress free state after the loads have been removed is impossible if the body is to remain continuous. The stresses that remain after all the loads have been removed are called residual stresses. Large residual stresses often exist in an object after manufacturing. These stresses can reduce the usefulness of the object by altering the performance in its intended use.

The prediction of residual stresses can be very important. This is especially true of welding even though the physical processes causing residual stresses are not limited to welding. Residual stresses are insidious and dangerous since they can cause the premature failure of ships, pipelines, reactor piping and many other welded structures. Failure to predict residual stresses can be costly.

In a weld, the primary cause of residual stresses is the thermal cycle. The heat source in a weld moves, melting some of the base metal and usually depositing more molten metal. Determining the temperature history of any point involves the analysis of transient, nonlinear 3D heat flow including phase transformations and latent heat effects. Prediction of the thermal history is a difficult problem in itself. The temperature increase which occurs as the source approaches any point causes thermal expansion of the material. Since the material close to the source is surrounded by colder material this expansion is frustrated

and stresses appear. The actual stress state is triaxial, with the major principal stress being compressive. The stresses are frequently large enough to cause a large volume of material to yield. Thus the material is usually said to be yielding in compression. This does not imply that a hydrostatic compressive stress is causing plastic deformation, nor that a state of uniaxial compression exists. This volume which has yielded is called the upset zone. This zone of plastically deformed material is usually several times larger than the volume of fused metal. For mild steels a rough rule-of-thumb states that the upset zone is approximately equal to the volume of material which experienced a temperature increase of more than 100 C. After the source has passed the material begins to cool and shrink. Because of the initial yielding in compression the body cannot return to its original shape, free of stress. As it shrinks tensile stresses appear in the upset zone. These stresses may again become larger than the yield stress causing more plastic deformation, this time in tension. Phase transformations such as the austenite to ferrite transformation in steel, can occur in some metals over fairly narrow temperature ranges. If phase changes occur as part of the thermal cycle then any volume change will further complicate the issue. Determining the residual stresses in welds requires a nonlinear, transient, thermal-stress analysis. As will be seen, it is necessary to analyse the transient stresses in three dimensions.

It is well known that residual stresses equal to or greater than the yield stress exist after the weld has completely cooled. Associated with these residual stresses are permanent deformations of the body. This deformation, known as welding distortion, is frequently large enough to be visible to the naked eye. More than this, it is frequently so large that the object cannot perform its intended function or fit its intended location. Residual stresses and distortion are two faces of the same problem, that of thermally induced plasticity. Prediction and control of residual stresses and distortion is crucial to the fabrication of welded structures.

In order to predict residual stresses it is necessary to investigate large deformation, thermal-stress problems where plastic deformation occurs. Analysis of this sort is quite complicated. Empirical and closed-form techniques were abandoned long ago in favour numerical techniques. The Finite Element Method (FEM) is one technique that has demonstrated promise in the analysis of this phenomena. The finite element algorithm presented here is intended to analyse three dimensional (3D), transient, thermal-elasto-plastic, large deformation and phase transformation phenomena. The class of problems in which temper-

ature changes and phase transformations cause large, plastic deformations in 3D objects is not limited to welding. Most of the discussion focuses on welds but the algorithm developed is not limited to welds.

The numerical analysis of welds has been continually evolving. Initially only linear elastic small deformation problems were soluble. The addition of efficient algorithms for the analysis of nonlinear problems like plasticity occurred only approximately 15 years ago. Since that time the numerical analysis of thermal stresses in welds has received considerable attention over the last decade. A complete list of the references is beyond the scope of this thesis. Since the original work of Hibbitt and Marcal^[1] progress has been steady^{[2],[3],[4],[5],[6],[7],[8],[9],[10],[11]}.

The difficulties to be overcome in the analysis of welds are common to many manufacturing processes. In welds the cause of all stress changes is the thermal cycle. The temperature field is transient, three dimensional and highly localized. It is well known that accurate predictions of the thermal history require that the temperature dependence of the properties be included^{[12],[13],[14]}. Prediction of the thermal history is not part of this thesis. It remains however a difficult and important problem. For the purposes of this work it will be assumed that the thermal history is known. The methods used to predict thermal histories may be found in^{[7],[11],[12],[15]}.

Once the thermal history is known it becomes possible to attempt to predict the thermal stress history of a weld. The stress analysis of a weld is also 3D, transient and nonlinear. Temperature changes cause thermal expansion. This expansion causes stresses to appear. The stresses may exceed the yield strength of the material, resulting in plastic deformation. Material properties are again temperature dependent. Deformation is frequently very large, leading to substantial changes in shape^{[16],[17]}. These may be the result of large deflections and large strains. Large in this sense means that they are clearly visible to the naked eye. This is even greater than what is normally considered large in continuum mechanics. In continuum mechanics, as will be seen, large means elongations greater than 10% or rotations greater than 1°. Phase transformations, occurring in certain temperature ranges cause volume changes and other complex phenomena. In order to rigorously model the physical processes associated with welding a 3D, nonlinear, transient, thermal-elasto-plastic, finite deformation, finite element algorithm is required.

4

In order to attempt to model the complexity of the problem various assumptions have been made by many authors. It has been very common to assume that longitudinal heat flow; i.e., flow along the weld centreline could be neglected. The consequences of this will not be dealt with here except to say that this condition is only satisfied for high speed welds^{[13],[18],[19]}. When it became possible to analyse the thermal stresses in welds superficially similar assumptions were made. By far the most common assumption^{[1],[2],[3],[4],[5],[6]} has been that the strain field is only two dimensional (2D). On occasion, the full plate is analysed under conditions of plane stress^[4]. In such an analysis it is assumed that temperatures, displacements and stresses are uniform through the thickness of the plate. This corresponds to the conditions expected in thin plates with full penetration welds. The more frequent assumption is that a state of plane strain or axisymmetry exists^{[1],[2],[3],[5]}. This reduces the problem to that of a cross-section of the plate. More recently so-called generalized plain strain analyses have been attempted^{[10],[20]}. This generalized plane strain is nothing more than the assumption that plane sections remain plane, that normal strains perpendicular to the cross-section are not zero and shear stresses between adjacent sections are zero everywhere. Both the generalized and traditional plane strain analyses assume that all quantities are uniform along the length of the weld.

These plane strain assumptions do not stand up to examination. They have become so well entrenched however that even some experimental measurements of residual stress assume it from the outset^[21]. The adoption of this idea most likely occurred because 2D plane strain analyses were all that could be afforded. The justification may have been that if the weld centreline does not change length during the thermal cycle then there must never be any longitudinal movement of the material. However, material moving forward and then backward, or a small net shift backward would not change the length.

What are the consequences of the plane strain assumptions? Consider a large plate with a weld along its centre. If a fully plane strain state exists then small slices taken perpendicular to the weld will never change thickness. The compressive stress which occurs in the upset zone early in the cycle will not create tensile stresses in the material far away from the weld centreline. This happens because the plane strain state in effect puts this thin slice between two rigid walls. After the weld has cooled, the ensuing shrinkage near the weld centreline creates tensile stresses inside the upset zone but is again completely unable to create compressive stresses outside the zone. We know that these compressive stresses

must exist in order that the body remain in equilibrium. Reconnecting all the slices to recreate the plate would result in a stress state requiring external tensile loads at the ends of the plate. This may partially explain the move to generalized plane strain analyses.

In plane strain, the direction along the weld centreline, perpendicular to the slice, must be a direction of principle stress and principle strain. Thus, another consequence of the plane strain assumptions is that there can be no shear stresses in the plane of the plate, parallel or perpendicular to the weld. This must be true at the plane of symmetry running along the centre of the weld. It is, however, unlikely to be true a short distance away. Consider a large plate. If a strip running along the centre of the plate were to shrink, then at the junction between the shrunken and unshrunken material there would exist shear stresses. They would join the region of tensile longitudinal stresses (inside the shrunken region) to the region of compressive longitudinal stresses (outside the shrunken region). Welds are more complex than this but they do share similarities with this simple mental experiment.

In a butt weld a plane of symmetry often exists along the weld centreline if two identical plates are joined. If the two halves of the plate were not joined together then the shrinkage occurring in the material near the weld would cause each half to bend in such a way that a convex cavity would exist between the two halves. Since this does not occur when the two halves are joined together, there must be forces acting on the plane of symmetry which prevent this. Only normal forces can exist on a plane of symmetry and thus transverse normal stresses appear along the weld centreline. These transverse stresses will be compressive near the ends of the plate and tensile in the middle regions of the plate. These conditions exist in all bead-on-plate welds but at this point let us consider only a bead-on-plate weld where the heat source fully penetrates the plate and there is no variation in any quantity through the thickness. Transverse stresses will still be present and still exhibit the same distribution. Andersson's^[4] experimental results possess exactly this pattern of transverse stresses. Furthermore, Andersson reports exactly the same pattern for his plane stress (thin plate) analysis where, again, all quantities are uniform through the thickness. Plane strain, and generalized plane strain; i.e., cross-sectional, analyses will never capture this phenomena because of the implicit assumption that all quantities are uniform along the length of the weld. It is a trivial matter to show that the presence of transverse normal stresses also requires the presence of shear stresses in the plane of the plate. This lack of

transverse stress and shear stress in the plane strain analyses will affect the onset of plastic deformation. The magnitude of these stresses will be very important in determining their significance. However, for the simple case of a full penetration, bead-on-plate weld, a plane strain analysis will not predict the transverse normal stresses, will not possess all the shear stresses present and will not predict plasticity correctly.

In any weld there will exist small regions on both ends where the thermal and stress histories differ from those in the mid-length because of transients caused by turning the source on and off or by the presence of the plate ends. The regions are known as end-transients. These end-transients are purely a 3D phenomenon. It is well known that the thermal end-transients are quite small and are usually ignored. Some experimental evidence suggests that the stress end-transients can be quite large. In a 1m bead-on-plate weld Andersson's^[4] experimental measurements show 20cm long regions on both ends of the weld where the stress state is significantly different from that of the mid-length.

Finally, consider the displacement pattern to be expected around the heat source. In an infinite plate a stationary heat source would cause a uniform thermal expansion in the material around it. It would be expected that radial displacements would occur in the material near the source. Neither full plane strain or even generalized plane strain would be appropriate assumptions. If the source started moving, one would still expect a general expansion in the material surrounding the source. Because of the motion one might expect these radial displacements to be skewed, thus enhancing backward displacements behind the source at the expense of forward displacements in front of the source. This displacement pattern still does not satisfy plane strain conditions.

Mental experiments are fine but they are not proof. If these conditions are expected then why haven't experiments shown them? The answer is partially that dynamic stress and displacement measurements near the weld centreline are very difficult. Another part of the answer is that the hold exerted by the plane strain assumption is so strong that very few experimenters have bothered to test it. Some measurements^{[9],[21]} of residual stress in welds start off with the assumption that two of the shear strain components are always zero. In fact, experiments capable of testing the plane strain assumption show it to be false. Duncan^[22] showed displacement fields incompatible with plane strain. Hernandez^[23] measured strains near the heat source in aluminum welds with a Moire fringe technique. The longitudinal strains measured were not zero, proving the plane strain assumption to be

invalid. The longitudinal strains did not vary linearly with the distance perpendicular to the weld line. Thus the assumption that plane sections remained planar was very unlikely.

The best demonstration of the error in these assumptions was that of Johnson^[24]. The strain patterns around welds in aluminum sheets were measured with a Moire fringe technique. It was found that straight lines scribed on the surface perpendicular to the weld centreline always bent away from the source. That is, they were bent a small amount in the direction of motion of the source as the source approached. Later, after the source had passed, these lines were then bent a much larger amount in the direction opposite to the motion of the source. Plane strain conditions did not exist during the weld and plane sections did not remain plane. It was also found that the directions of maximum shear strain in the plane of the plate were nearly parallel and perpendicular to the weld centreline. In fact, shear strains in the plane of the plate were much larger than the normal strains. The plane strain assumption is clearly not valid.

One prediction made by Johnson^[24] deserves special attention. In a stationary heat source the radially symmetric temperature field causes a general radial expansion in the material around the source. Thus radial displacements away from the source would be expected. The surrounding cooler material would resist this expansion to some extent and create a radially symmetric compressive stress field in the material near the source. In a moving arc, the temperature pattern becomes asymmetric, with the steepest thermal gradients directly in front of the source. This will increase the magnitude of the compressive stresses in front of the source and reduce them behind the source. Johnson^[24] predicted that plastic deformation under these conditions will cause a net backward shift of material. This backward shift did occur and was measured for a series of welds. For a constant weld bead size, it was found that the backward displacement was greater for faster moving welds and the upset zone was smaller. Welding speed seems, from these results, to strongly influence this backward shift. Thus conditions that reduced longitudinal heat flow increased longitudinal metal flow.

What the consequences of these plane strain assumptions are, is difficult to say. The work of Ueda^[20] and Leung^[10] both varied the longitudinal restraint in a 2D finite element analysis by comparing full plane strain to generalized plane strain conditions. Making generalizations from these results is dangerous since all cases neglect the shear strains. However, in both it was found that increasing the longitudinal restraint caused an increase

in the residual tensile stresses existing in the weld after cooling back to room temperature. This tendency is not too surprising since increases in longitudinal restraint should increase the amount of compressive yielding. However this seems subtly at variance with the experimental results of Johnson^[24] where, at first glance, it would seem that higher welding speeds ought to correspond to greater restraint.

It may well turn out that an analyst's requirements can be met adequately with a 2D stress analysis in most cases. Before this can be proven it is necessary to be able to analyse the stress and deformation in 3D. There will always be cases where a full 3D analysis is required just to capture the geometry of the object. Based on the experimental results presented above, a 3D analysis may be required in all cases. The analysis of 3D thermal-elasto-plasticity in welds would be a major step forward. Patel's^[7] algorithm was capable of 3D, thermal-elasto-plastic, small deformation, small strain analysis without transformation effects. For reasons of economy full 3D analysis of a weld was never attempted. Part of work reported here was also reported by Goldak^[11]. That paper is, to the author's knowledge, the first 3D, nonlinear, transient analysis of the stresses in a weld. It was only a small deformation, small strain analysis with no transformation effects. The only other attempt at full three dimensional (3D) analysis of a weld was a linear elastic analysis of Lai^[6].

Most authors recognize the need for temperature dependent material properties. The correct constitutive relation^{[6],[25]} is one which includes the effects of changes in the material properties. The correct addition of the temperature dependent material properties will almost certainly be as important to the stress analysis as it has been to the thermal analysis^{[12],[13],[14]}.

Phase transformations causing expansions and contractions of approximately 1% are well known. It is difficult to see how accurate stress predictions are possible if these are ignored. More unusual phenomena such as transformation induced plasticity^{[26],[27],[28],[29]} will also have important effects. Transformation plasticity occurs through the interaction of the stresses and the volume change during the transformation. Hibbitt^[1] speculated that the poor correlation between their experimental and analytical results was due to some additional, unexplained phenomena. As will be demonstrated, the differences are completely consistent with the effects of transformation plasticity. A few investigations have attempted to include the expansion due to austenite decomposition in steels^{[1],[2],[4],[5],[30]}. None have included the contraction during heating. The only attempts at including transformation

plasticity in welds have done so by specifying an artificially low yield stress during the transformation^{[6],[20]}. As will be demonstrated this is a heuristic assumption that has no relation to physical processes occurring during a transformation.

Weld deformations are often very large, so large in fact that the change in shape is clearly visible to the eye. Wan^[16], and Hogan^[17] measured camber in edge welded bars. Stainless steel bars only 1 m long showed centre deflections of up to 1.5 cm. Rotations in butt welds and tee joints of several degrees are common. Deflections like this fall into the region of large strain, large deflection problems. In spite of this most analysts use small strain, small deformation algorithms. Implicit in the mathematics of these codes is the assumption that the shape changes are negligible; that displacements and rotations are so small they can be ignored. Only Hibbitt^[1], in one of the earliest and most ambitious attempts, performed a 2D large deformation, small strain analysis.

Ample experimental evidence exists to indicate that the stresses and strains in welds are three dimensional, that deformations and strains are large, that material properties change during the analysis and that phase transformation effects are significant. Nevertheless, none of the past investigations have properly addressed these issues, much less attempted all four at the same time. What will be described here is an algorithm which brings together all these phenomena in order to analyse incremental, 3D, large displacement, large strain, thermal-elasto-plastic stress problems with transformation effects, and a temperature dependant constitutive relation.

2. Method

The general class of problems to be examined involve many time steps, large displacements and rotations, large strains, plasticity and phase changes. That is to say, temperature changes cause thermal expansion or contraction. This deformation may change the shape or orientation of the body significantly. This expansion also causes stresses to appear. These stresses may exceed the yield strength of the material and that yield strength may also have changed due to the change in temperature. In certain temperature ranges there will be phase transformations and associated volume changes. There may also be an interaction between stresses caused by the transformation and any applied stress leading to further plastic deformation. The general conditions to be analysed are that of a 3 dimensional body, with an initial stress state, subjected to a known temperature change and set of loads. During the time step there are potentially large strains and deformations. It is necessary to predict the deformations, strains and stresses that exist at the end of the time step taking into account thermal expansion/contraction, plasticity and phase changes.

To accomplish the task outlined above, an existing nonlinear, thermal-elasto-plastic finite element code *ASGARD*, was taken and modified. The original program was capable of 3D thermal stress analysis^[7] but a number of additions were made. These additions and changes were intended to achieve several goals; to reduce run-times and memory requirements, to change the basic formulation so that large deformation and strain problems could be analysed and to include additional physical phenomena.

Problems where significant changes in geometry occur are usually divided into 2 classes. An excellent introduction to these types of problems are available in Bathe^[31]. The first class is one in which large displacements cause only small strains. The second and more comprehensive class is one in which large displacements and large strains are possible. These types of problems have been called finite deformation problems^{[32],[33],[34],[35]}. It is this latter class of problems which will be examined here.

2.1 Finite Deformation

In designing any finite deformation algorithm, there are a number of important decisions that must be made at a very early stage. Excellent discussions of the nature and implications of these decisions are available in the literature^{[31],[33],[36]}. The choices to be

made are; the discretization model to be used, the deformation description and the stress rate description. These choices influence one another and place requirements on each other. The wide variety of mathematical formulations which exist for these problems^{[32],[33],[37]} is the result of making different choices from among the alternatives. For the most part these algorithms are valid. It is obvious that when making these choices the model selected must be mathematically correct. It is equally true that the type of physical phenomena under examination will also influence the choice. Beyond this, two relatively simple criteria have been used in making the choices. It was highly desirable that the algorithm be computationally efficient and easy to add to the existing finite element program. Of greater significance is the notion that it be compatible with the physical processes which are occurring. Algorithms which make use of unnatural or unreal quantities are bound to be more complex and unlikely to model reality correctly. It is clearly preferable to use an algorithm which, in addition to satisfying the mathematical requirements, is derivable from, or relatable to, physical conditions, quantities or processes. Such an algorithm is more likely to produce results of the desired form and make interpretation of these results more convenient.

The choice of discretization model is closely tied to the type of phenomena to be analysed. The so-called Lagrangian or material description is one in which the mesh follows the motion of the material as it deforms. This is the one most commonly used in solid mechanics because surfaces are followed naturally and history dependent variables such as plastic strain are easily passed on from step to step. In the alternative, known as a Eulerian or spatial description, the mesh remains fixed in space and the material flows through the volume occupied. This is commonly used in fluid mechanics. A more thorough discussion of this may be found in Belytschko^[32]. The class of problems of interest here may involve large changes in the shape of a solid object due to the action of surface or body forces. These forces will change their magnitude, direction and most importantly, their location with time. A Lagrangian formulation is appropriate. Because of the incremental nature of the loads, the object at any arbitrary time step may look nothing like its initial shape. An Updated Lagrangian formulation has therefore been selected. That is, at the end of every time step all static and kinematic variables are updated. This new geometry is used for the computation of element stiffness properties, derivatives, etc. in the next step. Incremental displacements, strains, stresses, etc. are computed with respect to this current geometry.

In its original form ASGAR had a Total Lagrangian formulation. All static and

kinematic variables were referred to the initial state. In other words element stiffnesses and strains were computed with respect to the geometry that existed at the beginning of the problem. The data requirements for incremental loads were defined in terms of the change in the load during the increment, not the actual load present during the increment. This was very inconvenient since external loads may appear, disappear, move and/or change magnitude and direction during the total time span being examined. To do incremental, large deformation problems in a Total Lagrangian formulation the data requirements would have been inconvenient and the code complicated and obtuse. It is far more convenient to use an Updated Lagrangian formulation; i.e., update all variables so that every time step is like a completely new problem. Every time step then has some initial geometry and stress state. Boundary conditions at the end of the step are known and the displacements and change in stress required to satisfy equilibrium are calculated.

Conversion of ASGARD from a Total Lagrangian to an Updated Lagrangian formulation required changes or additions to 45 subroutines involving approximately 17000 lines of code. A brief list of the work required would include: changes to update geometry, associated derivative operators, stresses and strains; a redefinition of load types into real forces and pseudo-forces; alterations to the radial-return plasticity routines and; modifications to the routines calculating residual forces. The line search subroutine using Crisfield's^{[38],[39]} method was also modified extensively. Although many of the changes were not large, considerable effort had to be expended to understand the underlying mathematics and modify the existing code. Although the discussion focuses on 3D thermal stress analysis, a full 2D large deformation thermal stress capability was included.

In its original form ASGARD was fully capable of 3D thermal-elasto-plastic stress analysis. For economic reasons Patel^[7] was unable to analyse a full 3D weld. Part of the work described here involved the addition, with McDill^[15], of the Isoparametric Mesh-Grading Hexahedrons developed by McDill^[40]. More than anything else, it is the mesh grading hexahedrons of McDill^{[15],[40]} that have made 3D stress analysis of welds possible. The significance of this advance cannot be over emphasized.

The choice of the deformation description and stress rate influence one another. Many exist. Some are better than others. The choice should be influenced by the ease of coding and the relation to reality and physical processes. The largest barrier to understanding the mathematics of large deformation is the fact that several of the parameters are not

easily relatable to well known measurable quantities. They exist only as mathematical abstractions.

Several valid measures of strain exist. They have names like Green-Lagrange strain, Almansi strain and Centred strain. They are all different from the well known infinitesimal strain, which in indicial notation is;

$$\epsilon_{ij} = \frac{1}{2} \left(\frac{\partial u_i}{\partial X_j} + \frac{\partial u_j}{\partial X_i} \right) \quad (1)$$

This measure of strain, is related to the derivative of the displacements, u , with respect to the initial geometry X . The difficulty with this measure of strain is that it neglects several higher order terms of the derivatives. Consequently, significant errors will appear whenever the derivatives become large. These derivatives may be large even if the strain is small. For instance, serious errors occur when rigid body rotation is present. Belytschko^[32] shows that an infinitesimal strain tensor produces spurious strains when rigid body rotations are present. These spurious strains are approximately;

$$\epsilon^{spurious} = \frac{1}{2} \Theta^2 \quad (2)$$

where Θ is the angle of the rotation in radians. Clearly even a small rotation can lead to serious errors. As an example, in steel, a 1° rotation translates into a stress of approximately 30 MPa. Belytschko^[32] also recommends that the infinitesimal strain measure not be used for strains of the order of 0.1 even if rotations are not present. In any weld or forging the deformations are often large enough to be visible. Camber in welded bars often causes rotations much larger than 1° .

These errors and spurious strains are solely the result of neglecting the higher order terms of the derivatives. If these higher order terms are included then one obtains either the Green-Lagrange strain;

$$E_{ij} = \frac{1}{2} \left(\frac{\partial u_i}{\partial X_j} + \frac{\partial u_j}{\partial X_i} + \frac{\partial u_k}{\partial X_i} \frac{\partial u_k}{\partial X_j} \right) \quad (3)$$

where again the derivatives of the displacements are being taken with respect to the initial geometry X , or the Almansi strain;

$$e_{ij} = \frac{1}{2} \left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} - \frac{\partial u_k}{\partial x_i} \frac{\partial u_k}{\partial x_j} \right) \quad (4)$$

where the derivatives of the displacements are being taken with respect to the final geometry x . Both the Green-Lagrange strain and the Almansi strain are perfectly valid measures of large strains. It will be obvious, however, that for a given initial geometry and set of displacements they will not give rise to the same strain tensors. Neither of these measures are even approximations to the True strain, which for a uniaxial stretch is;

$$\epsilon^{true} = \ln(l/l_0) \quad (5)$$

The name True strain does not imply that the other strain measures are false. This strain measure is obtained from integrating the instantaneous strain rate over the time step. The Centred strain, described later, was adopted in this endeavour. It will be shown to be a good approximation to the True strain.

Most engineers are familiar with the concept of true strain. It would be an advantage if the eventual code produced a strain that was similar to the true strain. There are also issues in the evaluation of the constitutive equations where true strain and strain rate are preferred. In a multistep problem, especially where reversed plasticity occurs, intuition suggests that the net strain ought to be the sum of the stepwise strains. Here lies one of the inconvenient aspects of both the Green-Lagrange and Almansi strain tensors. Consider the 2 step problem shown in Figure 1. A body with an initial length l is stretched to twice its original length, $2l$. For this increment there exists an initial state and a final state and all 4 strain measures could be computed. They will all give different values, but this is not a concern at this time. In the second step the body with an initial length $2l$ is compressed until it is again l . For this set of initial and final states one could again compute all 4 strain measures. It is obvious that the net strain accumulated over both steps is now zero. If, as intuition suggests, one simply adds the stepwise strains then one will not obtain a net strain of zero for the small strain, Green-Lagrange strain, or the Almansi strain. Needless to say the sum of the incremental true strains is zero. In an incremental code it will be important to use a strain measure which can give the correct results for this problem.

If this problem were run using many existing finite element codes a net strain of zero would be obtained. This is only possible because no matter what step was being examined the derivatives of the displacements were always being evaluated with respect to the geometry that existed at the beginning of the first step. This is clearly absurd. In any step, there exists a body with an initial geometry and stress state. This body is deformed during the

Figure 1. Multistep Finite Deformation.

step. It finishes that step with a new geometry and stress state having experienced some strain during the step. There is no theoretical or mathematical reason why the derivative of the displacements in n^{th} step should be evaluated with respect to a state which existed in the dim and distant past. Furthermore if the body had undergone a rigid body rotation during one of the intervening steps, then displacements in the horizontal direction would create strains that should not be related to the horizontal direction in the original state.

2.2 Stress Rate

The stress rate is the final choice to be made. Many viable stress rates exist. They produce different types of stress. As a minimum, there are 2 requirements that must be met to satisfy mathematical constraints. First, the stress and strain measures used must be conjugate; i.e., their inner product must equal work. Second, they must be objective or frame invariant. This second requirement states that if a rigid body rotation R is superimposed on some deformation pattern, then the relationship between the stress rate with the rotation $\dot{\sigma}'$ and the stress rate without the rotation $\dot{\sigma}$ must be;

$$\dot{\sigma}' = R \dot{\sigma} R^T \quad (6)$$

These are minimum requirements, they are necessary but not sufficient.

Various stress rates have been proposed that satisfy these minimum requirements. The names of a few of the more popular are the Jaumann stress rate, the Truesdell stress rate and the Green-Naghdi stress rate. A thorough discussion of these rates will not be included here since they are readily available^{[32],[34],[35]}. Instead, a general discussion of their advantages is all that will be included.

Perhaps the most popular of these stress rates is the Jaumann stress rate. Although it satisfies the minimum requirements it is not derived from a physically meaningful stress. Under some circumstances it suffers from an instability, which leads to incorrect results. Nagtegaal^[41] showed that large, monotonically increasing shear strains resulted in an oscillating stress response for kinematic hardening. Johnson^[34] showed that this happened even in hyper-elastic shear problems and that any instability existing in the elastic solution of the simple shear problem will also appear with kinematic hardening in elastic-plastic problems. A stress rate exhibiting such a response is unacceptable. Although Voyiadjis^[42] claims that the use of a Lagrangian description avoids this problem, no evidence is given

to support this contention. Indeed it is difficult to see how this can be true in the light of Johnson's work^[34].

The Truesdell stress rate, another popular rate is an expression of the rate of change of the symmetric Piola-Kirchoff stress^[34]. What is usually desired is the Cauchy, or True stress. Cauchy stresses are defined as the body internal forces per unit area in the configuration at the end of the increment; i.e., the force per unit area which exists in the body in the deformed state. Piola-Kirchoff stresses are not the same as Cauchy stresses although one may be calculated from the other. Symmetric and unsymmetric Piola-Kirchoff stress tensors exist and both are measures of the force existing in the deformed state per unit undeformed area. The actual, mathematically correct definitions are somewhat more complicated but it is obvious that both are mathematical abstractions only. They are not physically meaningful quantities. Readers interested in more detail on this subject are referred to Malvern^[36]. The chief drawback to this stress rate is that it too has not been derived from a physically meaningful stress. It is therefore discarded in favour of one which is.

The stress rate advocated here and elsewhere^{[33],[34],[35]} is the Green-Naghdi stress rate. It can be shown^[34] that this stress rate is in fact derived from a physically meaningful stress. It is related to the rate of change in the rotated Cauchy stress; i.e., the Cauchy stress rotated back to the initial orientation. This stress rate is preferred for several reasons. First, the fact that it is derived from a physically meaningful stress has to be a significant advantage both in terms of understanding the underlying principles, and in obtaining meaningful results. Second, constitutive relations for thermal-elasto-plasticity are written in terms of the rate of change of the Cauchy stress with respect to the strain rate. Third, a simple, straight-forward formulation is always easier to implement, operate and modify for future needs. When implemented, such a formulation virtually guarantees that the additional cost of doing a finite deformation problem will be negligible compared to a small deformation analysis.

The algorithm adopted shares many similarities with that of Hallquist^{[33],[44]}. The details will be described in greater depth later. In this work, the Green-Naghdi stress rate is evaluated using the thermal-elasto-plastic constitutive relation^{[6],[25]}. This ensures a proper evaluation of the effects of thermal expansion as well as the effects caused by the temperature dependence of the material properties. The method used is very easy to

add to an incremental, Updated Lagrangian small deformation finite element program. It avoids the instability shown in the Jaumann stress rate, produces Cauchy stresses directly and uses a strain measure which closely approximates the true strain.

Much of the following derivation has been reduced to exclude some of the more involved aspects of the mathematics. Mathematical proofs and more precise derivations are available elsewhere^{[34],[35],[36]}. Since one is trying to model physical processes, wherever possible the explanations will be in terms of those physical processes. As will soon become apparent, a consistent subscript notation is almost incomprehensible if it must include quantities that are tensors, vectors, transposed vectors, quantities that are evaluated at different points in time, or in different orientations or states, plus subtle distinctions between various types of similar variables. Some quantities like stress and strain may be used as vectors or tensors depending on the circumstance. For the most part tensor notation is limited to only a very few cases. Virtually all quantities are functions of time although this is explicitly denoted only to indicate terms evaluated at specific instants in time. Functional dependence, including time, is indicated using bracketed subscripts. Superscripts denote different types or classes of similar variables. In some instances, differences in orientation of a particular quantity are indicated with a prime; i.e.,

In any time step the deformation experienced by a body could be a combination of stretching, shearing and rotation. It is customary to assume at any instant in time γ , during any time step from n to $n + 1$, where $0 < \gamma < 1$, that the location x of a specific point is;

$$x_{(n+\gamma)} = (1 - \gamma)x_{(n)} + \gamma x_{(n+1)} \quad (7)$$

where $x_{(n)}$ is the position of the point at time n , $x_{(n+1)}$ is the location at time $n + 1$ and $x_{(n+\gamma)}$ is the location at time $n + \gamma$. This assumption states that during any time step all points move in straight lines with constant velocities. This is an important assumption because it makes a series of simplifications possible while, at the same time, placing certain limits on the amount of deformation allowable in any time step. A rigid body rotation of 180° cannot be represented by such an approximation. Any attempt to do so forces all the material to pass through one point in space at the mid-point of the step^[35].

During the step in question it is possible to evaluate the deformation gradient F at any

point γ in time;

$$F_{(n+\gamma)} = \frac{\partial x_{(n+\gamma)}}{\partial x_{(n)}} \quad (8)$$

In ASGARD the Jacobian J in indicial notation is defined as;

$$J_{ij} = \frac{\partial x_j}{\partial \xi_i} \quad (9)$$

Thus;

$$F_{(n+\gamma)} = J_{(n+\gamma)}^T (J_{(n)}^{-1}) \quad (10)$$

Using the polar decomposition method of Hoger^[43], this can be decomposed into;

$$F = RU \quad (11)$$

where U is a symmetric, positive-definite tensor related to the stretching which has occurred and R is the proper-orthogonal rotation tensor. For reasons which will become obvious, F is evaluated at the mid-point of the step and at the end of the step to produce rotation tensors for the orientation at the mid-point of the step and the orientation at the end of the step.

The Green-Naghdi stress rate is often described^{[33],[34],[35]} in terms of a rate of rotation tensor, the Cauchy stress tensor, and the rate of change in the Cauchy stress tensor. This is the most popular form of the equation but not the simplest form. In its most simple form, it can be shown^[34] that the Green-Naghdi stress rate $\dot{\tau}$ can be represented as;

$$\dot{\tau} = R\dot{\sigma}R^T \quad (12)$$

where $\dot{\sigma}$ is the rate of change of the rotated Cauchy stress. The Green-Naghdi stress rate is, therefore, the rotated Cauchy stress rate rotated to the current or final orientation. The rotated Cauchy stress is the Cauchy stress rotated back to the initial orientation. If the rate of change of the rotated Cauchy stress is evaluated and the stresses are updated in the initial orientation then the algorithm becomes very simple. The updated Cauchy stresses are then rotated to the final orientation. This procedure preserves objectivity, is very simple to implement and is, in fact easier to understand.

As recommended by Hallquist^[33] the stress rate will be evaluated in the initial orientation. From this, an increment to the rotated Cauchy stress will be evaluated and added

to the initial stress which existed at the beginning of the increment. The Cauchy stress at the end of the increment will be found by rotating this sum, to the final orientation. That is;

$$\sigma_{(n+1)} = R_{(n+1)}(\sigma_{(n)} + \Delta\sigma)R_{(n+1)}^T \quad (13)$$

where;

$$\Delta\sigma = \int \dot{\sigma} dt \quad (14)$$

The radial return algorithm with isotropic and kinematic hardening of Patel^[7] is used to account for plasticity occurring during the step. Quantities like the total strain, plastic strain, and back stress for kinematic hardening must also be rotated forward to the final orientation.

A one point rule will be used to integrate the stress rate. That is;

$$\Delta\sigma = \int \dot{\sigma} dt \approx \dot{\sigma}_{(n+\frac{1}{2})} \Delta t \quad (15)$$

This may appear to be a sweeping simplification. In fact it is fundamental to the best finite deformation theories currently used^{[33],[35]}. For elastic problems with temperature independent mechanical properties this condition must be true since eqn. 7 assumes that points move in straight lines at constant velocity. For nonlinear, thermal-elasto-plastic problems with temperature dependent mechanical properties this condition limits the size of the steps somewhat, but not seriously so. This implementation of the Green-Naghdi stress rate is virtually identical to the formulation found in NIKE2D^[33] and DYNA3D^[44] finite deformation codes.

In a finite deformation analysis, significant deformation of the elements is a distinct possibility. Element shape can affect the interpolation functions obtained in the physical domain. Thus element distortion can seriously affect performance. This has been noted by many authors^{[31],[45],[46],[47],[48],[49],[50],[51],[52],[53],[54]}. A need for some automatic indication of excessive element distortion exists. The distortion metric of Oddy^[55] was added to ASGAR. A complete description of this metric is contained in Appendix 1.

Constitutive Model

In the numerical analysis of elasto-plastic problems it is customary to assume initially that the deformation is elastic. If the resulting stresses exceed the yield point

then the stress is corrected for plasticity. Various different algorithms exist^{[56],[57]}. It is generally acknowledged that the best of these is the so-called Elastic Predictor, Radial Corrector^{[35],[44],[56],[57]}. This was originally added to ASGAR by Patel^[7].

Initially assuming that the body undergoes an elastic step leads to an incremental constitutive relation^{[6],[25]}. This differs from the common constitutive model in that it correctly accounts for changes in the mechanical properties due to temperature changes. In indicial notation, the incremental relation becomes;

$$\dot{\sigma}_{ij} = D_{ijmn}(\dot{\epsilon}_{mn}^{Tot} - \dot{\epsilon}_{mn}^{Th} - \dot{\epsilon}_{mn}^{Tr}) + \dot{D}_{ijmn}(\epsilon_{mn}^{El} + \Delta\epsilon_{mn}^{Tot} - \Delta\epsilon_{mn}^{Th} - \Delta\epsilon_{mn}^{Tr}) \quad (16)$$

where: $\dot{\sigma}_{ij}$ is the rate of change in the Cauchy stress; D_{ijmn} is the elastic property tensor and $\dot{D}_{ijmn}$ is the rate at which it is changing; ϵ_{mn}^{El} is the total elastic strain at the beginning of the increment; $\Delta\epsilon_{mn}^{Tot}$ and $\dot{\epsilon}_{mn}^{Tot}$ are the incremental total strain and total strain rate respectively; $\Delta\epsilon_{mn}^{Th}$ and $\dot{\epsilon}_{mn}^{Th}$ are the incremental thermal strain and thermal strain rate respectively; and $\Delta\epsilon_{mn}^{Tr}$ and $\dot{\epsilon}_{mn}^{Tr}$ are the incremental transformation strain and transformation strain rate respectively. These quantities will be described in greater detail later. The fact that the stress rate depends on the rate at which the material properties are changing as well as the strain rates is something frequently overlooked. All the terms in eqn 16 may change during the time step.

In equation 15 it was assumed that the stress increment could be approximated by the product of the stress rate at the mid-point of the step and the length of the time step. Switching to matrix notation, equation 16 becomes;

$$\Delta\sigma = \left(D_{(n+\frac{1}{2})} \left(\dot{\epsilon}_{(n+\frac{1}{2})}^{Tot} - \dot{\epsilon}_{(n+\frac{1}{2})}^{Th} - \dot{\epsilon}_{(n+\frac{1}{2})}^{Tr} \right) + \dot{D}_{(n+\frac{1}{2})} \left(D_n^{-1} \sigma(n) + \Delta\epsilon_{(n+\frac{1}{2})}^{Tot} - \Delta\epsilon_{(n+\frac{1}{2})}^{Th} - \Delta\epsilon_{(n+\frac{1}{2})}^{Tr} \right) \right) \Delta t \quad (17)$$

Material properties and their rate of change are therefore being evaluated at the mid-point of the time step. This is an important advantage. If too large a temperature change occurs during the step and material properties change too much it is possible for numerical instabilities to occur^{[6],[58]}. Properties evaluated at some intermediate point in time are known as finite approximation schemes^[6]. These generally yield better approximations to the solution path^[6]. Argyris^[58] states that for a linear approximation of the relation between time and state variables, the evaluation of constitutive properties at $\gamma \geq \frac{1}{2}$ leads to unconditional stability for the step.

In the following sections it will be assumed that all the strain components can be related to the strain rates in the form;

$$\Delta \epsilon_{(n+1)} \approx \dot{\epsilon}_{(n+\frac{1}{2})} \Delta t \quad (18)$$

It will be further assumed that;

$$\Delta \epsilon_{(n+\frac{1}{2})} \approx \frac{1}{2} \Delta \epsilon_{(n+1)} \quad (19)$$

This allows further simplification to eqn 17. During any time step the material properties are assumed to change smoothly. The temperature may vary linearly with time but the material properties may not vary linearly with temperature. Instead of assuming a linear variation of the properties with time it is assumed that;

$$\dot{D}_{(n+\frac{1}{2})} = \frac{D_{(n+1)} - D_{(n)}}{\Delta t} \quad (20)$$

Equation 17 reduces to;

$$\Delta \sigma = D_{avg} (\Delta \epsilon_{(n+1)}^{Tot} - \Delta \epsilon_{(n+1)}^{Th} - \Delta \epsilon_{(n+1)}^{Tr}) + (D_{(n+1)} - D_{(n)}) D_n^{-1} \sigma_{(n)} \quad (21)$$

where;

$$D_{avg} = \left(D_{(n+\frac{1}{2})} + \frac{D_{(n+1)} - D_{(n)}}{2} \right) \quad (22)$$

The algorithm will evaluate the constitutive properties using the temperatures at the start, middle and end of the time step.

Patel^[7] collected temperature dependent material properties for several steels. They form the basis of the material properties used here. The constitutive model used here is an elasto-plastic one. The materials used all soften considerably as the temperature rises. At high temperatures the materials behave as visco-plastic solids until they liquify at approximately 1500 C. A elasto-plastic material model will fail when the yield strength and elastic modulus become very small. The elasto-plastic material loses its strength and cannot sustain any load. To correctly model this material, a visco-plastic constitutive model is required. Fortunately the volume of material in the visco-plastic range is small in a weld. This is the approach taken by Hibbitt^[1] and Argyris^{[6],[58]}. Above the melting point even visco-plastic analyses may not be appropriate. In this regime the equations of fluid mechanics apply.

An alternative approach to the problem of softened material, one that would allow large volumes of liquid material to coexist, would be to use a separate substructure for all elements above some selected temperature. The total problem would be divided into two subproblems. The liquid material would have no effect on the stress analysis of the elasto-plastic region. The displacements in the elasto-plastic solid would be found first. To ensure that the liquid region followed the motion of the solid region, a separate analysis of the liquid region could be performed. It is important that the liquid volume remain attached to the solid and change shape as the solid boundaries do. To do this, the displacements found in the analysis of the solid, for nodes joining the two regions, would be applied to the liquid volume. An arbitrary set of material properties for a perfectly plastic material would enable the change in shape of the liquid region to be obtained using the standard elasto-plastic algorithm. Stresses in the liquid region are of no interest. All that is important is that the liquid region deform as the solid boundaries do in order that the liquid and solid remain attached. This would be very useful when determining the residual stresses in castings, where large volumes of liquid and solid material coexist.

Time did not permit the inclusion of a substructuring capability. The material properties collected by Patel^[7] include modifications designed to avoid these problems. Above an arbitrary temperature the properties listed in the data base have been fixed to specific values. In effect, material properties don't change above a cutoff temperature. Instead properties that existed at the cutoff temperature are substituted. This cutoff temperature varies depending on the property and material but is usually in the range from 800 C to 1200 C.

Total Strains

Strain and strain rate terms were present in eqn 17. It was from this that eqn 21 was derived. In order that eqn 21 give the proper stress increment it is necessary to evaluate the proper strain tensors. The instantaneous total strain rate;

$$\dot{\epsilon}_{ij}^{Tot} = \frac{1}{2} \left(\frac{\partial v_i}{\partial x_j} + \frac{\partial v_j}{\partial x_i} \right) \quad (23)$$

sometimes known as the velocity strain, may be evaluated at any point during the time step. It is the derivative of the instantaneous velocity v with respect to the instantaneous configuration x . If the body is rotating, then to get the Green-Naghdi stress rate it is

necessary to rotate the strain rate back to the initial orientation;

$$\dot{\epsilon}^{Tot} = R_{(n+\frac{1}{2})}^T \dot{\epsilon}^{Tot'} R_{(n+\frac{1}{2})} \quad (24)$$

It can be shown^[34] that the so-called Centred strain tensor^{[33],[34],[35]} is;

$$\Delta \epsilon^{Tot} = \dot{\epsilon}_{(n+\frac{1}{2})}^{Tot} \Delta t \quad (25)$$

which becomes;

$$\Delta \epsilon^{Tot} = R_{(n+\frac{1}{2})}^T \Delta \epsilon^{Tot'} R_{(n+\frac{1}{2})} \quad (26)$$

where;

$$\Delta \epsilon_{ij}^{Tot'} = \frac{1}{2} \left(\frac{\partial u_i}{\partial x_{j(n+\frac{1}{2})}} + \frac{\partial u_j}{\partial x_{i(n+\frac{1}{2})}} \right) \quad (27)$$

Therefore the algorithm takes the derivative of the displacements occurring in the step with respect to the geometry that existed at the mid-point, and rotates this tensor back to the initial orientation. Several authors^{[33],[35]} have shown that this closely approximates the equation for the True strain in a uniaxial stretch test. This is not surprising considering their common basis.

After the constitutive equation has been evaluated, the net strain; i.e., the sum of the increment in the total strain and the net strain at the start of the time step is calculated;

$$\epsilon_{(n+1)}^{Net} = \epsilon_{(n)}^{Net} + \Delta \epsilon_{(n+1)}^{Tot} \quad (28)$$

This new net strain is in the original orientation and must be rotated, onto the final orientation;

$$\epsilon_{(n+1)}^{Net''} = R_{(n+1)} \epsilon_{(n+1)}^{Net} R_{(n+1)}^T \quad (29)$$

In the two step, reversed elongation problem outlined earlier, the sum of the incremental total strains gives a zero net strain with this algorithm.

Thermal Strains

When the temperature of a body changes there is always a change in the dimensions as a result of the thermally induced vibrations of the atoms. This change is reversible and, in polycrystalline metals, isotropic. The expansion and contraction associated with the thermal cycle is the basic cause of plastic deformation, distortion and residual stresses in welds.

The property relating temperature change to the change in linear dimensions is called the linear coefficient of thermal expansion. This property is temperature dependent. Some care must be taken when examining the literature on this subject. Two types of coefficient exist, either the instantaneous coefficient of thermal expansion α , or an average coefficient of thermal expansion $\bar{\alpha}$. A more thorough discussion of this problem is available in Patel^[7]. The thermal strain rate in indicial notation is;

$$\dot{\epsilon}_{ij}^{Th} = \bar{\alpha} \dot{T} \delta_{ij} \quad (30)$$

where $\bar{\alpha}$ is related to α by;

$$\bar{\alpha} = \alpha + \frac{\partial \alpha}{\partial T}(T - T_{ref}) \quad (31)$$

Patel^[7] collected an extensive set of data relating α and temperature T . This format, convenient, appropriate and easy to add to, was retained.

In any time step from n to $n + 1$, at time γ , where $0 < \gamma < 1$, the temperature T of any point is assumed to be;

$$T_{(n+\gamma)} = (1 - \gamma)T_{(n)} + \gamma T_{(n+1)} \quad (32)$$

where $T_{(n)}$ is the temperature of the point at time n , $T_{(n+1)}$ is the temperature at time $n + 1$ and $T_{(n+\gamma)}$ is the temperature at time $n + \gamma$. Thus the temperature rate $\dot{T}$ is;

$$\dot{T} = \frac{T_{(n+1)} - T_{(n)}}{\Delta t} \quad (33)$$

It is also assumed that the material properties are a smooth function of the temperature, thus;

$$\frac{\partial \alpha}{\partial T} \approx \frac{\alpha_{(n+1)} - \alpha_{(n)}}{T_{(n+1)} - T_{(n)}} \quad (34)$$

Therefore;

$$\Delta \epsilon_{ij}^{Th} \approx \dot{\epsilon}_{(n+\frac{1}{2})}^{Th} \Delta t \quad (35)$$

Returning to indicial notation, it can be shown that;

$$\Delta \epsilon_{ij}^{Th} = \left(\alpha_{(n+\frac{1}{2})}(T_{(n+1)} - T_{(n)}) + (\alpha_{(n+1)} - \alpha_{(n)})(T_{(n+\frac{1}{2})} - T_{ref}) \right) \delta_{ij} \quad (36)$$

The form of this equation is identical to that of Patel^[7]. The only difference here is that some of the temperatures and coefficients are evaluated at the mid-point of the step, not at

the end. This conforms more closely to the ideal that all rates be evaluated at the mid-point of the step for use in a one point integration over the step.

An important consideration, when computing thermal strains is the order of the temperature field to be used. When an element type is selected; e.g., 8-node brick or 20-node brick, then a specific set of interpolation functions is applied to the volume occupied by the element. They will be linear for a 8-node brick and quadratic for the 20-node brick. These functions describe the variation of temperature and displacement in the element. The total strain as we have seen is a function of the derivative of those displacements. In most finite element codes the thermal strain is calculated directly from the temperature field. Thus the thermal strain field has the same order as the temperature field while the total strain field has an order one less than the displacement field. Since the temperature field is of the same order as the displacement field, the two strain fields will be incompatible. This could lead to convergence problems.

Consider the following example. A single, linear element is subjected to an arbitrary, but linear temperature field. It is allowed to expand freely. The total strain ought to be identical to the thermal strain because the stresses ought to be zero. Unfortunately, there is no set of nodal displacements that will create a total strain field in a linear element that is compatible with the thermal strain field. There will, therefore, always be a stress present. To the author's knowledge this has never been examined in depth. If the linear terms in the temperature field are dominant then large errors could result.

In this investigation the linear, grading hexahedrons of McDill^[40] were used exclusively. A linear variation in displacement was therefore used in any element. To avoid an incompatibility in the thermal and total strain fields the following limitation was applied to the temperature field. In any element, at any instant in time, the thermal strain at any point was computed from the temperature at the centroid of the element. This created a constant thermal strain field in the element. This does cause errors to creep into the solution. However, totally artificial convergence problems, caused only by incompatible strain fields, are avoided. Baldly stated, small errors were preferred to convergence problems for two reasons. In areas where the temperature gradient is high the elements will tend to be very small. Thus assuming constant thermal strains is not too bad. In a highly nonlinear analysis, the problems associated with difficulties in convergence could degrade the solution without the user being aware of it. Some tests were performed to compare the effects of us-

ing the full thermal strain field versus a constant thermal strain field. Only small increases in the out-of-balance nodal forces were observed when a thermal strain field of order equal to the displacement field was used. It is likely that larger errors could be obtained under different conditions.

Transformation Strains

Many materials experience a phase change when the temperature passes through some specific range. These allotropic changes are often accompanied by small, but significant changes in the volume of the body. This volume change is reversible in that if the body were to re-transform to its original structure there would be a volume change of opposite sign. The change is usually not isotropic but in polycrystalline materials it can be assumed to be so. Steels exhibit this phenomenon. In welds there are transformations to high temperature phases during heating and a subsequent re-transformations during cooling.

The kinetics of this transformation cycle are quite complicated and will not be dealt with here. Briefly, an initial ferrite-pearlite microstructure transforms to austenite as temperatures rise from 723°C to approximately 900°C. There is approximately a 1% decrease in volume associated with this change. During cooling the austenite decomposes into a number of possible microstructures. This re-transformation often doesn't occur until temperatures are well below 700°C. Because of this hysteresis and the difference in the coefficient of thermal expansion between austenite and ferrite-pearlite, the expansion on re-transformation varies somewhat. Transformation from austenite to a ferrite and iron carbide structure (pearlite, upper bainite, lower bainite) causes an approximately 1% increase in the volume. Transformation from austenite to martensite causes an approximately 4% increase in the volume^[59].

Phase transformations affect the solution in several ways. It is well known that material properties such as yield strength are strongly affected by the microstructure present. There are additional phenomena associated with the volume change that will directly affect the stresses present. Knowing what microstructure is present requires the ability to predict the transformation kinetics. Prediction of the transformation products is a difficult problem in its own right. Eventually the work of Henwood^[60] will have to be added to predict the microstructures present at any point in time. However this is beyond the scope of this endeavour. At this time, material properties are assumed only to be dependent on the

temperature and, for plasticity problems, strain hardening. The thermal history has no effect on the material properties.

Even just predicting the temperature at which the austenite decomposition occurs is quite difficult. The transformation temperature range depends on the cooling rate, composition and austenite grain size and the stress state. An excellent summary of the relation between transformation temperature and stress state is given by Denis^[61]. Much work remains to be done before a clear picture will emerge. Uniaxial tensile or compressive stresses appear to promote transformation^[61] by raising the M_s ; i.e., the temperature at which the first austenite transforms to martensite. Uniaxial tensile stresses appear to cause a greater effect than do compressive uniaxial stresses. High pressures appear to retard the transformation. Although it was not mentioned^[61], this hints that the hydrostatic and deviatoric components of the stress field have different effects. Deviatoric stresses appear to always promote transformation. Hydrostatic stresses may promote or hinder transformation depending on the sign of the stress and the volume change. For instance, compressive hydrostatic stresses appear to retard transformations involving expansion. The thermodynamic basis for understanding why hydrostatic stresses affect transformations is described by the well known Clausius-Clapeyron^[62] equation. Why deviatoric stresses affect phase transformations is less well known. Some evidence suggests that the nucleation rate is affected more than the growth rate^[61]. Eventually this dependence of the transformation on the stress state will have to be included when predicting residual stresses. Since the full relationship has not been discovered yet it will not be included in this analysis.

The reversible volume change during transformation can potentially have a significant effect on the residual stresses in welds. The inclusion of the volume change in the stress analysis of welds has been commonplace for some time^{[1],[3],[4],[6]}. This reversible volume change is not the only phenomena present however. There can also be some irreversible plastic deformation. This phenomena has been known for many years^{[26],[27],[28]}. Experiments have shown^{[26],[27],[28]} that a bar, subjected to a very small stress, changes length as it is passes through the transformation. Reversing the transformation does not reverse the deformation and repeated re-transformation only enhances the deformation. Two mechanisms have been put forward to explain this phenomenon. The first was originally expressed by Greenwood and Johnson^[26]. It stated that the deformation was the result of the gradual accumulation of small plastic deformations which occurred during each transformation.

This plastic deformation occurred because the volume change generated large stresses and thus plastic deformation. Even though the applied stress was small, the volume change raised the stress state above the yield point. The second, proposed by Magee^[63], theorized that, as the atoms rearrange themselves, some directions are preferred due to the presence of stresses. This favouring of some directions over others causes a change in shape. This second mechanism is particularly important in the transformation from austenite to martensite.

It is very important to understand that this phenomenon has nothing to do with creep. The original experiments of Greenwood^[26] and DeJong^{[27],[28]} showed that the deformation was proportional to the stress present but not the time taken to transform. The deformation was therefore not related to creep since it was not affected by the rate of the transformation and stopped when the transformation was complete. The phenomenon was called transformation plasticity. Plastic deformation caused by the transformation is quite distinct from transformations caused by plastic deformation. The work of Leblond^[29] is the first to place transformation plasticity on a sound theoretical basis. In that derivation both the Greenwood-Johnson and Magee mechanisms were present.

Modelling transformation plasticity^{[6],[30],[61],[64],[65]} has received much attention, especially in the last few years. However, it is only recently that any satisfactory analytical investigation of this subject has occurred^{[29],[66]}. Leblond's^[29] derivation discovered that both the Greenwood-Johnson and Magee mechanisms are present during any transformation. The Greenwood-Johnson mechanism would appear to predominate^{[61],[64]} especially when the stresses are large. The Magee type mechanism is more important in martensitic transformations. In order to understand the cause of transformation plasticity it is necessary to examine the micromechanics of phase transformations. Consider an appreciable volume of material that has partially transformed. It is a common fallacy to assume that all of the volume has partly transformed. In fact what has really occurred is that part of the volume has fully transformed. Thus small regions of fully transformed material lie next to regions of completely untransformed material. The volume change occurring during the transformation creates what Leblond^{[29],[66]} has aptly called microstresses. These stresses are self-equilibrating over very small distances; i.e., on the order of the grain size. They can't be measured by conventional techniques but are large enough to cause plastic deformation in the weaker material at the interface between transformed and untransformed

material.

Leblond^[29] has divided the mechanical response of a body undergoing a phase transformation into 2 aspects. The first, in terms of classical plasticity, examines the response of a mixture of two phases to variations in the applied stress or temperature. The presence of the two phases implies that large microstresses exist. The response of such a body to external forces will be altered. In other words the presence of a second phase and large microstresses changes the stress-strain response of the body even if the transformation does not proceed. Leblond^[66] has examined this. The second class of response describes how variations in the phase proportions affect the body. This second phenomenon is the transformation plasticity. Unfortunately no way of dealing with this second condition was given^{[29],[66]}. Since, in welds, the volume of material undergoing a transformation is usually fairly small and the transformation proceeds fairly rapidly, the transformation plasticity is more important than the mixed phase response.

As already stated, in a partially transformed body there are very large stresses at the interface between the transformed and untransformed material. As the transformation proceeds, these stresses at the interface cause plastic deformation in the weaker material. If there are no macroscopic stresses then the direction of this plastic deformation is random and no preferred macroscopic deformation is observed. In the presence of even very small macroscopic stresses, the plastic deformation occurring as the transformation proceeds is no longer randomly oriented. Thus there is some macroscopic deformation. If the macroscopic stresses are constant, then repeated transformation and re-transformation will cause this plastic deformation to accumulate.

It must be emphasized that the Greenwood-Johnson mechanism is, in fact, not a new phenomenon but merely the familiar plastic deformation operating at microscopic levels. Any finite element analysis where the elements were smaller than the size of the grains and the volume change were included as elements fully transform, would automatically create the proper microstresses and plastic deformation. The challenge is to capture the phenomenon without resorting to microscopic elements. Any relations which are used to introduce this phenomenon in large masses of material should be derived from the laws of plastic deformation. These relations must eventually cope with triaxial stress states, partial transformations, and changing stresses.

Are transformation phenomena important? In low carbon steel welds the volume plastically deformed by the thermal expansion is much larger than the region which transforms. In cases such as this, transformation effects may or may not be of great significance. However, for a variety of conditions, modelling transformation phenomena is crucial to the prediction of residual stresses. In an experimental examination of residual stresses in alloy steel welds Darbyshire^[67] showed that the composition difference between the base metal and the filler metal caused differences in the transformation temperatures. This in turn, led to very large residual stresses after post-weld stress relief. Simulated welds^[68] have also shown that volume changes and transformation plasticity can have very large effects on residual stresses. Other examples are welds in high strength steels where the volume transformed is a significant portion of plastically deformed volume. Hibbitt^[1] included the volume change due to transformation in the analysis of HY-80 and HY-130 welds. The predicted residual stresses were much larger than the measured residual stresses. At the time Hibbitt^[1] theorized that an additional unknown phenomenon or process was involved. The discrepancies in the results are completely consistent with the effects of transformation plasticity. Many authors^{[30],[61],[64],[65]} have demonstrated that modelling of the transformation plasticity is absolutely essential when analysing the stresses and strains that occur during the heat treatment of large steel bodies.

When transformation plasticity occurs, the effect is very similar to a sharp reduction in the yield strength. In a weld, the resulting plastic deformation serves to substantially reduce the stresses present^[68]. This probably explains the discrepancies in Hibbitt's^[1] results for the HY-80 and HY-130 welds. To include the transformation plasticity several authors^{[6],[30]} have artificially lowered the yield stress when the transformation occurs. This may capture some of the effects in a material model that includes visco-plasticity but will be less effective in an elasto-plastic analysis. Furthermore, the body does not lose all strength during a transformation and so this heuristic approach is inconvenient at best. As has been clearly pointed out^[64], such an assumption implies that transformation strains may be infinitely large and independent of the fraction transformed. All the experimental evidence shows that the transformation strains are finite and a function of the fraction transformed. A more rigorous approach, based on the concepts of continuum mechanics is required to accurately model this phenomenon.

To model both the volume change and transformation plasticity the transformation

strain ϵ^{Tr} has been divided into two components;

$$\Delta\epsilon^{Tr} = \Delta\epsilon^{TrV} + \Delta\epsilon^{TrP} \quad (37)$$

where $\Delta\epsilon^{TrV}$ is the reversible, hydrostatic volume change and $\Delta\epsilon^{TrP}$ is the irreversible, deviatoric plastic strain caused by transformation plasticity.

It was assumed that for polycrystalline bodies the volume change was isotropic. Thus;

$$\Delta\epsilon_{ij}^{TrV} = \frac{1}{3} \frac{\Delta V}{V} \delta_{ij} \quad (38)$$

where $\frac{\Delta V}{V}$ is the fractional change in volume. This change in volume has been included in many finite element weld analyses^{[1],[3],[4],[6]}. All of these attempts only included the volume change during the transformation on cooling. This of course is totally inadequate in that the volume of the body would grow larger during the thermal cycle. If it is to be included it must be included during heating and during cooling to conserve volume. Tests made with ASGARF confirmed that only including the volume change in the cooling transformation leads to inaccuracies. Including both the expansion and contraction will be particularly important in the analysis of multipass welds.

Two decades ago, Greenwood^[26] derived an expression for transformation plasticity that has still not been improved upon. Although other relations have been proposed it still gives the best agreement with experimental results^[61]. In that analysis, the fundamental equations of plasticity, namely the yield function and the flow rule were used to derive an expression for the transformation plasticity. For a uniaxial stress σ_{xx} , assuming a random orientation to the phase front, and a perfectly plastic material Greenwood^[26] obtained;

$$\Delta\epsilon_{xx}^{TrP} = \frac{5}{6} \frac{\Delta V}{V} \frac{\sigma_{xx}}{Y} \quad (39)$$

$$\Delta\epsilon_{yy}^{TrP} = -\frac{1}{2} \Delta\epsilon_{xx}^{TrP} \quad (40)$$

$$\Delta\epsilon_{zz}^{TrP} = -\frac{1}{2} \Delta\epsilon_{xx}^{TrP} \quad (41)$$

where Y is the flow stress. The flow stress can be obtained from the radius of the yield surface in the deviatoric stress space.

The assumption of a random orientation to the phase front may seem excessive. It is a common fallacy to assume that a single, smooth, continuous phase front sweeps through

the body like an isothermal surface. Continuous is used here in the sense that only 1 surface separates all the transformed material from all the untransformed material. For alloys with significant 2 phase regions in their phase diagram, intuition suggests that a single continuous, smooth phase front never exists. The experiments of Dadian^[69] clearly demonstrate this. For steels, the transformation front is highly irregular, and not continuous. Islands of transformed material, nucleated at grain boundaries, etc., appear and grow in random directions until the transformed regions link up. Dadian^[69] also showed that for Armco iron and pure iron the phase front was continuous and smooth.

The equation proposed by Greenwood^[36] is not able to deal with triaxial stresses and partial transformations. Various heuristic, semi-empirical relations have been proposed to get around these difficulties. A common format^{[61],[64]} is;

$$\Delta \epsilon_{ij}^{TrP} = K S_{ij} f(x) \quad (42)$$

where; K is an empirical constant for the particular alloy under examination, S_{ij} is the deviatoric stress tensor, and $f(x)$ is a scaling factor relating x , the fraction transformed, to the fraction of the full transformation plasticity. The appearance of the deviatoric stress tensor should be of no surprise since the Von Mises yield function was used to derive equation 35. It is well known that only deviatoric and not hydrostatic stresses cause plastic deformation in metals. The use of empirical constants like K is a significant disadvantage. For one thing, empirical constants always indicate areas where understanding is incomplete and thus are eventually replaced as knowledge increases. A more practical consideration is that different values of K are required for different materials. Even for the same material, the austenite to ferrite/cementite transformation will have a different value of K than the austenite to martensite transformation, owing to the different volume change. Complex experiments are required to measure this material factor. The fundamental properties underlying this phenomena may already be known or more easily measured. Thus the use of K would represent a step backward. The need to relate triaxial stresses to transformation strains appears to be the reason for its use. It is not required. A very simple modification to the original formulation eliminates that need. Converted to a more useful and compact form, in indicial notation equations 39, 40 and 41 become;

$$\Delta \epsilon_{ij}^{TrP} = \frac{5}{4} \frac{\Delta V}{V} \frac{1}{Y} S_{ij} \quad (43)$$

The derivation is trivial. Equation 43 may be obtained directly from eqns 39, 40 and 41 or it may be obtained for triaxial stresses by superimposing the effects of the three principal stresses. In either case exactly the same equation is obtained. Equation 43 has the definite advantage that it does not require any empirical constants. It has been derived from fundamental properties using the fundamental laws of plasticity.

Minor changes are required for materials exhibiting isotropic or kinematic hardening. Transformation plasticity is governed by the same laws as regular plastic deformation. It arises from the interaction of the highly irregular stress field caused by the volume change and the macroscopic stress field caused by the external loads, etc. For materials showing isotropic or kinematic hardening the flow stress Y is easily obtained from the radius of the yield surface in hyperstress space; i.e.,

$$Y = \sqrt{\frac{3}{2}} R \quad (44)$$

where R is the radius of the yield surface. For kinematic hardening, the deviatoric stress S_{ij} is obtained in the normal way. The back stress is subtracted from the full stress and the resultant separated into hydrostatic and deviatoric components.

Equations 38 and 43 need further changes. Both must be modified to account for partial transformations and equation 43 should be modified for cases where the stress changes significantly during the time step. Leblond^[29] has proven that a linear relation exists between the volume change and the fraction transformed. Experimental studies seem to indicate that a parabolic relationship exists between the transformation plasticity and z , the fraction transformed^{[64],[61]}. A popular relation seems to be;

$$f(z) = z(2 - z) \quad (45)$$

-- As will be seen a great deal of further work is still required before the definitive relation is found. For this reason, and others to be outlined later, a linear relation between z and the strain will be assumed for both $\Delta\epsilon^{TrV}$ and $\Delta\epsilon^{TrP}$. Therefore;

$$\Delta\epsilon_{ij}^{TrV} = \frac{1}{3} \frac{\Delta V}{V} z \delta_{ij} \quad (46)$$

$$\Delta\epsilon_{ij}^{TrP} = \frac{5}{4} \frac{\Delta V}{V} \frac{z}{Y} S_{ij} \quad (47)$$

It will be recalled that the constitutive relation requires strain rates as well as strains. Conversion to time dependent formulae is a matter of some controversy. Denis^[61] uses a heuristic justification to derive a relation of the form;

$$\dot{\epsilon}_{ij}^{TrP} = K' S_{ij} g(z) \dot{z} \quad (48)$$

They argue that there can be no terms dependent on $\dot{S}_{ij}$ because that would inevitably lead to transformation plasticity when no transformation occurred. This is very attractive and has been confirmed by the analytical investigation of Leblond^[39]. However, the terms derived are not in a useable form.

The assumption of a linear relation between the strain and z allows, for a constant stress;

$$\dot{\epsilon}_{ij}^{TrP} = \frac{5}{4} \frac{\Delta V}{V} \frac{\dot{z}}{Y} S_{ij} \quad (49)$$

The basic form of the heuristic, semi-empirical relations is recovered but a theoretical basis now exists. The linear form of $f(z)$ may appear a drawback. However, further modifications are required in order to avoid serious numerical instabilities. These modifications are required no matter what form $f(z)$ takes. They far outweigh any disadvantages associated with a linear $f(z)$. A linear $f(z)$ allows some simplifications. Once again it will be assumed that;

$$\Delta \epsilon_{(n+1)}^{Tr} \approx \dot{\epsilon}_{(n+\frac{1}{2})}^{Tr} \Delta t \quad (50)$$

Any other form of $f(z)$ will not give this. During a time step Δt , if a fraction Δz transforms then;

$$\Delta \epsilon_{ij}^{TrV} = \frac{1}{3} \frac{\Delta V}{V} \Delta z \delta_{ij} \quad (51)$$

$$\Delta \epsilon_{ij}^{TrP} = \frac{5}{4} \frac{\Delta V}{V} \frac{\Delta z}{Y} S_{ij} \quad (52)$$

Equation 52 is dependent on the macroscopic stress existing in the object.³³ This stress does not always remain constant. Jones^[68] did an experiment in which a small bar at 1300°C was fixed between 2 rigid plates and allowed to cool. Owing to the thermal contraction a tensile stress developed. During the austenite decomposition, stresses dropped rapidly and became slightly compressive. It was claimed that the volume change alone should have caused compressive stresses one order of magnitude larger to appear. Further it was claimed that only the inclusion of the effects of transformation plasticity would explain the results

observed. Unfortunately, insufficient data on the material properties and transformation kinetics, as well as difficulties in knowing the actual experimental conditions preclude any precise numerical simulation of the phenomenon. It does illustrate however that stresses can change dramatically during the transformation. This is especially true when the material is highly restrained. Any transformation plasticity will serve to reduce the magnitude of the existing macroscopic stresses. This is precisely the condition observed by Hibbitt^[1] in the HY-80 and HY-130 welds.

As a body transforms, the potential exists for dramatic changes in the stress state. Numerical analyses^[61] of the martensitic transformation have shown that the volume change alone can cause a 1000 MPa change in the stresses during a 50°C temperature change. If S_{ij} changes greatly then modelling the transformation plasticity requires that very small time steps be used. When studying transformation plasticity in quenched objects in 2D, Denis^[61] used some 25 time steps during a temperature change of 150°C. This may be acceptable for 2D analyses but cannot be tolerated in 3D analyses. In a 3D weld analysis there is always some material behind the source where the austenite is decomposing. Dramatic changes in the stress state will be occurring in this material. Limiting the time steps so that temperatures never change by more than 5-10°C is clearly impractical.

What are the alternatives? Since the transformation plasticity is dependent on the stress and the stress depends on the transformation plasticity, one is trying to solve a nonlinear relation. Another iteration loop could be added to determine what the average stress was during the time step. Thus the transformation strains could be modified, leading to modifications in the stress and so on until convergence occurred. This would be expensive and very inconvenient.

If the initial stress state were used to compute ϵ^{TrP} it would be tantamount to assuming that the stress state did not change during the time step. When this was attempted numerical instabilities resulted. Consider a simulation of the experiment of Jones^[68]. Let material properties for a low carbon steel be adopted. A bar, initially at 1300°C, fixed between two plates, cools and tensile stresses appear. When the transformation starts the tensile stresses drop rapidly. If large time steps are used, then using the initial stress state overestimates the transformation plasticity. It becomes possible for the transformation plasticity to drive the stress state deep into compression. The next time step starts with a very large compressive stress. Transformation plasticity causes even larger tensile stresses

to appear in the next step.

When simulating the conditions in the experiments of Jones^[68], a transformation range from 600°C to 300°C was assumed. If temperature steps larger than 20°C were allowed during the transformation, then the stress state alternated from tension to compression but gradually converged to zero. If temperature changes of 50°C were used then the analysis became unstable. The stress state became more tensile or compressive with each successive step until all accuracy had been lost. In a numerical analysis of any weld, it is very possible that a large time step will cover a large portion of the transformation range. In many cases there will be a large stress change as the transformation proceeds. Some modification to avoid this instability is required.

In a weld, the volume inside the heat-affected zone (HAZ) is highly restrained. It is surrounded by a large mass of untransformed material. The material transformed may start with a large stress that rapidly changes due to the transformation plasticity and volume change. The effect of the transformation plasticity in a highly restrained body is to reduce the magnitude of the stresses. If the stress were reduced to zero then no further transformation plasticity would occur. Clearly, the stress change caused by the transformation plasticity cannot exceed the initial stress. Under conditions of high restraint this produces the condition;

$$|D\epsilon_{ij}^{TrP}| \leq |S_{ij}| \quad (53)$$

If this is expanded, a factor Φ is obtained;

$$\Phi = \frac{5}{4} \frac{\Delta V}{V} \frac{\Delta z}{Y} \frac{|DS_{ij}|}{|S_{ij}|} \quad (54)$$

such that;

$$\Phi \leq 1 \quad (55)$$

This automatically produces a scaling factor that can be used to limit ϵ^{TrP} . If the value of Φ exceeds 1 then ϵ^{TrP} is divided by Φ .

Before closing this section on transformation effects a summary is needed. Transformation kinetics are influenced by composition, cooling rate, and stress state. Prediction of these is very complex and not included here at all. Eventually the work of Henwood^[60] should be added to predict the phases present and the transformation kinetics. In this work such information was unavailable so an approximation had to be used. Material properties

are assumed to be solely dependent on temperature. For those materials which experience a phase change the following approximations were used. The phases present during heating and cooling transformations were assumed to be linear functions of the temperature between transformation start and a transformation finish temperatures. The transformation to austenite was assumed to start at 723°C and end at 900°C. This corresponds roughly to the transformation ranges for the low carbon steels in the ASGARD data base. For the low carbon steels in the ASGARD database, the decomposition of the austenite during cooling was assumed to start at 800°C and end at 600°C. A 1.4% volume change was assumed to occur during the transformation.

The volume change during transformation was included in both the heating and cooling portions of the thermal cycle. Failure to include it in the heating half leads to a net increase in volume during the weld. In multipass welds it is particularly important that this spurious volume increase be avoided. During heating, the transformation occurs in material with virtually no strength and thus very low stress. Consequently the transformation plasticity is of little importance during the heating half of the cycle. Transformation plasticity effects are calculated only during the austenite decomposition in the cooling half of the thermal cycle. The equations used to predict transformation plasticity are rooted in the fundamental laws of plasticity. Wherever possible, semi-empirical equations have been avoided. The resulting equations are very similar in appearance to these empirical relations. Some rather sweeping assumptions have been made but these are overshadowed by conditions imposed to avoid numerical instabilities and to keep computing costs at an affordable level.

A significant amount of space has been devoted to the examination of transformation phenomena. This is indeed unfortunate. It may lead to the erroneous impression that this thesis is primarily an investigation of the numerical analysis of phase transformation effects. This imbalance has occurred because the theory of thermal-elasto-plasticity is well established and does not need elaboration. Transformation plasticity, on the other hand, is much newer and not nearly so well understood. There can be no doubt that transformation effects are very important in the analysis of residual stresses. Thus some attempt must be made to include them. However, the theoretical basis for understanding them is not sufficiently developed. Much more work, both analytical and experimental, is required before the definitive numerical analysis of transformation effects is possible.

2.3 Incremental Theory

The fundamental principles behind all finite element stress analyses are the conservation of momentum, compatibility and the constitutive relations. Conservation of momentum is sometimes called the equilibrium equation. In this, as in all other finite element algorithms we seek to solve the weak form of the equilibrium equation. That is, equilibrium is satisfied only in the sense that the forces at the nodes are in equilibrium and only at the end of each time step. The equation used is usually derived from the principle of virtual work. Stated in words this principle says that at the end of the time step some stress state exists in the body, in its deformed shape. If a set of virtual; i.e., small displacements are superimposed on this body then the internal work done by the resulting virtual; i.e., small strains is equal to the external work done by the external loads and the virtual displacements. As mentioned earlier; the stresses and virtual strain measures used must be conjugate. Their inner product must equal work. In this thesis all stresses are Cauchy stresses. The virtual strains, since they must be small, then become the common small strain relation;

$$\epsilon^{\text{virtual}} = B u^{\text{virtual}} \quad (56)$$

After some rearrangement of the terms the fundamental relation is obtained;

$$\int B^T \sigma d\Omega = \mathfrak{R} \quad (57)$$

where σ is the Cauchy stress that exists in the body at the end of the increment, B is the usual derivative operator, evaluated in the geometry existing at the end of the increment and $\mathfrak{R}$ is the vector of nodal loads on the body in the deformed state at the end of the increment due to tractions and body forces.

A few words of clarification are necessary. The virtual strains are, by definition, small. They are not the strains which occurred as the body deformed during the time step. Those strains, which may be very large, are only used in the evaluation of the constitutive relation, eqn 21. It will be seen that the algorithm also calculates pseudo-loads due to initial stresses or thermal strains. Loads like this have no place in $\mathfrak{R}$. The force vector $\mathfrak{R}$ includes only real forces. Real forces are those surface and body forces which are applied to the body, and are capable of accelerating it; i.e., changing the momentum of the body.

In any iteration there exists some guess as to the deformed shape and thus the displacements u' which occurred during the time step. For those displacements eqn 21 could

be used to determine the current guess as to the stresses σ' existing at the end of the time step. Naturally, in a nonlinear analysis, the current guess σ' usually does not satisfy eqn 57. Let us assume that the stress which does satisfy eqn 57 is related by;

$$\sigma = \sigma' + \sigma^{corr} \quad (58)$$

That is, the correct stress σ is the sum of the current guess σ' and a correction σ^{corr} . Equation 57 becomes;

$$\int B^T \sigma^{corr} d\Omega = \mathfrak{R} - \int B^T \sigma' d\Omega \quad (59)$$

The residual Ψ is commonly defined as;

$$\Psi = \mathfrak{R} - \int B^T \sigma' d\Omega \quad (60)$$

This residual represents the out-of-balance nodal forces for the body with stress state σ' and loads $\mathfrak{R}$. Let us further assume that the correction σ^{corr} may be related to a correction in the strain ϵ^{corr} by;

$$\sigma^{corr} = D\epsilon^{corr} \quad (61)$$

and that ϵ^{corr} may be related to a correction in the displacement u^{corr} by;

$$\epsilon^{corr} = Bu^{corr} \quad (62)$$

Thus eqn 48 becomes;

$$\int B^T DBd\Omega u^{corr} = \mathfrak{R} - \int B^T \sigma' d\Omega \quad (63)$$

or in other words;

$$Ku^{corr} = \Psi \quad (64)$$

This process is very similar to the algorithm of Patel^[7] in which an elastic prediction of the displacements was made. From this, strains were computed. Stresses corrected for plasticity were computed and equilibrium checked. If the residual was too large then another iteration sought to correct the difference by computing a correction to the displacements. If plasticity had occurred in an element then an elasto-plastic constitutive matrix D_{ep} was used to compute a new elasto-plastic stiffness matrix K_{ep} for that element. The algorithm proposed here is not very different. An elastic, small strain prediction of the displacements is made. The strain associated with that step is calculated using the centred strain. This

properly calculates the strains associated with a step involving large displacements and large strains; i.e., a finite deformation. The stress increment is calculated using the constitutive relation and taking into account the rotations as described above. Corrections to the final stress due to plasticity are computed using the Elastic-Predictor, Radial-Corrector algorithm used by Patel^[7], in the initial orientation, before being rotated on to the final orientation. Residuals are calculated with eqn 60, using the current guess of the stress at the end of the step and the current guess of the geometry at the end of the step. The 3 norms used by Patel^[7] are computed. Any or all of these norms can be used to define acceptable convergence. The 3 norms correspond to an energy norm, a force norm and a displacement norm. If the norms are larger than preset limits, then a correction to the displacements is computed and the process repeated.

Equation 64 uses the residual calculated in the current estimate of the geometry at the end of the step to compute corrections to the geometry at the end of the step. The residual was calculated using the current estimates of the stresses at the end of the step and the geometry at the end of the step. Therefore, it would seem that all quantities being used to calculate K should be obtained from conditions existing at the end of the step. Thus the matrix D in eqn 63 is not necessarily the same as D_{avg} , used to evaluate the stress increment in the constitutive relation, eqn 21. Here D has the material properties that exist at the end of the step. As in Patel^[7], at the start of any time step the body is assumed to behave elastically. If stresses exceed the yield point in a particular element then the stiffness matrix for that element is reformed assuming an elasto-plastic response and using D_{ep} . The differences between D and D_{ep} will not be outlined here. A more detailed description of the differences may be obtained in Patel^[7] and Bathe^[21]. This new K is used in subsequent iterations.

The B matrix and thus K should probably be re-evaluated in each iteration since the geometry may change significantly. This would correspond to a Full Newton-Raphson solution. The alternative, a Modified Newton-Raphson solution would evaluate the stiffness matrix K once, with the initial geometry. Full Newton-Raphson techniques are usually more costly since they require a new stiffness matrix to be calculated at every iteration. They can also cause numerical instabilities^[70] in nonlinear problems possessing extrema or limit points. When the incremental displacements are very large it may be necessary to use a Full Newton-Raphson iteration scheme. Modified Newton-Raphson schemes are generally

cheaper and the benefits of line searches are more pronounced for Modified Newton-Raphson schemes than for Full Newton-Raphson schemes^{[38],[39]}.

Initially, it was believed that a Full Newton-Raphson scheme would be necessary for all finite deformation problems. It eventually became apparent that for many problems the deformation was large enough to require the use of the finite deformation algorithm, but not so large that the Full Newton-Raphson scheme was required. As in Patel^[7] some alterations have been made to the Modified Newton-Raphson scheme to facilitate convergence in the presence of plasticity. For finite deformation problems the element stiffness matrices are evaluated at the beginning of every increment. When using a Modified Newton-Raphson scheme the element stiffness matrices are re-evaluated during an increment only when the element response becomes elasto-plastic.

As expected, using a Modified Newton-Raphson scheme retarded convergence somewhat in some tests. For most problems the Modified Newton-Raphson scheme was sufficient. For the tests examined, the Full Newton-Raphson scheme took approximately 20% longer than the Modified Newton-Raphson without substantially reducing the number of iterations. Of course, this saving will depend on the problem being analysed. Some problems involving extremely large deformations cannot be solved without using the Full Newton-Raphson scheme. As will be seen using a Modified Newton-Raphson scheme lead to a finite deformation algorithm that was little more expensive to use than the existing small deformation algorithm.

The first iteration of any increment represents a special case where some small modifications need to be made to the algorithm outlined above. It might appear that no guess as to the geometry at the end of the increment exists. In fact one does. The only available description of the body is the one which exists at the beginning of the step. If it were assumed that this were an estimate, however inaccurate, of the geometry at the end of the step then the algorithm outlined above could be applied. In this case the current estimate of σ' , the stress at the end of the increment is again;

$$\sigma' = \sigma_{(n)} + \Delta\sigma \quad (65)$$

Assuming that the initial geometry is some approximation to the final geometry, is identical to assuming that there are no displacements during the increment and, therefore;

$$\Delta\sigma = -D_{avg}(\Delta\epsilon^{Th} + \Delta\epsilon^{Tr}) + (D_{(n+1)} - D_{(n)})D_n^{-1}\sigma_{(n)} \quad (66)$$

and again;

$$K u^{corr} = \Psi \quad (67)$$

but now, in the first iteration;

$$\Psi = \mathfrak{R} - \int B^T \sigma_{(n)} d\Omega + \int B^T (D_{avg}(\epsilon^{Th} + \epsilon^{Tr}) + (D_{(n+1)} - D_{(n)}) D_n^{-1} \sigma_{(n)}) d\Omega \quad (68)$$

Thus the initial stress becomes an initial stress pseudo-load;

$$\mathfrak{R}_I = \int B^T \sigma_{(n)} d\Omega \quad (69)$$

and the thermal and transformation strains become a thermal strain pseudo-load;

$$\mathfrak{R}_{Th} = \int B^T (D_{avg}(\Delta \epsilon^{Th} + \Delta \epsilon^{Tr}) + (D_{(n+1)} - D_{(n)}) D_n^{-1} \sigma_{(n)}) d\Omega \quad (70)$$

These pseudo-loads are applied only in the first iteration and are not included in the equilibrium equation, eqn 57, or the calculation of the residual at the end of any iteration, eqn 60. They do not appear as part of $\mathfrak{R}$ because at that point they are already present, hidden in σ' . It might appear that the first iteration of an increment was being treated differently from any other iteration. As has been shown this is really not the case.

3. Validation

An important part of the description of a new algorithm is the proof that it correctly predicts the phenomena being modelled. Proofs of this sort usually take two forms. Either a comparison of the numerical results is made with some analytical solution or a comparison to experimental results is made. As will be seen these two different types of tests are not equivalent. In fact they actually test different things. Both have limitations which must be fully understood before the comparison is made.

Comparisons of numerical results to analytical solutions test the accuracy and correctness of the algorithm. They test the ability of the algorithm to correctly solve the mathematics that it was designed to model. Analytical solutions, by their very nature are devoted to the solution of specific, fairly simple problems. Even so, they always include assumptions. These assumptions and their consequences are sometimes very subtle and often not expressly stated. The analytical solutions used for comparison should be complete solutions of the problem and not approximations themselves. Thus for stress analysis, solutions based on simple beam theory should be discarded in favour of stress function solutions which do satisfy equilibrium, compatibility and the stress-strain laws. Finding such solutions can be more difficult than it first appears. Many so-called exact solutions involve hidden approximations which are only apparent after a thorough examination of the solution method. Examples of some approximations are: solutions which are valid in the near field only; e.g. around a crack tip; solutions valid only at some distance from a discontinuity such as a point load; boundary conditions which are impossible to achieve in reality such as perfectly sharp re-entrant corners which do not create infinite stresses. Thus, it is almost always impossible to speak of an exact solution. Analytical solutions are exact only to the extent that their underlying assumptions are true. In spite of this analytical solutions still provide the best test that a program is functioning correctly.

The purpose behind creating computer programs has always been to analyse or predict some complex physical phenomenon. Ultimately the numerical results must be compared against experimental results. If the program does not agree with reality then either the program is not functioning properly or the underlying theory fails to capture some significant aspect of the physics. This would seem to place a heavy prejudice in favour of comparing the numerical results to experimental results. Great care must be taken however. It is not

always appropriate or easy to make such a comparison. Experimental evidence also has flaws, most of which are never discussed.

Let us deal first with the circumstances which make comparison to experimental results difficult. Experimental results on appropriate simple tests are not easily available. When they are, descriptions of the experimental conditions are often not sufficiently precise to allow accurate simulation. Constitutive properties, especially their dependence on temperature, are almost always unavailable for the materials used in the experiments. The properties of similar materials may be available but the differences may be important. Some experiments, like those of Hibbitt^[1] are special cases where additional exotic phenomenon dominate.

When comparisons between experimental and numerical results are made it is frequently assumed that any differences are caused solely by inadequacies in the numerical algorithm. Experimental results are assumed to be exact measurements. This is neither reasonable nor completely true. Experimental results are, in fact, always approximations to the real results. As such, they should be treated with the same skepticism applied to numerical results. Differences may be due to flaws in either the experimental or numerical technique.

Experimental error has many sources. Errors associated with the actual measurement are usually very small. It is by no means the only source of errors present. In a weld it is very difficult to control the experimental conditions. Welding speed, arc voltage and current are not constant but vary slightly. They are seldom measured. In addition, arc efficiency is never known exactly. Arc efficiency, and its variability depend on the process used. For manual arc welding Christensen^[71] showed that arc efficiency can vary by $\pm 10\%$. TIG welds have arc efficiencies that can vary by $\pm 20\%$. Submerged arc efficiencies may vary by $\pm 5\%$. Thus the heat input is not known precisely and probably also varies during the test. It will, therefore, be extremely difficult to produce two identical welds. Furthermore, there will always be differences between the desired experimental conditions and the actual conditions.

Experimental techniques always contain assumptions. The extent to which these assumptions are true has a direct bearing on the validity of the experimental results. Often these assumptions are very subtle. For example Ueda's^{[9],[20],[21]} measurements of residual

stresses in welds assume that two of the shear stresses are zero everywhere. Theoretical and experimental evidence outlined above makes this unlikely. The consequences of this assumption are unknown. Residual stress measurement techniques involving the removal of material assume that an elastic unloading occurs. Mitter^[72] has shown that this is not always the case. Assuming elastic unloading may not be valid when measuring stresses of yield point magnitude in objects that have experienced significant plasticity and kinematic work hardening. The presence of several phases with different yield strengths will be even more likely to experience inelastic unloading.

Making accurate measurements of weld related phenomena is also very difficult. When measuring thermal histories, errors in location of as little as 0.5 mm can lead to noticeable differences in the thermal history. Control of the heat source location or thermocouple location is rarely possible to this degree. When examining experimental results, it is necessary to bear in mind the variation to be expected for measurements of supposedly identical quantities, in supposedly identical welds. If the numerical results fall within these limits then the numerical prediction is as good as it is possible to measure. Agreement must then be considered as being exact. Sometimes more than one experimental technique is employed to measure the residual stress field in a single weld^[73]. These tests are particularly interesting. A significant difference often exists between the sets of experimental results. Experimental errors should probably be assumed to be at least as large as the difference between the two sets of results. The difference in the experimental results also proves the fallacy in expecting an exact match between numerical and experimental results.

Acceptable experimental results are difficult to find. Before comparisons are made to experimental results it is necessary first to prove that the algorithm correctly solves the kind of problems it has been designed for. Whether or not the mathematics is an accurate description of the physics is the second test. In this section the algorithm will be tested on three simple analytical test cases. The three are: a rigid body rotation, a hyperelastic tensile test, and a thermo-elastic deformation test. A fourth test to demonstrate the effect of transformation plasticity will also be included.

3.1 Rigid Body Rotation Test

Rigid body rotations and translations do not cause stresses. As has been outlined earlier, small deformation strain tensors cause spurious strains to appear whenever rigid

body rotations occur. These spurious strains may destroy the accuracy of any solution. In this section all comparisons will be made on the basis of the magnitude of the spurious stress which appears as a result of the presence of the spurious strains. This will give a better idea of the seriousness of the errors. As mentioned earlier in a steel body, small deformation theory causes a spurious stress on the order of 30 MPa for a 1° rigid body rotation.

Table 1. Comparison of Small and Finite Deformation
Stresses for a 10 Degree Rigid Body Rotation

<i>Deformation</i>	σ_{xx} Pa	σ_{yy} Pa	σ_{zz} Pa	σ_{xy} Pa	σ_{yz} Pa	σ_{zx} Pa
<i>Theory</i>						
<i>Small</i>	-6.0×10^9	-6.0×10^9	-3.6×10^9	1.4×10^{-7}	1.1×10^{-6}	1.7×10^2
<i>Finite</i>	1.0×10^3	-1.6×10^2	2.7×10^2	1.1×10^2	1.5×10^1	1.7×10^2

An extremely simple test would be to rotate a single element and compare the stress state obtained using small deformation theory to that obtained with the finite deformation algorithm described above. It will be recalled that the finite deformation theory proposed above assumed that points move in straight lines and that strains and stresses may be obtained from the one point integration of their respective rates. The assumption of straight line motion is not true during a rotation. It will be particularly interesting to see what effect this has.

For this test a 1 m cube of steel was selected, Young's modulus of 2.0×10^{11} Pa and Poisson's ratio of 0.3 were selected as material properties. Figure 2 shows the element before and after the rotation. Plastic deformation was not permitted by selecting a large value for the yield stress. A rigid body rotation about one axis was applied by prescribing the nodal displacements on a single 8-node brick element. A 10° rotation about the z axis was

Small Deformation

$$\sigma_{zz} = -6.0 \times 10^9 \text{ Pa}$$

Large Deformation

$$\sigma_{zz} = 1 \times 10^3 \text{ Pa}$$

Figure 2. 10° Rigid Body Rotation.

selected. Very large rotations must be avoided in the finite deformation theory proposed here. However, 10° was felt to be a moderate rotation. This test will demonstrate the inability of small deformation theory to deal with rigid body rotations as well as give some idea as to the seriousness of the assumption of straight line motion in the finite deformation theory.

The results of this test were very encouraging. The typical Gauss point stresses obtained may be seen in Table 1. In the small deformation case all Gauss points had identical stresses. The finite deformation stresses tended to vary slightly from one Gauss point to another. In the small deformation test, very large spurious normal stresses of approximately 6 GPa appeared when the cube was rotated 10° . The finite deformation rigid body rotation test resulted in spurious stresses always less than 1 KPa. A 10° rotation would seem to be acceptable in finite deformation analyses.

3.2 Hyperelastic Tensile Test

A proper demonstration of the finite deformation capability should include a test where deformations and strains are large. Various analytical solutions of such problems exist. As outlined above it is very important to fully understand all the assumptions inherent in the analytical solution. The more complicated the problem is, the more difficult understanding the solution is. Ensuring that the mesh is sufficiently fine to completely capture the displacement field and that element distortion is kept low are important considerations. If either of these two problems appear then the numerical solution will lose accuracy not because of any weakness in the algorithm but merely because the mesh does not possess interpolation functions capable of adequately approximating the displacement field. Element distortion is particularly insidious in large deformation test cases. When it occurs, initially acceptable meshes degrade and become inaccurate as the deformation proceeds. Automatic remeshing algorithms are usually required when this happens. It would be an advantage if the problem selected were relatively simple and could be modelled with a small number of elements.

The finite deformation test case selected here is that of a hyperelastic uniaxial tensile test. Using the true strain equation and an elastic material it can be shown that the

relationship between force F and length l is;

$$F = A_0 E \left(\frac{l_0}{l} \right)^{2\nu} \ln(l/l_0) \quad (71)$$

where; A_0 is the original cross-sectional area, l_0 is the original length, E is Young's Modulus, and ν is Poisson's ratio. The derivation is trivial and will not be shown here. This is a geometrically nonlinear test case which involves both large deformations and large strains. This test is particularly attractive because its simple displacement field can be captured exactly by one 8-node brick. In addition, no matter how large the deformation becomes, the interpolation functions in the physical domain will still be able to capture the displacement field exactly.

In this problem, the stresses will continue to increase as the elongation proceeds. At some point however, the cross-sectional area starts to decrease faster than the stresses increase. The force will be maximum at this point, and decline for subsequent elongation. An elastic instability exists in this problem in that applied forces greater than this maximum will result in infinite elongations. This allows two numerical simulations of this test to be performed. The first is the numerical prediction of the force for prescribed elongations. The second is the prediction of the elongation for applied forces less than the maximum value. Both these tests were performed on an initially 1 m cube of steel. Young's Modulus and Poisson's ratio were again 2.0×10^{11} Pa and 0.3 respectively. For this material it can be shown that the maximum force occurs when the length is approximately 5.3 times the original length. The analytical solution of this problem and numerical results for prescribed displacements and prescribed forces are all shown in Figure 3. A convergence tolerance of 1% on the energy norm was used in both tests. The shape change occurring in each step is very large. During each step the stiffness matrix will also change a great deal. Thus use of a full Newton-Raphson iteration scheme might have been more appropriate. The results presented for this test were obtained using a Modified Newton-Raphson scheme with line search. Crisfield's^{[38],[39]} line search algorithm was necessary in both tests.

The results for the numerical simulation with prescribed deflections are in excellent agreement with the analytical solution. Prescribed elongations of up to 8 times the original length were applied. Transverse deflections due to the Poisson effect and reactions were calculated by the finite element program. With Crisfield's^{[38],[39]} line search all increments except the first, took only two iterations to converge. The first increment took 3 iterations

Figure 3. Hyperelastic Tensile Test, Force vs Elongation.

to converge. The largest difference between the analytical and numerical solutions occurs in the first step where the incremental strain is the largest. There, even a 100% elongation in one step has not caused excessive errors to appear.

In the applied force test it is again apparent that excellent agreement with the analytical solution was obtained. It was not possible to accurately predict the deflection at the maximum force. This is not surprising. At the maxima the slope of the force/deflection curve is zero. Thus terms in the stiffness matrix relating longitudinal deflections to longitudinal forces are approaching zero. Numerical results are therefore increasingly inaccurate and the solution becomes unstable as the maxima is approached. In this test, all increments converged in 2 iterations.

3.3 Thermal-Elastic Deformation Test

Since this work is devoted to thermal stress analysis it is necessary to perform a large deformation thermal stress test. Great care must be taken when selecting the test in order to understand any hidden assumptions. As will be seen, these assumptions are insidious. In order to validate finite deformation formulations a common test is that of a very long, slender beam subjected to a constant bending moment^{[74],[75]}. A test very similar to this is used here. Assumptions implicit in the derivation of these analytical solutions limit them to large deformation, small strain cases. Answers may be obtained for large strain situations but there will be some error associated with this.

In this test a long, slender cantilever beam has been subjected to a temperature field which varies linearly through the depth of the beam. Plane sections are assumed to remain plane. This assumption is only strictly true for cases where the deformation corresponds to that of pure bending^[76]. These conditions were satisfied for this problem. With such a temperature field, linear through the depth and constant along the length, the beam bends into an arc of constant curvature. At this point it is usually assumed that the strain varies linearly through the depth. This assumption is only true for small strains. It is true that the stretching or compression of the fibres varies linearly if plane sections are to remain planar. However, it should be recalled that the true strain is not a linear function of the elongation. This is not a fatal problem however. A large deformation, small strain test is still useful. This is probably why tests such as this typically select beams that have lengths 50 to 1000 times the beam depth. In the thermal stress test used here, if the total strain

is linear through the depth then it follows that the stresses will be zero. The geometry of this type of test is shown in figure 4. The angle of rotation for a small strain case with a linear temperature field may be obtained from Roark^[77];

$$\Theta = \frac{\alpha y}{d}(T_2 - T_1) \quad (72)$$

where y is the distance from the fixed end, α is the average coefficient of thermal expansion, d is the depth of the beam and T_1 and T_2 are the temperatures on the left and right edges. If the beam bends into a circular arc then the radius of curvature r is;

$$r = \frac{d}{\alpha(T_2 - T_1)} \quad (73)$$

The horizontal displacement of the free end Δx is;

$$\Delta x = r(\cos(l/r) - 1) \quad (74)$$

and the vertical displacement of the free end Δy is;

$$\Delta y = l - r(\sin(l/r)) \quad (75)$$

Figure 4 shows the mesh used to test this case. A steel beam $10 \text{ mm} \times 1 \text{ mm} \times 1 \text{ mm}$ of initially uniform temperature is subjected to a temperature field which varied linearly from one edge to the other. For this beam Young's modulus was $2.0 \times 10^{11} \text{ Pa}$, Poisson's ratio was 0.3 and the average coefficient of thermal expansion α was 1.463×10^{-6} . Initially the beam was at a uniform 0°C . The left-hand edge was cooled in 200°C steps for 10 steps while the right hand edge was heated in 200°C steps for 10 steps. At the end of each step a linear temperature field was therefore created. Shrinkage on the left edge and expansion on the right edge will cause the beam to bend to the left. These large temperature changes are required in order to obtain large displacements for this rather short, thick beam. The free end rotates about 4° per increment. Different material properties could have been used to generate the same effects. It was decided however that a better understanding of the situation would be obtained by keeping to familiar materials. The full linear temperature field in 20-node bricks was used to model this test. The conditions analysed represent a very severe degree of bending in a relatively short thick beam.

This case represents an extremely difficult test of the finite deformation algorithm. Normally in an incremental, finite deformation test the stresses continue to increase as

Figure 4. Thermal-Elastic Bending Test.

Figure 5. Thermal-Elastic Bending Test. Displacement vs Temperature Difference.

the deformation increases. Thus errors in the stress field do not tend to accumulate as additional increments are analysed. In this problem the stress is always supposed to be zero. Increasing the through-thickness temperature difference will increase the deformation but the stresses should still be zero. Any errors in the stress will tend to accumulate, making each step worse than the last. Thus the steps must be fairly small and great care must be taken when selecting the convergence tolerance. Enforcing too strict a convergence tolerance can cause a failure to converge. Too lax a tolerance and the results will be inaccurate. In this problem it was assumed that the error in the stress would accumulate. A 1% tolerance in Patel's^[7] force norm was used for the first increment. In subsequent iterations the force norm tolerance was allowed to increase by 1% per increment.

Figure 4 shows the initial geometry and the unmagnified deformation after 10 increments. Figure 5 compares the analytical and numerical horizontal and vertical displacements of the free end. A reasonable agreement between the numerical and analytical results is evident. It would appear that the tolerance on the second increment was too lax. Thus errors in the displacement which appeared in the second increment are passed on to all subsequent increments. Incremental displacements for all subsequent steps appear to be fairly accurate. It would not appear that either excessive element distortion, or strains large enough for the small strain assumption to be invalid, have occurred.

3.4 Transformation Plasticity Test .

Testing the transformation plasticity code is an inescapable requirement. Analytical solutions for transformation plasticity problems do not exist. Unfortunately, good experimental results, including temperature dependent material properties, known temperature fields and thermal histories, transformation rates and volume changes are also not available. Lack of reliable constitutive properties is a common problem. The experiments of Jones and Alberry^[68] come very close to being good enough but are not sufficient to allow an accurate test. However, a simulation of that experiment using available data should show similar trends.

In the experiments of Jones and Alberry^[68] several 12 mm long bars of various alloys were heated until the middle of the bar was at approximately 1300°C. At this temperature the bar was rigidly fixed between two water cooled grips which extracted heat and ensured that the length could not change. Thus the temperature was not uniform down the length of

Figure 6. Transformation Plasticity Test. Axial Stress vs Temperature. The Transformation occurred between 600°C and 300°C.

the bar. The constitutive properties and transformation kinetics were quite different for the alloys chosen. As the bar cooled a tensile stress developed. This stress increased until the transformation from austenite to ferrite/cementite started, at a temperature between 600°C and 500°C depending on the alloy. During the transformation there was a dramatic drop in the stress. After the transformation was complete, at about 300°C, there was a small, compressive longitudinal stress present. Continued cooling and further shrinkage caused a rapid reappearance of a very large tensile longitudinal stress. At room temperature the resulting tensile stress was as large as it had been before the transformation for one alloy and much larger than before the transformation for another. The large differences in the responses of these alloys clearly demonstrates the impossibility of simulating these tests without accurate constitutive data. Be that as it may, all the bars which did transform exhibited a similar decrease in stress during the transformation. Alloys which did not transform did not show this change. If similar circumstances for a known material are modelled then a similar variation of stress with temperature would be expected.

In the finite element test case analysed, a bar at 1300°C was fixed between two rigid plates. The temperature dependent material properties for 0.23% carbon steel^[7] were used. For this simulation of the conditions of Jones and Alberry^[68] the transformation from austenite to ferrite/cementite was assumed to be a linear function of temperature, starting at 600°C and ending at 300°C. A 1.4% volume expansion on cooling through the transformation was assumed. The bar was cooled by prescribing the nodal temperatures at all times. In this way it became possible to impose a uniform temperature on the body. Several tests, with varying sizes of temperature step through the transformation, were performed to test the stability of the algorithm. As outlined earlier, not limiting the transformation plasticity leads to an unstable response. When the temperature steps are too large and the stress is changing rapidly, an overestimate of the transformation plasticity is made and the stresses oscillate back and forth from tension to compression. Without limiting the transformation plasticity, it is dangerous to allow temperature steps of more than about 20°C in this case.

Stress versus temperature profiles for temperature step lengths of 20°C, 50°C and 100°C through the transformation and one 300°C step across the transformation range are shown in figure 6. The stress vs temperature curve for 20°C steps through the transformation exhibits a response identical in form to the experimental results of Jones and Alberry^[68].

When the transformation starts there is a rapid reduction in the tensile stress. As the transformation ends a small compressive stress exists. As will be recalled, experimental evidence^{[64],[81]} suggests that the transformation plasticity should be a quadratic function of the fraction transformed, not a linear function as has been assumed here. Had the algorithm been designed with such a relation then the reduction in the tensile stress would have occurred even more rapidly. The linear dependence of the volume change on the fraction transformed might then have resulted in a larger compressive stress at the end of the transformation. In this test there was sufficient shrinkage after the transformation is complete to cause plastic deformation. The final room temperature stresses are all approximately the same because they are limited by the plastic flow stress. In other circumstances this might not be so.

The results do not exhibit the instability associated with large steps through the transformation. Errors caused by using large steps do not appear to be overly serious. Even for one step completely across the transformation range the stress at the end of the transformation does not fatally affect the analysis.

4. Results

4.1 Comparison with Experiment

It is desirable to compare residual stress and deformation predicted numerically with that obtained experimentally. Although a reasonable amount of experimental residual stress data exists in the literature^{[1],[3],[6],[10],[16],[78]} it is necessary to proceed cautiously when comparing numerical and experimental residual stresses for two reasons. Very broadly, these reasons involve the reliability of the experimental data and the ability to adequately simulate those welds numerically.

In this work, a full 3D thermal-elasto-plastic analysis of a short edge weld will be performed. The predicted residual stresses will be compared to experimental measurements of residual stress^[78]. The experimental data were obtained by machining thin layers from the welded edge and measuring the bar deflection and strain at the remote edge caused by the removal of these stressed layers. From these data it was possible to calculate the average longitudinal stress existing in the uncut bar. This was done for two identically welded bars. By measuring strain changes and deflections in two bars it was possible to obtain 4 values for the average stress in any layer.

Measuring residual stresses in welds is extremely difficult. The stress fields change rapidly with position. Thus great accuracy is required in determining the position of the point where the stresses are being measured. Residual stress measurements that involve the use of strain gauges, only measure the average strain change in the area covered by the gauge. When the stress gradients are steep this averaging may be important. Some residual stress measurement techniques, like the hole drilling method, assume that an elastic unloading occurs around the hole as material is removed. When the stresses are very large and equal to the local yield value this is not a good assumption. X-ray or neutron diffraction measurements are perhaps most likely to give the best results.

An expensive and not necessarily definitive test of residual stress measurement techniques is to attempt to measure identical quantities several times. If the same result is obtained it proves that the method is precise but not necessarily accurate. If the method fails this test then it is not precise and can only be used with caution when measuring residual stresses. Tests of this sort are rarely done. When they are, the results are not

encouraging. Leung^[10] reported longitudinal stresses on both sides of a single pass butt weld using an X-ray diffraction technique. A difference of approximately 100 MPa was found between stresses measured on opposite sides of the weld in an areas where average longitudinal stress was approximately 200 MPa. Since the stress field was supposed to be symmetric either the measured stresses contain substantial errors or some unknown, asymmetric, condition existed in the experimental weld. Rybicki^[5] measured residual stresses in multipass pipe welds and also found considerable variation in the stresses obtained. For supposedly identical stresses of approximately 200 MPa, the range of experimental values was frequently more than 100 MPa. At the weld centreline the range was over 300 MPa. No reasons for these large ranges were given. It is possible that this variation was due to a lack of radial symmetry in the restraint or transverse symmetry in the order in which the beads are laid down. This doesn't seem likely however. Certainly differences of this magnitude are completely unexpected. For unknown reasons the experimental measurement of residual stresses in welds is very difficult and would seem to be unable to produce precise results. The experimental results shown below have a greater variation in the longitudinal stresses than would be desired. Unfortunately they are typical of most published values.

- ① Real welds are often several meters in length. On both ends of any weld there will be regions where the stress field is different from that existing over the majority of the weld length. These regions have been called end transients. The difference between the thermal history of points at the ends of the weld and points in the middle of the weld is an important cause of these end transients. Differences in restraint also contribute however. The size of these regions is dependent on the weld parameters but can be quite large. In a 1 m butt weld Andersson's^[4] experimental measurements show 20 cm long zones on both ends of the weld where longitudinal and transverse normal stresses were noticeably different. In a series of 2D, in-plane analyses of butt welds Chakravarti^[80] found that the weld had to be at least 60 cm long in order for the longitudinal residual stresses to stabilize. At the current moment the cost of full 3D analysis of long welds makes it unattractive to do many tests of long welds. The adaptive, automatic mesh refinement and coarsening techniques of McDill^[18] will soon reduce this difficulty. However at this time it is not possible to analyse welds of sufficient length to guarantee that the end transients are a small portion of the total length. Thus even if reliable experimental data were available, comparisons of numerical and experimental data will be difficult unless the experiments are on short welds.

It is well known that very large distortion frequently occurs in welds. For such welds it will be absolutely essential to use a finite deformation algorithm. However, is it always required, even for welds where the deformation is moderate? A comparison of large deformation to small deformation results will not be made here because it is a pointless exercise. When the deformation is small these two analysis types give identical answers. There will always be problems where the large deformation option is not required because the deformations are indeed small. When the deformations are not small the small deformation algorithm can give wildly wrong answers. Every one of the large deformation tests in the previous section could have been run as a small deformation problem. In that case there would not have been any nonlinearity in the problem and convergence in 1 iteration to a completely wrong answer would have resulted. Using the small deformation algorithm when it is not appropriate is dangerous. There is no way for the user to know that the results are incorrect. There are no errors; no convergence problems; and no warnings. In fact using the small deformation algorithm when it is inappropriate removes the geometric nonlinearity from the problem and actually speeds convergence. Tests show that the additional expense of using the large deformation algorithm over the small is negligible. In most tests, the large deformation option increased the CPU time taken to solve the problem by less than 2% per iteration. In any case the reason for going to a large deformation analysis is not for economy but for accuracy. If the large deformation analysis is no more expensive but always more accurate then the small deformation option is not required.

A 3D large deformation analysis of a short weld was performed. Transformation phenomena; i.e., volume change and transformation plasticity, were included. The weld examined was that of a short edge weld. One symmetric half of a $10.16 \text{ cm} \times 5.08 \text{ cm} \times 1.27 \text{ cm}$ ($4.0 \text{ in} \times 2.0 \text{ in} \times 0.5 \text{ in}$) bar was analysed. Figure 7 shows the mesh used in this analysis. The back face in Figure 7 is the plane separating the symmetric halves of the edge weld. In effect, half of the heat source travels along the top, rear edge. The temperature dependent mechanical and thermal properties of an AISI 1020 carbon steel were used. These properties came from the data base built into ASGARD. The data base was originally collected from various sources by Patel^[7]. The initial temperature of the bar was 20°C . The original experimental weld was a de-focussed electron beam weld used to create a temperature field similar to that of manual arc weld. The electron beam parameters possessed a voltage of 70 kV a current of 30 mA and the welding efficiency of the source was 85 %. Thus the power

input of the welding source to the bar was 1785 Watts. The welding speed was 0.508 cm/s (12 ipm). The source was present for 10 s and travelled approximately 5.08 cm. The first two time steps were quite small, each was only 0.1 s. Following this 50 steps each 0.2 s long occurred as the source moved along the bar. The double ellipsoidal heat source model^[70] was used. In this case the dimensions chosen for the source resulted in a hemispherical source with a radius of .25 cm^[71]. Of course by modelling only half the bar, only half the hemisphere was present at any time. After the source was turned off a further 12 steps of increasing length, totalling some 20 minutes, were imposed to cool the bar until a uniform temperature was achieved. The process used in the thermal analysis of this weld is identical to that reported by Goldak^{[11],[81]}.

In the mechanical analysis, no displacements are permitted perpendicular to a plane of symmetry. Thus, as in past analyses^{[11],[81]}, displacements perpendicular to the plane of symmetry were not permitted. As a result the bar was free to exhibit the camber typical of edge welds. Convergence in any increment was assumed to occur whenever the energy norm^[7] became less than 5%.

The results from this test may be found in figures 8 to 15. The source moved in the x direction from left to right along the top, rear edge of the mesh shown in figure 7. The y direction, "up" in figure 7, lies in the plane of the bar, perpendicular to the motion of the source. The z direction is the through-thickness direction in figure 7. Since the heat source travels along the plane of symmetry, temperatures and displacements are largest on that plane. In some of the figures the bar has been rotated to show features on the plane of symmetry in the edge weld. Thus the thermal contours in figure 9 and deformed shapes in figures 11 and 12 show the most interesting features; i.e., those on the plane of symmetry. On the other hand, for clarity the stress contour plots (figures 13, 14) show the stresses on the outside face of the unrotated bar.

Figure 8 shows the isotherm pattern on the outer surface 7.6 s. after the weld has started. Figure 9 shows the temperature distribution on the plane of symmetry, also at 7.6 s. Figure 10 shows the displacement pattern as seen for the outside of the bar at 7.6 s, magnified 50 times. Figure 11 shows the displacement pattern for the same instant in time but on the plane of symmetry. Figure 12 shows the displacement pattern after the bar has completely cooled, also as seen on the plane of symmetry and also magnified 50 times. From the outside, assuming that plane sections remain plane might appear reasonable. However,

Figure 7. Finite Element Mesh used in Edge Weld Analysis

CONTOUR VALUES

A = $2.00E+02$ B = $4.00E+02$ C = $6.00E+02$ D = $8.00E+02$ E = $1.00E+03$ F = $1.20E+03$ G = $1.40E+03$ H = $1.60E+03$

Weld Direction

Figure 8. Temperature Distribution Around Heat Source, Time = 7.6 s; Edge Weld, Outer Surface. Contours in °C.

CONTOUR VALUES

A= 2.00E+02

B= 4.00E+02

C= 6.00E+02

D= 8.00E+02

E= 1.00E+03

F= 1.20E+03

G= 1.40E+03

H= 1.60E+03

Figure 9. Temperature Distribution Around Heat Source, Time = 7.6 s; Edge Weld, Symmetric Plane. Contours in °C.

Figure 10. Deformation Around Heat Source, Time = 7.6 s; Edge Weld, Outer Surface.

Figure 11. Deformation Around Heat Source, Time = 7.6 s; Edge Weld, Symmetric Plane.

Figure 12. Permanent Deformation pattern after complete cooling; Edge Weld, Symmetric Plane.

CONTOURS OF XX-STRESS

MIN = -0.473E+09 IN ELEMENT 477.
 MAX = 0.870E+09 IN ELEMENT 69

CONTOUR VALUES

A = 0.00E+00
 B = 1.00E+08
 C = 2.00E+08
 D = 3.00E+08

Figure 13. Longitudinal Residual Stress (σ_{xx}), Edge Weld. Contours in Pa.

TIME = 0.10000E+21
 CONTOURS OF XY-STRESS
 MIN = -0.153E+09 IN ELEMENT 425
 MAX = 0.940E+08 IN ELEMENT 29

CONTOUR VALUES

A = -2.00E+08
 B = -1.50E+08
 C = -1.00E+08
 D = -5.00E+07
 E = 0.00E+00
 F = 5.00E+07
 G = 1.00E+08
 H = 1.50E+08
 I = 2.00E+08

Figure 14. Residual Shear Stress σ_{xy} , Edge Weld. Contours in Pa.

Figure 15. Experimental and Numerical Longitudinal Residual Stress σ_{xx} , Edge Weld.

displacements occurring inside the bar makes this assumption unrealistic.

Depicting the residual stress field in a readily comprehensible fashion for a 3D body is quite difficult. Each of the 6 independent members of the stress tensor changes with position. Contour plots of surface values for some stresses are presented. For figures 13 and 14 the source travelled along the top edge of the bar. Although all six independent members of the stress tensor are calculated only two will be shown for this bar. Figure 13 shows the longitudinal (σ_{xx}) residual stress pattern on the outside of the bar after it has completely cooled. Figure 14 shows one of the shear stresses neglected in 2D analyses of welds. Figure 14 contains the (σ_{xy}) residual shear stress pattern on the outside of the bar after it has completely cooled.

Before discussing the stress contours it is necessary to comment on the nature of stress contour plots. It is well known that displacement based FEM always produces continuous displacement fields from one element to the next. However the stress field produced is frequently discontinuous even though the principles of continuum mechanics require that it be continuous. This discontinuous stress field is the result of the division into discrete elements. These discrete elements each possess polynomial interpolation functions of finite degree to describe the variation in the displacements within the element. The displacement field is guaranteed to be continuous from one element to the next. The derivatives of the displacements, from which strains and stresses are calculated, are not necessarily continuous. The displacement field is said to be C^0 continuous.

This discontinuous stress field will be especially noticeable when stresses are changing rapidly and meshes are coarse. Most graphics programs which produce stress contours take the Gauss point stresses and convert them to nodal stresses using some sort of projection and smoothing algorithm. These smoothing algorithms usually have no rigorous basis in continuum mechanics. They are, in fact, merely a filter which makes the stress field easier to visualize. The least-squares projection method of Patel[7] was used here. Unfortunately the subelements contained in the 8- to 26-node linear grading bricks^{[15],[40]} used here are not ideally suited to the available stress smoothing algorithms. The contours produced are sometimes slightly discontinuous. This is not the fault of the analysis, but rather the fault of the graphics programs used.

Another difficulty with stress contours sometimes arises when the mesh is too coarse.

On free surfaces and planes of symmetry some members of the stress tensor must be zero. The stress projection algorithm takes Gauss point stresses and projects them to the nodes where the contributions from adjacent elements are summed. Gauss points are not on these free surfaces and planes of symmetry and their stresses may not be zero. The projection to the nodes sometimes produces stress contours that are not zero on these special surfaces. This can easily occur when the mesh is rather coarse.

In figure 13 it is evident that the end transients for stresses are very large. The weld examined may or may not have been long enough. A very short region, slightly to the right of the midlength may have attained or nearly attained the quasi-stationary steady state. The familiar pattern of tensile stresses greater than yield point magnitude near the weld is evident. Moving away from the centreline, the longitudinal stresses drop rapidly to become mildly compressive. At the bottom edge of the bar, the longitudinal stresses are tensile because of the bending which has occurred.

Figure 14 shows the residual σ_{xy} shear stress. This is one of the shear stresses that all 2D plane strain analyses presume is zero. The shear stress levels are large enough to be significant. The largest shear stresses are small compared to the largest longitudinal stresses. However, in the regions where the shear stress is large, it is between a quarter and a third of the local longitudinal stress. They will have an effect on the plasticity in that region. The presence of shear stresses of such magnitude raises some questions as to how appropriate the plane strain assumptions are.

Figure 15 compares the average longitudinal stress measured experimentally with the numerical predictions. Average longitudinal stresses are plotted against the distance from the weld. The experimental results are shown as two lines. One line represents the largest value measured at each point, the other represents the smallest value. It would be hoped that the range between these limits would not be large and that the FEM results would lie somewhere between them. The range between the limits is not as small as one would like. The magnitude and distribution of the stresses as predicted by the numerical process is in reasonable agreement with that measured experimentally. A rapid drop in stress very close to the fusion zone and subsequent elevation a short distance further out is present in both the experimental limits and the numerical results. This is very interesting. Classical stress analysis of welds contains no theoretical justification for such a phenomenon. It is tempting to suggest that its existence is due to some transformation effect.

The lack of agreement between the numerical and experimental results may have several sources. First, the experimental technique only measures the average stress in a layer. This is not particularly accurate when stresses change rapidly. The technique itself, where thin layers are removed, requires great control to remove a layer of even thickness along the full length of the curve. The numerical analysis has some bad points too. The stresses do appear to change rapidly over very small distances. To capture this it would seem that a finer mesh should have been used. Perhaps several elements are required across the heat-affected zone. It is also possible that the numerical simulation took place on too short a bar. In that case the differences shown may only be the difference between the end-transient and quasi-stationary zone. The approximations inherent in modelling the transformation plasticity may also be causing the differences. Clearly better experimental and numerical results are desirable.

4.2 Butt Weld

Even without a comparison to experimental data the algorithm described here can still be used as a tool to examine the nature of residual stresses in welds. Within the framework of the assumptions inherent in the underlying mathematics, and bearing in mind the types of phenomena included, it is possible to assess the importance of various factors and assumptions frequently made in the analysis of welds.

One key assumption made by the welding community is that a state of plane strain exists in any weld. Occasionally the generalized plane strain condition^{[10],[20]} is used. In the plane strain state, some deformation modes are suppressed and some shear stresses are forced to be zero. If, in a 3D analysis, the displacement patterns do not agree with the full plane strain condition or the generalized plane strain condition then serious doubts as to the validity of those assumptions will exist. Furthermore, if these other shear stress components are not zero then their relative magnitude will be important in assessing how accurately plasticity is predicted in these 2D analyses. We have seen that these assumptions are not necessarily valid for edge welds. Edge welds do not occupy a significant place in industrial applications of welding. Butt joint welds are frequently used to join two plates together. There are reasons to believe that the approximations to the stress field inherent in 2D analyses are even less valid for butt joint welds. With very little effort it is possible to repeat the edge weld analysed above, and with some changes in the boundary conditions,

obtain the residual stresses around a butt joint weld.

The thermal conditions outlined earlier were reapplied to the mesh in figure 7. For ease of viewing the mesh has been rotated to expose the top surface; the surface upon which the heat source is applied in a butt weld. This view is shown in figure 16. The boundary conditions for the stress analysis were chosen to correspond to that of half a butt weld. In this case, the plane of symmetry becomes the near edge of the bar shown in figure 16. Again displacements perpendicular to the plane of symmetry are not permitted. All other modes of deformation were permitted.

A 3D large deformation analysis of a short butt weld was performed. Transformation phenomena; i.e., volume change and transformation plasticity, were included. The weld examined was one symmetric half of a $10.16\text{ cm} \times 10.1\text{ cm} \times 0.635\text{ cm}$ ($4.0\text{ in} \times 4.0\text{ in} \times 0.25\text{ in}$) plate was analysed. In effect, half of the heat source travels along the top, front edge. The temperature dependent mechanical and thermal properties of an AISI 1020 carbon steel were used. The initial temperature of the bar was 20°C . As before, the power input of the welding source to the bar was 1785 Watts. The welding speed was 0.508 cm/s (12 ipm). The source was present for 10 s and again travelled approximately 5.08 cm . The first two time steps were quite small, each was only 0.1 s. Following this 50 steps each 0.2 s long occurred as the source moved along the bar. The double ellipsoidal heat source model^[79] was used. In this case the dimensions chosen for the source resulted in a hemispherical source with a radius of $.25\text{ cm}$ ^[71]. Of course by modelling only half the plate, only half the hemisphere was present at any time. After the source was turned off a further 12 steps of increasing length, totalling some 20 minutes, were imposed to cool the bar until a uniform temperature was achieved. The process used in the thermal analysis of this weld is identical to that reported by Goldak^{[11],[81]}. The thermal history of this butt weld was identical to the thermal history of the edge weld mentioned earlier. Only the boundary conditions for the stress analysis were different. As before the convergence tolerance on the energy norm was 5%.

The results of the butt weld analysis can be seen in figures 17 to 26. Figures 17 and 18 show the deformed shape at 7.6 s and the permanently deformed shape respectively. In both cases the displacements have only been magnified 50 times. Figure 19 shows a contour plot of the longitudinal residual stress σ_{xx} on the top surface of the plate. Figure 20 shows the residual shear stress σ_{xy} pattern. This is one of the shear stresses that the 2D plane

Figure 16. Finite Element Mesh used in Butt Weld Analysis. Rotated to Expose Top.

strain analysis presumes to be zero. Figures 21 to 26 show the stress contours on a cross section of the plate taken at the midlength of the weld; i.e., at a distance of 5.08 cm (2.0 in) from both ends of the plate.

The displacement patterns seen in figures 17 and 18 again quite clearly demonstrate that plane sections do not remain planar. The displacement patterns measured experimentally by Johnson^[24] are clearly evident. In front of the source lines perpendicular to the path of the source bend slightly forward. After the source has passed there is substantial motion to the rear, away from the source. The deformation clearly shows that the assumption of plane strain is invalid, and that plane sections do not remain plane. On the plane of symmetry it is evident that the rearward shift is not uniform through the thickness. The well known thickening of the plate near the weld is also evident. The actual path followed by any point must be fairly complicated. In order for the thickening to be uniform along the length of the weld the additional material that ends up near the weld must have come from the volume on either side of the centreline. That is, in effect some of the material must be pulled towards the weld centreline, perpendicular to the direction of motion of the source. This motion is evident in both figures. This material that moves towards the weld also receives a substantial push in the direction opposite to the motion of the source. Very large triaxial strains and displacements must be present to create the deformation patterns observed.

Figures 17 and 18 also show clear evidence that the mesh used was too coarse. Even though lines perpendicular to the weld will move forward and back, they should still remain perpendicular to the weld centreline, where they join the centreline. That is, they should show an "S" shaped curve when seen from above. If they do not it implies that the shear stresses are not zero on the plane of symmetry. This is not possible. The recurvature required to intersect the centreline at 90 degrees might occur quite rapidly, over a very short distance near the centreline, if the stress gradients were high. This error due to poor discretization will serve to elevate some stresses and reduce others since the stresses in any FEA are always a least-squares approximation. We will return to this point when examining the stress contour plots.

The longitudinal stress pattern shown in figure 19 would seem to indicate that the weld is almost long enough to get away from the end-transients. In fact it may be slightly on the short side. The general pattern of large tensile stresses near the weld and much lower

Weld Direction

DISP. SCALE FACTOR = 0.500E+02

Figure 17. Deformation Around Heat Source, Time = 7.6 s; Butt Weld.

Weld Direction

DISP. SCALE FACTOR = 0.500E+02

Figure 18. Permanent Deformation Pattern after Complete Cooling; Butt Weld.

CONTOURS OF XX-STRESS

MIN=-0.228E+09 IN ELEMENT 396
MAX= 0.878E+09 IN ELEMENT 241

CONTOUR VALUES

A=-2.00E+08
B=-1.00E+08
C= 0.00E+00
D= 1.00E+08
E= 2.00E+08
F= 3.00E+08

Figure 19. Longitudinal Residual Stress (σ_{xx}); Butt Weld. Contours in Pa.

CONTOURS OF XY-STRESS

MIN=-0.970E+08 IN ELEMENT 211
 MAX= 0.125E+09 IN ELEMENT 225

CONTOUR VALUES

A=-5.00E+07
 B= 0.00E+00
 C= 5.00E+07
 D= 1.00E+08

Figure 20. Residual Shear Stress σ_{xy} ; Butt Weld. Contours in Pa.

CONTOURS OF XX-STRESS

MIN = -0.228E+09 IN ELEMENT 396
 MAX = 0.878E+09 IN ELEMENT 241

CONTOUR VALUES

A = -2.00E+08
 B = -1.00E+08
 C = 0.00E+00
 D = 1.00E+08
 E = 2.00E+08
 F = 3.00E+08

Figure 21. Longitudinal Residual Stress (σ_{xx}) on Cross-section at Midlength; Butt Weld.
 Contours in Pa.

CONTOURS OF YY-STRESS

MIN = -0.514E+09 IN ELEMENT 225
 MAX = 0.831E+09 IN ELEMENT 448

225
 448

CONTOUR VALUES

A = -2.00E+08
 B = -1.00E+08
 C = 0.00E+00
 D = 1.00E+08
 E = 2.00E+08
 F = 3.00E+08

Figure 22. Transverse Residual Stress (σ_{yy}) on Cross-section at Midlength; Butt Weld.
 Contours in Pa.

CONTOURS OF ZZ-STRESS

MIN=-0.582E+09 IN ELEMENT 262

MAX= 0.439E+09 IN ELEMENT 136

CONTOUR VALUES

A=-2.00E+08

B=-1.00E+08

C= 0.00E+00

D= 1.00E+08

E= 2.00E+08

F= 3.00E+08

Figure 23. Through-thickness Residual Stress (σ_{zz}) on Cross-section at Midlength; Butt Weld. Contours in Pa.

2

of/de

2

CONTOURS OF XY-STRESS

MIN=-0.970E+08 IN ELEMENT 211
 MAX= 0.125E+09 IN ELEMENT 225

CONTOUR VALUES

A=-5.00E+07
 B=-2.50E+07
 C= 0.00E+00
 D= 2.50E+07
 E= 5.00E+07

Figure 24. Residual Shear Stress (σ_{xy}) on Cross-section at Midlength; Butt Weld. Contours in Pa.

CONTOURS OF YZ-STRESS

MIN=-0.641E+08 IN ELEMENT 483
 MAX= 0.599E+08 IN ELEMENT 216

CONTOUR VALUES

A=-5.00E+07
 B=-2.50E+07
 C= 0.00E+00
 D= 2.50E+07
 E= 5.00E+07

Figure 25. Residual Shear Stress (σ_{yz}) on Cross-section at Midlength; Butt Weld. Contours in Pa.

CONTOURS OF ZX-STRESS

MIN=-0.197E+09 IN ELEMENT 478
 MAX= 0.239E+09 IN ELEMENT 441

CONTOUR VALUES

A=-5.00E+07
 B=-2.50E+07
 C= 0.00E+00
 D= 2.50E+07
 E= 5.00E+07

Figure 26. Residual Shear Stress (σ_{xz}) on Cross-section at Midlength; Butt Weld. Contours in Pa.

compressive stresses far away from the weld is evident.

The contours of σ_{xy} , shown in figure 20 are rather complex and difficult to comprehend. However, upon close inspection it may be seen that regions with moderate shear stresses exist in the material on both sides of the weld. Since there are no external loads on the plate these shear stresses are required if shrinkage at the weld centreline is to create stresses at large distances from the centreline. They are also required whenever transverse σ_{yy} stresses are present. Arguments like this suggest that σ_{xy} should be particularly intense near the ends of the weld. Figure 18 shows a small area of elevated σ_{xy} at the left-hand; i.e., starting end of the weld.

The cross-sectional, stress contour plots in figures 21 to 26 show a stress state that is more complicated than expected. At first glance some of them appear to be impossible or wrong if they are approached in a 2D frame of mind. Before dismissing them out-of-hand it is necessary to recall certain facts. The stress state shown in these cross-sections is not constant along the weld length. Figures 19 and 20 demonstrate this. This may be part of the explanation for some things like the region of tensile σ_{yy} present between regions of very low σ_{yy} . Furthermore, since the plastic deformation occurred without the presence of external loads, the deformation and stress fields may be more complex than those usually encountered.

Several possible sources of error exist in the analysis. The weld analysed may be too short to get beyond the end-transient regions. The mesh used was probably too coarse near the weld centreline. As with all numerical analyses, the discretization in time and space creates some errors. Gradients in the through thickness; i.e., z , direction are likely to be very steep, especially for σ_{zz} . This plus the admitted problems with the stress projection algorithm in regions with very high stress gradients is probably the reason why some stresses are not zero on appropriate surfaces. Convergence tolerances may have been too lax. These may be partial explanations for some of the unexpected results. A complete understanding and interpretation of the information in these figures will take much more effort.

The stress state shown in figures 19 to 24 does indicate substantial tensile longitudinal stresses near the centreline. It also shows very large compressive transverse and through-thickness normal stresses. Large differences in the temperature and plastic deformation over short distances perpendicular to the weld centreline could create sizeable through-thickness

normal stresses. This stress must be zero on the external surface but need not be zero inside. If they are present then a very fine mesh is required to capture them accurately.

Shear stresses, particularly σ_{xy} , are far from zero. In some regions they are as large as one quarter of the yield stress, and one sixth of the local normal stresses. As already seen the mesh is probably too coarse to accurately capture the shear stresses in general and σ_{xy} in particular. Nevertheless, even with the crude mesh used the shear stresses present are large enough to be significant. The actual shear stresses are probably somewhat larger than shown.

5. Conclusions

- 1) A 3D, finite deformation, thermal-elasto-plastic finite element algorithm has been developed. It has been tested by using it on several analytical solutions. The numerical results agree quite well with the analytical solutions, indicating that the algorithm is capable of solving the mathematics it was designed for. Rigid body rotations of up to 10° per increment do not create significant stresses. In practice it is recommended that rotations be kept to not more than 3 or 4 degrees per step in order that the assumption of straight line motion not be violated. Very large deformations, including either large or small strains are handled very accurately.
- 2) The importance of McDill's^{[15],[40]} 3D mesh grading hexahedrons cannot be overstated. They were the critical step that made any 3D analysis possible at all.
- 3) Transformation plasticity and volume changes associated with phase transformations, are simulated to some extent. It is not possible to definitively state how accurately these processes are simulated because of the lack of experimental data. Stress changes associated with transformations are approximately the correct magnitude, in the correct direction, and at approximately the correct time. These processes have important effects on the residual stress state in welds and are too important to ignore. In the past it has been seen that ignoring transformation plasticity can seriously reduce the accuracy of any residual stress prediction. This is especially true in high strength, heat treatable alloys. Much more work on the simulation of transformation plasticity is required.
- 4) Predictions of the longitudinal residual stresses obtained numerically are in tolerable

agreement with experimental measurements. Until the consistency of experimental residual stress measurements improves it is unlikely that any better agreement will ever be obtained. Thus numerical predictions are as accurate as it is currently possible to measure.

- 5) Deformation patterns predicted numerically, exhibit patterns identical to those measured experimentally. Around the heat source in a weld there is a general expansion due to the temperature field. Thus points ahead of the source are pushed slightly forward, in the direction of motion of the source. Behind the source, the material moves a significant distance backward.
- 6) Even though the finite element mesh used was probably slightly too coarse, it is clear that deformation patterns, stress and strain fields are strongly three dimensional. Plane strain conditions are not present and plane sections do not remain plane. The degree to which plane strain or generalized plane strain analyses can be used to predict residual stresses in welds is open to question.

6. Future Work

- 1) The welds analysed in this thesis should probably be redone with a finer mesh near the weld centreline. This would produce contour plots able to capture some of the very high stress, strain and deformation gradients present.
- 2) A great deal of effort should be expended on the development of methods to predict transformation plasticity. The current methods do not adequately predict what occurs under triaxial stresses and large time steps. The current methods are not rigorously based on the principles of continuum mechanics.
- 3) The 3D analysis of metal movement and residual stresses in welds is still too expensive to do more than examine fairly short welds. Even finer meshes than those used are required. Currently a fine although graded mesh is required everywhere that the source will ever go. Adaptive mesh refinement schemes that would allow the very fine mesh to be present only around the current location of the source are needed.
- 4) A visco-plastic material or a substructuring capability to better model high temperature material would be very useful. The analysis of residual stresses in castings will be difficult without some ability to analyse low strength materials.

- 5) Several assumptions inherent in the 2D plane strain analyses do not appear to be completely justified. However, the errors introduced by these incorrect assumptions may or may not affect the prediction of residual stresses. A comparison of the residual stresses obtained in a full 3D analysis to those obtained in a 2D plane strain analysis for identical meshes and thermal histories should be attempted. This would help determine whether or not 3D analyses are always required and what errors are introduced by assuming plane strain conditions.

7. REFERENCES

- [1] Hibbitt, H.D., Marcal, P.V., A Numerical Thermo-Mechanical Model for the Welding and Subsequent Loading of a Fabricated Structure, *Comp. & Struct.*, Vol. 3, 1973, pp. 1145-1174.
- [2] Muraki, T., Bryan, J., Masubuchi, K., Analysis of Thermal Stresses and Metal Movement During Welding, *Trans. ASME, J. Eng. Mat. Tech.*, 1975, pp. 81-91.
- [3] Andersson, B.A.B., Thermal Stresses in a Submerged-Arc Welded Joint Considering Phase Transformations, *J. Eng. & Tech., Trans. ASME*, Vol. 100, 1978, pp. 356-362.
- [4] Andersson, B., Karlsson, L., Thermal Stresses in Large Butt-Welded Plates, *J. Thermal Stresses*, Vol. 4, 1981, pp. 491-500.
- [5] Rybicki, E.F., Stonesifer, R.B., Computation of Residual Stresses due to Multipass Welds in Piping Systems, *Trans. ASME, J. Press. V. Tech.*, Vol. 101, 1979, pp. 149-154.
- [6] Argyris, J.H., Szimmat, J., Willam, K.J., Computational Aspects of Welding Stress Analysis, *Comp. Meth. Appl. Mech. Eng.*, Vol. 33, 1982, pp. 635-666.
- [7] Patel, B., Thermo-Elasto-Plastic Finite Element Analysis of Deformations and Residual Stresses Due to Welds, Ph.D. Thesis, Carleton University, 1985.
- [8] Lai, C.K., Koenig, H.A., Morral, J.E., Three Dimensional Thermal Stress Analysis of a Welded Plate by the FEM, *Trans. CSME*, Vol 10., No. 3, 1986, pp. 153-165.
- [9] Ueda, Y., Kim, Y.C., Umekuni, A., Measurement of Three-Dimensional Welding Residual Stresses due to Electron Beam Welding, *Trans. JWR*, Vol. 15, No. 1, 1986, pp. 125-131.
- [10] Leung, C.K., Pick, R.J., The Use of Generalized Plane Strain Elements in the Prediction of Residual Stresses in Welded Flat Plates, *Proceedings of the Symposium on The Computer Modelling of Fabrication Processes and Constitutive Behavior of Metals*, Ottawa, 1986, pp. 563-648.
- [11] Goldak, J., Oddy, A., McDill, M., Bibby, M., House, R., Progress in Computing Residual Stress and Strain in Welds, Presented at the International Trends in Welding Research Conference, Gatlinburg, Tennessee, May, 1986. pp. 523-527.

- [12] Goldak, J., McDill, M., Oddy, A., Bibby, M., House, R., Chi, X., Computational Heat Transfer for Weld Mechanics, Presented at the International Trends in Welding Research Conference, Gatlinburg, Tennessee, May, 1986. pp. 15-19.
- [13] Oddy, A.S. The Effect of Heat Flow Approximations on the Thermal History of Edge Welds, M. Eng. Thesis, Carleton University, Ottawa, 1980.
- [14] Westby, O., Temperature Distribution in the Workpiece by Welding, Department of Metallurgy and Metals Working, The Technical University of Norway, 1968.
- [15] McDill, J.M., An Adaptive Mesh-Grading Algorithm for Three-Dimensional Finite Element Analysis, Ph.D. Thesis, Carleton University, Ottawa, 1987.
- [16] Wap, Y.L., Camber in Welded Mild Steel Sections, M. Eng. Thesis, Carleton University, Ottawa, 1979.
- [17] Hogan, S., Structural Deflections of Stainless Steel Sections due to Welding, B. Eng. Thesis, Carleton University, Ottawa, 1979.
- [18] Oddy, A.S., McDill, J.M.J., Bibby, M.J., The Significance of Longitudinal Heat Flow in Welds, Trans. CSME, Vol. 10, No. 3, 1986, pp. 129-134.
- [19] Jhaveri, P., Moffatt, W.G., Adams, C.M., The Effect of Plate Thickness and Radiation on Heat Flow in Welding and Cutting, AWS Welding Journal, Research Supplement, Jan. 1962, pp. 12a-16a.
- [20] Ueda, Y., Nakacho, K., Distributions of Welding Residual Stresses in Various Welded Joints of Thick Plates. Trans. JWRI, Vol. 15, No. 1, 1986, pp. 113-124.
- [21] Ueda, Y., Fukuda, K., Fukuda M., A Measuring Theory of Three Dimensional Residual Stresses in Long Welded Joints, Trans. JWRI, Vol. 12, No. 1, 1983, pp. 113-122.
- [22] Duncan, D.B., Geidt, W.H., Prediction of Deformations During Gas-Tungsten-Arc Stationary Welds, ASTM STP-765, 1982, pp. 185-201.
- [23] Hernandez, I.E., North, T.H., The Influence of External Local Heating in Preventing Cracking During Welding of Aluminum Alloy Sheet, AWS Welding Journal, Welding Research Supplement, March, 1984, pp. 84-90.
- [24] Johnson, L., Formation of Plastic Strains During Welding of Aluminum Alloys, AWS Welding Journal, Welding Research Supplement, 1973, pp. 298a-305a.
- [25] Allen, D.H., Computational Aspects of the Nonisothermal Classical Plasticity, Comp.

- & Struct., Vol. 15, No. 5, 1982, pp. 589-599.
- [26] Greenwood, G.W., Johnson, R.H., The Deformation of Metals Under Small Stresses During Phase Transformations, *Procs. Roy. Soc.*, Vol. 283, 1965, pp. 403-422
 - [27] De Jong, M., Rathenau, G.W., Mechanical Properties of Some Iron Alloys While Undergoing Allotropic Transformations, *Acta Met.*, Vol. 7, 1959, pp. 246-253.
 - [28] De Jong, M., Rathenau, G.W., Mechanical Properties of an Iron-Carbon Alloy During Allotropic Transformation, *Acta Met.*, Vol. 9, 1961, pp. 714-720.
 - [29] Leblond, J.B., Mottet, G., Devaux, J.C., A Theoretical and Numerical Approach to the Plastic Behaviour of Steels During Phase Transformations - I. Derivation of General Relations, *J. Mech. Phys. Solids*, Vol. 34, No.4, 1986, pp. 395-409.
 - [30] Rammerstorfer, F.G., Fischer, D.F., Mitter, W., Bathe, K.J., Snyder, M.D., On Thermo-Elastic-Plastic Analysis of Heat-Treatment Processes Including Creep and Phase Changes, *Comp. & Struct.*, Vol. 13, 1981, pp. 771-779.
 - [31] Bathe, K.J., *Finite Element Procedures in Engineering Analysis*, Englewood Cliffs, N.J., Prentice-Hall, 1982.
 - [32] Belytschko, T., Hughes, T., *Computational Analysis for Transient Analysis*, Vol. 1, *Computational Methods in Mechanics*, North-Holland, New York, 1983, pp. 22-37.
 - [33] Hallquist, J.O., NIKE2D - A Vectorized, Implicit, Finite Deformation, Finite Element Code for Analysing the Static and Dynamic Response of 2-D Solids. Lawrence Livermore National Laboratory, UCID- 19677, 1983, pp. 6-12.
 - [34] Johnson, G.C., Baumann, D.J., A Discussion of Stress Rates in Finite Deformation Problems, Sandia National Laboratories, SAND82-8821, 1982.
 - [35] Hughes, T.J.R., Belytschko, T.B., Recent Advances in Nonlinear Finite Element Methods for Structural Dynamics and Fluid Mechanics, Vol. 1, Workshop on Theoretical Foundation for Large-Scale Computations of Nonlinear Material Behavior, Northwestern University, Evanston, Ill., 1983.
 - [36] Malvern, L.E., *Introduction to the Mechanics of a Continuous Medium*, Prentice-Hall, Englewood Cliffs, New Jersey, 1969.
 - [37] Bell, J.F., Contemporary Perspectives in Finite Strain Plasticity, *Int. J. Plast.*, Vol.1, No. 1, 1985, pp. 3-27.

- [38] Crisfield, M.A., Variable Step-Lengths for Non-Linear Structural Analysis, TRRL Laboratory Report 1049, 1982.
- [39] Crisfield, M.A., Accelerated Solution Techniques and Concrete Cracking, *Comp. Meth. Appl. Mech. Eng.*, Vol. 33, 1982, pp. 582-607.
- [40] McDill, J.M., Goldak, J.A., Oddy, A.S., Bibby, M.J., Isoparametric Quadrilaterals and Hexahedrons for Mesh Grading Algorithms, *Comm. Appl. Num. Meth.*, Vol 3, 1987, pp. 155-163.
- [41] Nagtegaal, J.C., De Jong, Some Aspects of Nonisotropic Work-Hardening in Finite Deformation Plasticity, *Proceedings of the Workshop on Plasticity of Metals at Finite Strain: Theory, Experiment and Computation*, Stanford Univ. 1981, publ. by Division of Applied Mechanics, Stanford Univ., 1982, pp. 65-102.
- [42] Vóyiadjis, G.Z., Navaee, S., Large Elasto-Plastic Deformations of Work Hardening Metal Alloys, *Trans. CSME*, Vol 8, No. 2, 1984, pp. 77-83.
- [43] Hoger, A., Carlson, D.E., Determination of The Stretch and Rotation in the Polar Decomposition of the Deformation Gradient, *Quart. Appl. Math.*, 1984, pp. 113-117.
- [44] Hallquist, J.O., DYNA3D Course Notes, Fall 1982, CE 3151-12, Lawrence Livermore National Laboratory, UCID- 19899, Rev. 1, 1986, pp. 75.
- [45] Carey, G.F. Oden, J.T. 1983: *Finite Elements, A Second Course*, Vol. II, Englewood Cliffs, N.J.: Prentice-Hall, pp. 84.
- [46] Fried, I. (1972) Accuracy of Complex Finite Elements. *AIAAJ*, 10, No. 3, pp. 347-349.
- [47] Fried, I. (1973) Accuracy and Condition of Curved (Isoparametric) Finite Elements, *J. Sou. Vib.*, 31, No. 3, pp. 345-355.
- [48] Cook, R.D. 1981: *Concepts and Applications of Finite Elements*, 409 and 211, New York, N.Y.: J. Wiley and Sons, pp. 409, 211.
- [49] Turke, D.J., McNeice, G.M. (1974) Guidelines for Selecting Finite Element Grids based on an Optimization Study, *J. Comp. St.*, Vol 4, pp. 499-519.
- [50] Zienkiewicz, O.C. 1976: *The Finite Element Method*, 3rd ed., London, U.K.: McGraw-Hill, pp. 186.
- [51] Gifford, L.N. (1979) More on Distorted Isoparametric Elements, *Int. J. Num. Meth. Engng*, Vol. 14, pp. 290-291.

- [52] Nicolas V.T., Citipitioglu, E. (1977) A General Isoparametric Finite Element Program SDRC SUPERB, Comp. St., 17, No. 2, pp. 303-313.
- [53] Striklin, J.A., Ho, W.S., Richardson, E.Q., Haisler, W.E (1977) On Isoparametric vs. Linear Strain Triangular Elements, Int. J. Num. Meth. Engng, Vol. 11, pp. 1041-1043.
- [54] Backlund, J. (1978) On Isoparametric Elements, Int. J. Num. Meth. Engng., Vol. 12, pp. 731-732.
- [55] Oddy, A., Goldak, J., McDill, M., Bibby, M., A-Distortion Metric for Isoparametric Finite Elements, Submitted to CSME Transactions, Oct., 1987.
- [56] Krieg, R.D., Krieg, D.B., Accuracies of Numerical Solution Methods for the Elasticly-Perfectly Plastic Model, Trans ASME, J. Press. V. Tech., Vol. 99, No. 4, 1977, pp. 510-512.
- [57] Schreyer, H.L., Kulak, R.F., Kramer, J.M., Accurate Solutions for Elastic-Plastic Models, Trans. ASME, J. Press. V. Tech., Vol. 101, 1979, pp. 226-234.
- [58] Argyris, J.H., Vaz, L.E., Willam, K.J., Improved Solution Methods for Inelastic Rate Problems, Comp. Meth. Appl. Mech. Eng., Vol. 16, 1978, pp. 231-277.
- [59] Reed-Hill, R. E., Physical Metallurgy Principles, New York, N.Y., Van Nostrand, 1973, pp. 729-733.
- [60] Henwood, C., An Analytical Model for Computing Weld Microstructures, M. Eng. Thesis, Carleton University, Ottawa 1987.
- [61] Denis, S., Gautier, E., Simon, A., Beck, G., Stress Phase Transformation Interactions: Basic Principles; Modelization and Their Role in the Calculation of Internal Stresses, International Symposium on the Calculation of Internal Stresses in Heat Treatment of Metallic Materials, Linkoping, Sweden, Vol. 1, 1984, pp. 157-190.
- [62] Van Vlack, L.H., Materials Science for Engineers, Reading, Mass., Addison-Wesley, 1970, pp. 343-344.
- [63] Magee, C.L., Ph.D. Thesis, Carnegie-Mellon University, 1966.
- [64] Leblond, J.B., Mottet, G., Devaux, J., Devaux, J.C., Mathematical Models for Description of Anisothermal Phase Transformations in Steels and Their Consequences on the Plastic Behaviour, International Symposium on the Calculation of Internal Stresses in Heat Treatment of Metallic Materials, Linkoping, Sweden, Vol. 1, 1984, pp. 191-220.

- [65] Sjoström, S., Interactions and Constitutive Models for the Calculation of Quench Stresses in Steel. International Symposium on the Calculation of Internal Stresses in Heat Treatment of Metallic Materials, Linköping, Sweden, Vol. 1, 1984, pp. 221-246.
- [66] Leblond, J.B., Mottet, G., Devaux, J.C., A Theoretical and Numerical Approach to the Plastic Behaviour of Steels During Phase Transformations - II. Study of Classical Plasticity for Ideal-Plastic Phases, J. Mech. Phys. Solids, Vol. 34, No.4, 1986, pp. 411-432.
- [67] Darbyshire, J.M., Probert, K.S., Townshend, R.D., The Effect of Austenite Decomposition on Residual Stresses in Normalised Ferritic Steel Weldments, Residual Stresses in Welded Construction, an International Conference, London, Vol 1, 1977, pp. 241-250.
- [68] Jones, W.K.C., Alberly, P.J., A Model for Stress Accumulation in Steels During Welding, Residual Stresses in Welded Construction, an International Conference, London, Vol 1, 1977, pp. 15-26.
- [69] Dadian, M., The Contribution of Welding to the Understanding of Metallurgical Phenomena, Presented at the International Trends in Welding Research Conference, Gatlinburg, Tennessee, May, 1986.
- [70] Rice, J.R., Numerical Methods, Software and Analysis, IMSL Reference Edition, New York, N.Y., McGraw-Hill, 1983, pp 229.
- [71] Christensen, N., Davies, V. de L., Gjermundsen, K., Distribution of Temperatures in Welding, British Welding Journal, Vol. 12, 1965, pp 54-75.
- [72] Mitter, W., Rammerstorfer, F.G., Grundler, O., Weidner, G., Explanation of Discrepancies Between Calculated and Measured Residual Stresses in a Quenched Pure Iron Cylinder, International Symposium on the Calculation of Internal Stresses in Heat Treatment of Metallic Materials, Linköping, Sweden, Vol. 1, 1984, pp 118-133.
- [73] Argyris, J.H., Szimmat, J., Willam, K.J., Finite Element Analysis of Arc-Welding Processes, Chapter 1, Numerical Methods in Heat Transfer, Vol. III, Ed. by R.W. Lewis, J. Wiley and Sons Ltd, 1985, pp 30-33.
- [74] Kanok-Nukulchai, W., Taylor, R.L., Hughes, R.L., A Large Deformation Formulation for Shell Analysis by the Finite Element Method, Presented at the Symposium on Computational Methods in Nonlinear Structural and Solid Mechanics, Oct. 1980,

Washington, D.C., Pergamon Press, pp. 19-27.

- [75] Murakawa, H., Reed, K.W., Atluri S.N., Rubeñstien, R., Stability Analysis of Structures Via a New Complementary Energy Method, Presented at the Symposium on Computational Methods in Nonlinear Structural and Solid Mechanics, Oct. 1980, Washington, D.C., Pergamon Press, pp. 11-18.
- [76] Fung, Y.C., A First Course in Continuum Mechanics, Englewood Cliffs, N.J., Prentice-Hall, 1969, pp. 280-281.
- [77] Roark, R.J., Young, W.C., Formulas for Stress and Strain, Fifth Edition, McGraw-Hill, New York, N.Y., 1975, pp. 107.
- [78] Oddy, A.S., Residual Stresses in the Vicinity of a Weld, B.Eng. Thesis, Carleton University, 1979.
- [79] Goldak, J.A., Chakravarti, A.P., Bibby M.J., A New Finite Element Model for Welding Heat Sources, Met. Trans. B, Vol. 15B, 1984, pp 299-305.
- [80] Chakravarti, A.P., Malik, L.M., Goldak, J.A., Prediction of Distortion and Residual Stresses in Panel Welds, Proceedings of the Symposium on The Computer Modelling of Fabrication Processes and Constitutive Behavior of Metals, Ottawa, 1986, pp 547-561.
- [81] Goldak, J., McDill, M., Oddy, A., Bibby, M., House, R., Mashaie, M., 3D Thermo-Elasto-Plastic Analysis of Arc Welds, Proceedings of the Symposium on The Computer Modelling of Fabrication Processes and Constitutive Behavior of Metals, Ottawa, 1986, pp 529-545.

8. Appendix 1

A DISTORTION METRIC FOR ISOPARAMETRIC FINITE ELEMENTS

Alan Oddy, John Goldak, Moyra McDill, Malcolm Bibby

October 1986

Department of Mechanical and Aeronautical Engineering

Carleton University, Ottawa, Canada, K1S 5B6

ABSTRACT

Distorted isoparametric elements are frequently encountered in finite element analyses. It is well known that element distortion affects not only the interpolation functions in the physical (x, y, z) domain and convergence rates but can also reduce the accuracy of an analysis. Error indicators and bounds can't be computed until after a solution has been obtained because the accuracy of a Finite Element Analysis (FEA) depends on the problem, the governing differential equations and the mesh. A one-to-one relationship between accuracy and element distortion, therefore, does not exist. Since error measures can't be computed *a priori*, determining them is relatively expensive. It is generally accepted however, that element distortion should be minimized, except in rare circumstances such as fracture mechanics where it is carefully controlled. A parameter which measures distortion would be useful in assessing the possible performance of distorted elements or where remeshing is required. In addition, knowing when and where to remesh because of excessive distortion is critical in automatic mesh generation or rezoning programs. A measure of element distortion that could be computed *a priori* would be useful in automatic mesh generation and rezoning programs. The basic criteria which must be met by useful measures of element distortion are described. A distortion metric based on an invariant of the isoparametric transformation is proposed. The proposed metric can be applied to two- or three-dimensional elements. Values of the distortion metric obtained for typical distorted two-dimensional elements, as well as examples of how solution accuracy degrades with increased distortion are presented.

(Submitted to: CSME Transactions, October 1987)

NOMENCLATURE

C_{ij}	the right Cauchy strain
D	the distortion metric
e_{ij}	deviatoric strain tensor
h	convection coefficient
J_{ij}	Jacobian matrix
J'_{ij}	normalized Jacobian matrix
$ J $	determinant of the Jacobian matrix
J_2	second invariant of deviatoric strain tensor
l	element length
w	element width
x, y, z	coordinate in the physical domain
δ_{ij}	Kronecker delta
ϵ_{ij}	Green's strain tensor
ζ	coordinate in the computational domain
η	coordinate in the computational domain
θ	acute corner angle of a parallelogram
ξ	coordinate in the computational domain

INTRODUCTION

The great strength of isoparametric elements is their ability to approximate complex shapes. To accomplish this a computational (ξ, η, ζ) domain is created and the element stiffness matrices, etc. are obtained by a numerical integration over the volume of the element. In the computational domain these isoparametric elements have fixed shapes and interpolation functions. When the shape of the element in the physical (x, y, z) domain is not the same as it is in the computational domain the element is said to be distorted. The analysis of real problems often requires the use of a mesh composed of distorted elements. This element distortion may be the result of the mesh generation process or, as in the case of problems involving large deformations, it may be the result of the deformation which occurs during the analysis. When large deformations occur it is often necessary to create a new mesh at some stage in the analysis. One critical problem in automatic mesh regeneration and rezoning is to know where and when it is required, before errors degrade the solution. A parameter which describes the amount of element distortion and is applicable to both automatic mesh generation schemes and automatic rezoning schemes is described here.

Solutions in three function spaces must be considered. The exact solution lies in an infinite dimensional Hilbert space. The FEA solution for a mesh without distortion is in a finite dimensional subspace. The interpolation functions are piecewise polynomials. Errors can not be avoided if the interpolation functions in the physical domain cannot capture the exact solution. The interpolation functions in the physical domain for an arbitrarily distorted mesh lie in a space of rational functions that is not easily predicted beforehand. However, it is known that with distortion the polynomial obtained in the physical domain changes^{[1][2][3]}. For most problems distorted elements cause the convergence rate to decrease^{[2][4]} and the error to increase. Occasionally, element distortion produces superior performance in particular problems. Degenerate, isoparametric crack tip elements^[5] are a good example of this. In this case severe distortion creates an element function space in the physical (x, y, z) domain which contains terms dependent on the square root of the radial distance from the stress singularity. Enlarging the element while keeping the shape constant does not change the nature of the interpolation functions. The way in which the function space changes as elements are distorted is complex. For most distorted elements

the analyst has no knowledge of the nature of the interpolation functions in the physical domain.

An acceptable mesh is one which provides answers of sufficient accuracy at reasonable cost. Accuracy and error are functions of the geometry of the domain, the boundary conditions, the governing differential equations and the geometry of the finite element mesh. There is, therefore, no unique relationship between distortion and error. Element distortion is only one factor influencing accuracy. Large distortion usually, but not always, increases the level of error present. As a matter of fact, large element distortions of a particular form might be tolerable in one problem but very wrong in another. For instance, Turek^[6] showed that elements with large aspect ratios give rise to smaller errors when aligned perpendicular to the strain energy density contours i.e.; isoenergetics. If the boundary conditions and thus the isoenergetics change then the level of error present will change. In other cases, severe element distortion like that of the degenerate isoparametric crack tip element is used to create special function spaces that increase the accuracy of the solution. For a given problem it will always be possible to compute the nodal positions *a posteriori*, producing distorted elements which minimize the error in some sense. Error measures computed *a posteriori* will be more accurate in determining the error present but they will be more expensive in that a re-solution may be required. The most frequently used strategy is to seek a mesh with minimum distortion *a priori*. To apply this strategy a measure of element distortion is needed; i.e., some parameter sensitive to the element shape and thus the change in the function space in the physical (x, y, z) domain. This would alert the user to elements which are suspect, or at least not suitable for general purpose use. If the analyst cannot predict the desired form of element distortion, then minimizing the element distortion is a conservative strategy.

Empirical rules governing acceptable levels of distortion in general purpose elements are common. For example, it is frequently stated that for quadrilateral elements the aspect ratio should be less than 3:1 for accurate stress calculations or 7:1 for accurate displacements^[4]; and that corner angles should be between 45 and 135 degrees^[7]. Another rule states that the midedge nodes of 8-node, quadratic, quadrilateral elements should always be located in the middle third of the element edge^[8]. Element curvature is also known to affect the accuracy^[9]. At best, these heuristic rules are a crude measure or description of the degree of distortion in an element, at worst they could lead to a false sense of security when more

than one type of distortion is present. Clearly, an element where all types of distortion are combined will be worse than an element which has only one type of distortion. Merely satisfying the separate rules is not adequate because it neglects the coupling effect between the various types of distortion. The question then becomes; what degree of stretching is acceptable when the corner angles are not 90 degrees? It is conservative to assume and indeed highly probable that combined distortion degrades the element more than does either stretching or shearing alone. In addition to completely ignoring the coupled effects of various types of distortion, these empirical rules are awkward to apply to automatic mesh generation or rezoning schemes.

A number of measures of element distortion are available. Each, however has limitations. Nicolas et al. [10] proposed a parameter which is not able to detect distortion in any parallelogram. Using this parameter, parallelograms with aspect ratios of 1000:1 and/or corner angles of as little as 1 degree are indistinguishable from square elements. Therefore this parameter is not sensitive to the uniform stretching or shearing of elements. From the metric tensor, Kerlick et al. [11] compute separate measures of element aspect ratio, orthogonality, orientation, and volume for finite difference analysis. These separate measures cannot describe the combination of stretching and shearing of elements and are therefore awkward to apply to cases where both types of distortion are present. In addition, aspect ratio and orthogonality are, individually, only partial descriptions of element shape. Describing the shape of an element and thus the degree of distortion present requires both. To attempt to relate element performance to these quantities separately is fundamentally wrong. It may be more descriptive to know them separately but element performance depends on them both together. Saltzman and Brackbill [12] defined three functionals for mesh smoothness, orthogonality and grading. They minimized these functionals to compute an 'optimum' mesh for finite difference analysis. This is equivalent to optimizing the nodal positions for a given problem. This has merit when many analyses must be performed on the same structure, such as aerodynamic design of an aircraft. This strategy is rarely used in FEA. In automatic mesh generators and rezoning programs, the decision to accept or reject a given element must ultimately be reduced to a Boolean decision based on the evaluation of a single scalar measure.

The size of any element must not affect the distortion metric. Whether an element is too large or too small depends on the problem in which it is used. Changing an element's size

does not change the form of the interpolation functions in the physical (x, y, z) domain. For example, degenerate, crack tip elements whose interpolation functions involve the square root of the radial distance from one corner of the element will always have that dependence, regardless of element size. This type of element is inappropriate for general use. When it is appropriate, element size becomes a separate consideration, based on the problem at hand.

To be valid then, a distortion metric should meet the following criteria. It must:

- 1) measure the degree of deviation of the element in the physical (x, y, z) domain from its assumed shape in the computational domain (ξ, η, ζ) ;
- 2) be sensitive to combined stretching and shearing of the element;
- 3) not be affected by rigid body displacements and rotations;
- 4) be independent of the element size; and
- 5) enable a Boolean rejection or acceptance decision to be computed conveniently.

Ideally this parameter would be applicable to two-dimensional (2D) and three-dimensional (3D) elements. Such a distortion metric, based on an invariant of the isoparametric transformation, is presented in this paper. It is computed using the Jacobian of the transformation from the computational domain to the physical domain. This metric meets the criteria listed above and can be applied to 2D and 3D elements.

METRIC FORMULATION

The first step in obtaining a distortion metric is to normalize the Jacobian matrix to remove the effect of element size. Using the standard indicial summation convention:

$$J'_{ij} = \frac{J_{ij}}{|J|^{1/n}} \quad (1)$$

where n is the number of dimensions.

An analogy can be drawn between element distortion and strain. If a body with the shape of the element in the computational (ξ, η, ζ) domain were deformed until it had the shape of the element in physical (x, y, z) domain then elements with large distortions would experience the analogue of large strains. Green's strain is an appropriate measure to examine. If:

$$C_{ij} = J'_{ki} J'_{kj} \quad (2)$$

then

$$e_{ij} = \frac{1}{2}(C_{ij} - \delta_{ij}) \quad (3)$$

where

$$i, j = 1, 2 \text{ in 2D and}$$

$$i, j = 1, 2, 3 \text{ in 3D.}$$

It should be noted that in 2D the tensor has a rank of 2, whereas in 3D the rank is 3. In both 2D and 3D the first invariant of these tensors is a measure of the local change in area or volume. This will be removed since it is the change in shape, not the change in area or volume which is of interest. In 3D this is analogous to computing the deviatoric strains.

$$e_{ij} = \frac{1}{2}(C_{ij} - \frac{1}{n}C_{kk}\delta_{ij}) \quad (4)$$

The second invariant of this tensor J_2 , in 2D or 3D is:

$$J_2 = \frac{1}{2}e_{ij}e_{ij} \quad (5)$$

or,

$$J_2 = \frac{1}{8}(C_{ij}C_{ij} - \frac{1}{n}(C_{kk})^2) \quad (6)$$

The factor of 8 in the denominator serves no purpose in a distortion metric except to scale the values obtained. It is therefore dropped. The distortion metric becomes:

$$D = C_{ij}C_{ij} - \frac{1}{n}(C_{kk})^2 \quad (7)$$

Examination of this distortion metric will show that it is a function of the terms of the Jacobian to the fourth power. Thus D grows rapidly as distortion increases.

EVALUATION

As has already been discussed, it is not possible to relate a given value of distortion to some level of error a priori. This is due to the fact that the error in any analysis is highly

dependent on the problem being analysed. It is well known however, that highly distorted elements, in general, tend to reduce the accuracy. Knowing the value of D which represents excessive distortion would be useful. Three examinations of D were performed. The first draws on the experience contained in the empirical rules mentioned earlier. The second re-examines some tests of element accuracy and distortion contained in the literature. The third demonstrates the effects of combined stretching and shearing of elements.

The aforementioned empirical rules governing acceptable distortion limits were the result of much experimentation and are a reflection of the performance of distorted elements in a wide variety of situations. Although they are by no means definitive rules, they have long been useful in assessing how much distortion should be accepted. The experience they represent can be adapted to the present needs. Computing the value of D for elements that embody these limits will enable some comparison to be drawn between D and element performance. Since the Jacobian matrix is constant everywhere in a parallelogram, it would be useful to compute the value of D for various parallelograms. Figure 1 shows a typical parallelogram of length l , width w and acute angle θ . Table 1 shows the values of D obtained for various values of aspect ratio l/w and acute angle.

The new distortion metric is of most value when dealing with the combination of both stretching and shearing. When such elements are encountered, the value of D computed can be used in conjunction with Table 1 to estimate element performance. Although Table 1 does not constitute proof of poor performance, elements with combined distortion are not likely to perform better than those which were either stretched or sheared alone.

To assess the effect of element distortion on solution accuracy, several authors [7][13][14] have presented load-point displacements for a cantilever beam modelled with various distorted isoparametric quadrilateral elements. These investigations were intended to illustrate how element distortion affects the accuracy of the solution. Elements distorted to varying degrees can be found in them. The meshes used are shown in Figure 2. The results for the tip deflection of these cantilever beams using distorted 8-node quadrilateral elements and 3x3 Gaussian quadrature are reprinted in Table 2. Also included are the largest Gauss point value of D computed for each element, an l_{∞} norm. A correlation between D and solution error will be drawn from these results.

RESULTS AND DISCUSSION

The results presented in Table 1 can be applied in several ways. Before using an automatic mesh generator the user could specify some limit of allowable distortion and be reasonably sure of obtaining a specific level of element quality, regardless the type of distortion present. In large deformation analyses, where the mesh changes shape significantly, it could be used to automatically initiate mesh rezoning at some specified degree of element distortion, before errors degrade the solution. There are two principal advantages in using a single measure of distortion as opposed to several unrelated measures. Most important is the fact that with D the user can be confident that combinations of stretching and shearing do not create elements of dubious quality. Second, aspect ratio and acute angle are individually only partial descriptions of the shape. Both are required to define the shape of an element and therefore it is necessary to combine them when trying to estimate element performance. Either before, during or after an analysis, Table 1 and D allow the user to qualitatively assess the elements present in the mesh.

Many metrics could be used to measure D ; e.g., $L_1, L_2, L_\infty, l_1, l_2, l_\infty$ could be used. It may well turn out that the variation of D within an element is as important as its magnitude. In spite of this we have chosen l_∞ because it is cheap to compute. In a finite element analysis it is most convenient to compute the value of the distortion metric at the Gauss points, whereas, for mesh generation algorithms it may be more convenient to calculate the distortion at the nodes. Since D is based on a point-wise measure of strain it will be a strong function of the element geometry and, when calculated at the Gauss points, a weak function of the integration order. It is possible to encounter elements with zero or negative values of $|J|$. For such elements, D is arbitrarily assigned an infinite positive value.

The comparison of D and cantilever tip deflection is found in Table 2. Before undertaking a detailed examination of the results presented in Table 2, a general discussion of errors and element distortion would be useful. The first example in Table 2 shows a relatively highly distorted element providing adequate results for tip deflection. If the boundary conditions were changed so that instead of bending the beam there was a uniform horizontal load on the end, one would find that there was virtually no error in the tip displacement. In that particular case, the degraded interpolation functions of the distorted element are

capable of capturing the displacement field present in the exact solution. In another case, with different boundary conditions, there might be a substantial error associated with this element. Large values for D do not guarantee large errors under all conditions. A large value of D only indicates that the shape has changed considerably, and with it the interpolation functions in the physical domain. Whether this is good, bad or of no consequence depends on the problem. It would appear that for the problem examined in Table 2, stretching the element causes lower errors than does shearing the element. In other problems the reverse will be true. In any case, the general trend that large distortions are sources of potential error is borne out in Table 2.

Characterizing the elements shown in Figure 2, on the basis of orthogonality or aspect ratio and estimating their performance would be difficult since many elements exhibit combined distortion. Comparing the second example to the third; and the fourth example to the fifth shows that increasing the element distortion while maintaining the total number of elements present can increase the error. In addition, replacing a small number of elements with a larger number of more highly distorted elements does not necessarily reduce the error. This is evident in the second and fifth examples. The best results are obtained by using a large number of undistorted elements. Even so, the first example clearly demonstrates that a relatively highly distorted element may behave well in some problems. As previously discussed, the assumption that large distortion leads to large errors is not always true. It is however, a guideline used by prudent analysts. A very bad element may perform quite well in specific circumstances where the interpolation functions obtained in the physical (x, y, z) domain match the exact functions in the problem at hand.

As an additional example of the usefulness of D , consider the 2D heat conduction mesh shown in Figure 3. The left edge is at 200C and the right edge has a convective boundary condition where the convection coefficient is 1000 W/m²C and the ambient temperature is 20C. The upper and lower edges are insulated. The conduction coefficient was 1.0 W/mC. Also shown are the isotherms from 50C to 175C in 25C intervals. The elements used were 8-node quadrilaterals with 3x3 Gaussian quadrature. The mesh does not appear to have elements which are excessively distorted. In spite of this there is clearly something wrong with the isotherms in the element in the upper left corner. The isotherms in that element are kinked. In this case the combination of stretching and shearing is starting to affect the accuracy of the solution. In that element the largest Gauss point value of D was 6.0 while

the aspect ratio was approximately 2.5:1 and the acute corner angle was approximately 45 degrees. Examining aspect ratio and angle separately, one might not have suspected that this element would behave poorly. Using D one would know that the combination of stretching and shearing was becoming more than might be acceptable.

In many problems involving large deformations it is necessary to create a new mesh part way through the analysis. A variety of remeshing, rezoning algorithms exist. Some wait until after the analysis is complete. The problem is then restarted at some intermediate stage where the operator decides that a new mesh is needed because distortion was excessive^{[15][16]}. This is expensive and inconvenient. The goal of automatic mesh generation and rezoning demands that the algorithm be capable of rezoning the problem without the assistance of the operator. Fully automatic schemes need some criterion to determine where and when remeshing and rezoning is required. Some use element aspect ratios and angles to initiate rezoning^[17]. As has been demonstrated, the use of two or more separate parameters is inconvenient and can lead to a false sense of security. The parameter proposed here would greatly facilitate automatic remeshing. Either the maximum value of D or the average value in an element or group of elements could be used to alert remeshing algorithms to potential problems. What strategy is best will be largely decided by the requirements of the remeshing and rezoning algorithms. It is important to know when and where excessive distortion exists. How to eliminate it is beyond the scope of this paper.

CONCLUSIONS

1. The proposed distortion metric satisfies the five conditions necessary for a useful distortion measure.
2. A large value of D in an element will render that element unsatisfactory for general purpose use, even though it may perform adequately in specific cases.

ACKNOWLEDGEMENTS

The financial support of NSERC A2992 (J. Goldak), NSERC Strategic grant G1259 (J. Goldak), NSERC A4601 (M. Bibby), NSERC Postgraduate Scholarships (A. Oddy) and

Ontario Graduate Scholarships (M. McDill) are acknowledged.

Table 1. Distortion Metric for
2-D Parallelograms

Aspect Ratio l/w	Acute Angle Θ	Distortion Metric D
1 : 1	90	0.00
3 : 1	90	3.56
5 : 1	90	11.52
10 : 1	90	49.01
100 : 1	90	4999.00
1 : 1	60	0.67
1 : 1	45	2.00
1 : 1	30	6.00
1 : 1	15	27.86
1 : 1	5	261.29
3 : 1	45	9.11

Table 2. Cantilever Tip Deflection and
Element Distortion Metric D

Element Configuration	Tip Deflection mm	D			
		Element 1	Element 2	Element 3	Element 4
0.03054	49.0	—	—	—
0.03721	11.5	11.5	—	—
0.00844	38.5	38.5	—	—
0.03877	4.6	4.6	4.6	—
0.01768	16.5	22.0	16.5	—
0.03670	21.0	21.0	3.6	20.0

$P = 100 \text{ N}$, $E = 1.0 \times 10^7 \text{ MPa}$, $\nu = 0.3$, Beam theory tip deflection = 0.040 mm

LIST OF FIGURES

- Figure 1.. Length l , width w and acute angle Θ for a typical parallelogram.
- Figure 2. Cantilever meshes, only corner nodes shown, (dimensions in mm, unit thickness).
- Figure 3. Two dimensional heat conduction example. Isotherms from 50C to 175C in 25C intervals. Note kinked isotherm in top left element.

REFERENCES

- [1] Carey, G.F., and Oden, J.T., Finite Elements, A Second Course, Prentice-Hall, Englewood Cliffs, N.J., Vol. II, 1983, 84.
- [2] Bathe, K. Finite Element Procedures in Engineering Analysis, Prentice-Hall, Englewood Cliffs, N.J., 1982, 220-231.
- [3] Fried, I. Accuracy of Complex Finite Elements. AIAAJ, 10, No. 3, 1972, 347-349.
- [4] Cook, R.D., Concepts and Applications of Finite Elements, J. Wiley and Sons, New York, N.Y., 1981, 409 and 211.
- [5] Barsoum, R.S., On the use of Isoparametric Finite Elements in Linear Fracture Mechanics. Int. J. Num. Meth Engng., 10, 1976, 25-38.
- [6] Turke, D.J., and McNeice, G.M., Guidelines for Selecting Finite Element Grids based on an Optimization Study, J. Comp. St., 4, 1974, 499-519.
- [7] Gifford, L.N., More on Distorted Isoparametric Elements, Int. J. Num. Meth. Engng, 14, 1979, 290-291.
- [8] Zienkiewicz, O.C., The Finite Element Method, 3rd ed., McGraw-Hill, London, U.K., 1976, 186.
- [9] Fried, I., Accuracy and Condition of Curved (Isoparametric) Finite Elements, J. Sou. Vib., 31, No. 3, 1973, 345-355.
- [10] Nicolas V.T., and Citipitioglu, E., A General Isoparametric Finite Element Program SDRC SUPERB, Comp. St., 17, No. 2, 1977, 303-313.
- [11] Kerlick G., and Klopfer G., Assessing the Quality of Curvilinear Coordinate Meshes by Decomposing the Jacobian Matrix, Numerical Grid Generation, J.F. Thompson, ed., 1982, 787-807.
- [12] Saltzman J., and Brackbill, J., Applications and Generalizations of Variational Methods for Generating Adaptive Meshes, Numerical Grid Generation, J.F. Thompson, ed., 1982, 865-884.
- [13] Striklin, J.A., Ho, W.S., Richardson, and E.Q., Haisler, W.E, On Isoparametric vs. Linear Strain Triangular Elements, Int. J. Num. Meth. Engng, 11, 1977, 1041-1043.

- [14] Backlund, J. On Isoparametric Elements, Int. J. Num. Meth. Engng., 12, 1978, 731-732.
- [15] Thorne, B.J., and Holdridge, D.B., The TOOREZ Lagrangian Rezoning Code, Sandia Labs., SLA-73-1057, 1974.
- [16] Swegle, J.W., TOODY IV - A Computer program for Two-Dimensional Wave Propagation, Sandia Labs., SAND-78-0552, 1978.
- [17] Yarrington, P., and Byers, R.K., AUTOREZ: A Two-Dimensional Lagrangian Automatic Rezoning Wavecode, Sandia Labs., SAND-79-1914, 1980.

Figure 1. Length l , width w and acute angle Θ for a typical parallelogram.

Figure 2. Cantilever meshes, only corner nodes shown, (dimensions in mm, unit thickness).

Figure 3.

Two dimensional heat conduction example. Isotherms from 50C to 175C in 25C intervals. Note kinked isotherm in top left element.

END

08-08-88

FIN